Discovery of novel variants underlying inherited bleeding and platelet disorders by next generation sequencing

A thesis submitted to University College London in accordance with the requirements for award of the degree of Doctor of Philosophy in March 2018

Tadbir Kaur Bariana

tadbir.bariana@nhs.net

Supervisors:

Dr. Keith Gomez

Consultant Haematologist and Honorary Senior Lecturer, Royal Free Campus, University College London Medical School, k.gomez@ucl.ac.uk

Professor Willem Ouwehand

Professor of Experimental Haematology, University of Cambridge, who1000@cam.ac.uk

Abstract

Genetic variants that affect megakaryopoiesis and platelet formation result in inherited bleeding and platelet disorders (BPD). Only 40-60% of cases will receive a diagnosis indicating the pathway at fault [1, 2]. This thesis is the result of work undertaken to discover novel variants causative of BPD, as part of the NIHR BioResource-Rare Diseases Study (NIHR BR-RD). 1,213 BPD cases of uncertain genetic aetiology were recruited from 31 international centres. 687 samples were sent for whole exome sequencing and 1118 for whole genome sequencing. Cases were systematically phenotyped and genotype-phenotype relationships were assessed to detect causal variants in known and novel candidate BPD genes, including *KDSR* and *ABCC4*.

KDSR is an early enzyme in the de novo sphingolipid synthetic pathway, and several cases have recently been reported with *KDSR* variants, severe skin pathology and thrombocytopenia, without evaluation of the mechanism of thrombocytopenia [3, 4]. I report a pedigree in which novel compound heterozygous variants in *KDSR* cosegregate with a severe phenotype of neonatal-onset thrombocytopenia in two siblings and juvenile myelofibrosis in the older sibling, without significant skin pathology. Cellular studies support roles for KDSR in the regulation of growth and apoptosis, megakaryocyte yield, maturity, size and proplatelet formation.

This thesis also reports on the study of a pedigree in which coinheritance of homozygous loss of function variants in novel candidate BPD gene *ABCC4* and heterozygous variant in known BPD gene *P2RY12* cosegregate with a lifelong mild-to-moderate bleeding disorder. *ABCC4* is a membrane-bound transport glycoprotein with broad substrate specificity that includes cAMP, a powerful inhibitor of platelet activation, and intra-platelet cAMP was elevated in affected individuals. The role of *ABCC4* in megakaryopoiesis is also explored using CRISPR-Cas9-mediated inhibition of *ABCC4* expression in cellular models of megakaryopoiesis.

This study describes collaborative approaches to the analysis of high throughput sequencing data for the discovery of potentially pathogenic variants, and emphasises the necessity of functional validation of hypothetical associations.

Impact statement

This thesis summarises my involvement in the NIHR-BioResource Rare Diseases Project, which is part of the 100,000 Genomes Project. I have recruited 116/ 1213 (10%) cases on the Bleeding and Platelet Disorders (BPD) arm of the study. I contributed to the design and implementation of a custom phenotyping process for bleeding and platelet disorders that was subsequently adapted and used for all arms of the project, and to the prioritisation of variants identified by statistical genomics approaches. This collaborative strategy has resulted in the detection and feedback of 177 new genetic diagnoses to date in a clinically accredited format to recruiting clinicians. Several novel gene discoveries were made (*SRC* [5], *DIAPH1* [6] and *TPM4* [7]) and further manuscripts have been published describing novel variants and genotype-phenotype analyses in cases with variants in known BPD genes [8-10]. The impact of novel variant discovery is to allow improved diagnosis, tailored management, screening of family members, improved family planning, prediction of clinical progression and detection of serious consequences in a timely manner. Understanding the genetic basis of BPD also improves our knowledge of normal platelet development and physiology and provides potential cardiovascular drug targets.

Further variants are under evaluation by co-segregation, extended cellular phenotyping and cellular modelling, including *KDSR* and *ABCC4* which are the subjects of Chapters 4 and 5.

KDSR is reported to be associated with severe skin disorders and thrombocytopenia [3, 4]. I describe a pedigree in which novel *KDSR* variants co-segregate with a severe thrombocytopenia with minimal skin involvement. The proband has the additional novel phenotype of juvenile onset myelofibrosis. The sphingolipid profile in individuals with *KDSR* variants is reported for the first time, showing loss of function and suggesting pathways by which loss of function is compensated *in vivo*. Lentiviral expression of wild-type *KDSR* in proband –derived induced pluripotent stem cells (iPSC) was followed by reprogramming to megakaryocytes to assess the mechanism by which these variants cause disease. I have shown that *KDSR* variants are associated with increased apoptosis, reduced growth and megakaryocyte yield, transcriptional dysregulation of genes involved in the regulation of the cell cycle and impaired proplatelet formation. These findings are supported by collaborative work with the laboratory of Prof. Kathleen Freson, showing impaired proplatelet formation in bone marrow $CD34^+$ cell-derived megakaryocytes, and thrombocytopenia in zebrafish following *kdsr* morpholino knockdown, supporting the role of *KDSR* as a novel regulator of thrombopoiesis.

I also show the first evidence to my knowledge of human platelet pathology associated with *ABCC4* variants. A pedigree is described in which a homozygous single base pair deletion in gene *ABCC4* is associated with reduced platelet ABCC4 protein expression. Digenic inheritance with a novel P2RY12 receptor variant co-segregates with a mild bleeding disorder, and both genes are shown to contribute to increased intra-platelet cAMP. Further evidence for a key role for ABCC4 in thrombopoiesis is provided by identification of an MK-specific superenhancer in *ABCC4* containing known GWAS variants

that influence platelet crit and count and an expression quantitative trait locus that exhibits allele-specific binding for transcription factor MEIS1 in a megakaryocytic cell line. This suggests a mechanism by which MEIS1 may repress *ABCC4* expression in order to influence thrombopoiesis.

Declaration

I declare that the work in this thesis is my own, unless explicitly stated otherwise. Where information has been derived from other sources, I confirm that this has been indicated in the thesis. Work performed in collaboration with or with the assistance of others is indicated as such.

Word count: 49,658 words

SIGNED:

DATE:

Acknowledgements and dedication

I thank my supervisors, Dr. Keith Gomez, Professor Willem Ouwehand and Professor Amit Nathwani for this opportunity, and for their encouragement, enthusiasm, kindness and patience. I thank the British Society of Haematology and NHS Blood and Transplant for funding this research.

This project was the result of multiple collaborations. I would like to acknowledge the scientific and personal support of Dr. Mattia Frontini, Professor Kathleen Freson, Dr. Cedric Ghevaert, Dr. Andrew Mumford, Professor Mike Laffan, Dr. Kate Downes, Dr. Jose Guererro and Dr. Albert Koulman and their research groups, including Ms. Isabel Rosa, Ms. Rita Tome, Ms. Samantha Farrow, Ms. Frances Burden, Dr. Luigi Grassi, Mr Roman Kreuzhuber, Dr. Denis Seyres, Dr. Romina Petersen, Dr. John Lambourne, Dr. Thomas Moreau, Dr. Amanda Evans, Dr. Holly Foster, Dr. Annett Muller, Ms. Joana Batista, Ms. Carly Kempster, Mrs. Harriet McKinney, Dr. Karyn Megy, Mr Jonathan Stephens, Ms. Chantal Thys, Dr. Will Astle, Dr. Sergio Rodriguez-Cuenca and Dr. Ben Jenkins. Assistance in statistical genetics was gratefully received from Dr. Ernest Turro, Dr. Daniel Greene, Dr. Kathy Stirrups, Dr. Chris Penkett and Mr Tony Attwood. Thanks also go to Dr. Pratima Chowdary and Dr. Ann Riddell of the coagulation laboratory of the Royal Free NHS Trust.

This work is dedicated to my family. To my parents, Mr. Hardial Singh Bariana and Mrs. Mohinder Pal Kaur Bariana, my husband, Mr. Ralph Coulbeck and my mother-in-law, Mrs Susan Coulbeck, all of whom worked so hard to look after my children when I could not. And of course, equally to my children, Cordelia and Samuel, for their love and their understanding.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1 Introduction

1.1 Platelets

Historically platelets were thought to be cellular 'dust.' We now understand that adequate platelet number and function are vital for normal haemostasis, and that platelets have further roles to play in inflammation, wound healing and angiogenesis [11]. Human platelets are small, discoid, anucleate blood cells that measure approximately 3.0 x 0.5μm and have a mean volume of 7-11 fL. The number of platelets in healthy adult blood ranges between 150-450 x 10^9/L, with a mean lifespan of 10 days in circulation [12]. Approximately 10^{11} platelets are formed every day from the cytoplasm of their precursor cells, the megakaryocytes (MK), though production can be increased more than tenfold under conditions of stress such as during immune-mediated thrombocytopenia [13].

1.2 The origin of platelets

1.2.1 Megakaryopoiesis

The cell of origin of all human blood cells is the haematopoietic stem cell (HSC). HSC self-renew in order to remain multi-potential and provide a continuous supply of blood cells throughout life. MK derive from HSC and are terminally differentiated myeloid cells that produce platelets. They are the largest cells in the bone marrow (50-100μm) and constitute <0.01% of nucleated bone marrow cells [14, 15]. The bone marrow is the major site of production, however MK are also found in the lung, spleen and liver. The classical model of megakaryopoiesis states that HSC-derived multi-potential progenitors commit to the myeloid lineage, form bipotent megakaryocyte-erythrocyte progenitors (MEP) that subsequently mature to MK and erythroblasts. However, several recent studies have contended that multiple pathways may produce MK, that these pathways may not require an MEP stage and that MK can arise directly from HSC [16]. A myeloid-restricted subpopulation of HSC has been identified that gives rise directly to an MK-restricted lineage that, interestingly, increases with age [17, 18]. Lineage commitment and differentiation are regulated by an orchestra of transcription factors, including GATA1, FOG1, RUNX1, FLI1, TAL1, EV1, MYB, NF-E2 and MEIS1 [19-21]. Growth factors also play a major role, the most important of which is thrombopoietin. The maturation process of MK is unique and remarkable, involving polyploidization by endomitosis, the formation of an extensive, intricate invaginated membrane system, the formation and packaging of platelet-specific organelles and the synthesis of proplatelets that finally release platelets into the circulation [15].

1.2.2 Thrombopoiesis

Dynamic reorganization of the MK cytoskeleton gives rise to proplatelets which are long, branching pseudopodial projections comprising cytoplasm encased in excess membrane from the invaginated membrane system that contain a central core of microtubular bundles. Each MK extends 10-20 proplatelets, carrying cytoplasm and organelles, including pre-packaged specific granules, from the body of the MK along the shaft to the tip of the proplatelet (Figure 1a) [22, 23]. Studies of impaired proplatelet formation in congenital thrombocytopenia caused by variants in *MYH9* [24], *ACTN1* [25], *FLNA* [26] ,

TUBB1 [27], *DIAPH1* [6], *WAS* [28] and *RUNX1* [29] have identified these genes as regulators of this process, and both platelet number and shape are abnormal in these conditions. Interactions with the bone marrow microenvironment are also regulatory, via fibrinogen [30], von Willebrand factor [31], fibronectin [32] and by podosome formation. Podosomes are small, round actin-rich plasma membrane protrusions that adhere to and break down the extracellular matrix to allow proplatelets across the basement membrane and into the bone marrow sinusoid [33]. Blood components also appear to have a role to play. For example, gradients of the plasma sphingolipid sphingosine-1-phosphate (S1P) have been shown to guide proplatelet extensions into the bone marrow sinusoids. S1P receptor 1 (S1PR1) -deficient mice have reduced proplatelet formation and platelet release, severe thrombocytopenia, and S1PR1 activation directly leads to platelet release [34]. Avecilla *et al.* also showed that chemokines FGF-4 and SDF-1 promote MK localization in the vascular niche as well as platelet release [35]. Shear forces in the vascular compartment are also likely to assist in proplatelet breakdown to form platelets [36]. The initial fragments released from MK vary in size and shape, and continue to fragment in the peripheral circulation to achieve mature platelet size and structure [37, 38]. The cytoplasm of the MK is gradually consumed by this process until only bare nuclei remain, that are then degraded.

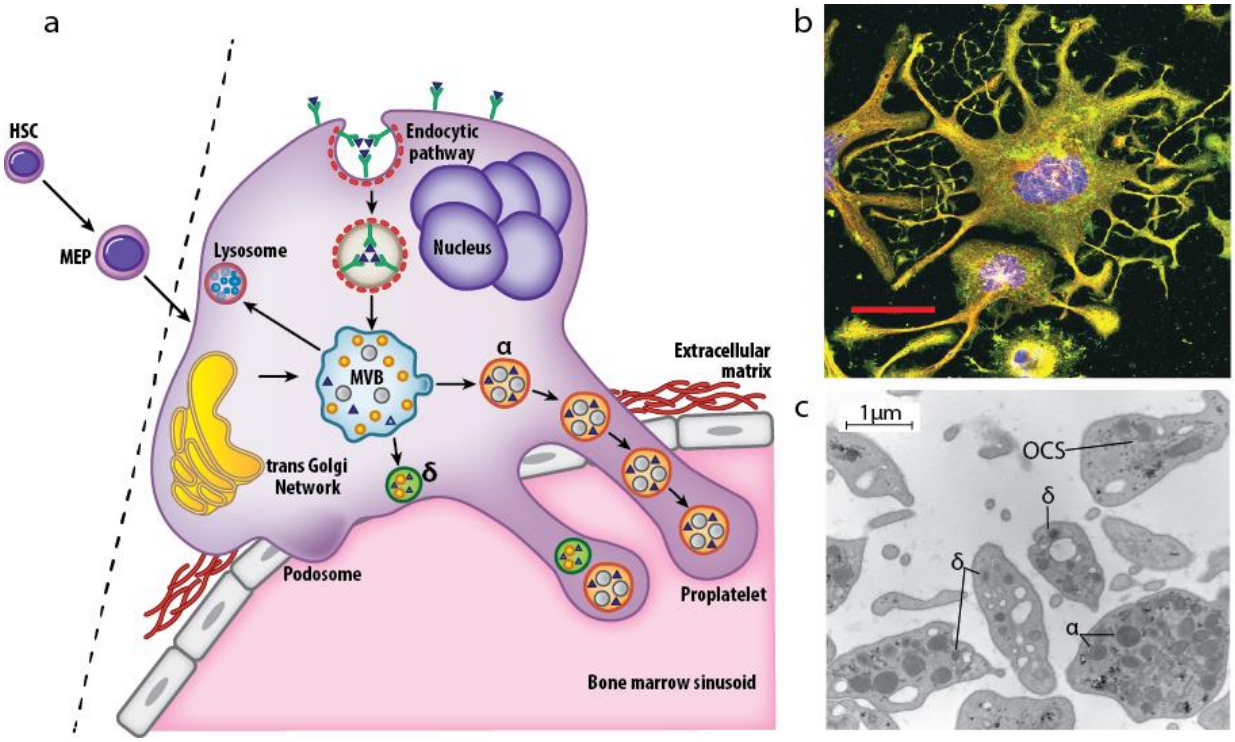


Figure 1: Megakaryocyte, proplatelet and platelet structure.

(a) The cartoon summarises the differentiation of haematopoietic stem cell (HSC) to megakaryocyte-erythroid precursor (MEP) to megakaryocyte (MK). In the MK the dual contribution of the endocytic pathway and the trans-Golgi network to the development of the multivesicular body (MVB) during the synthesis of specific granules is shown, with subsequent differentiation into α-, δ- and lysosomal granules. The characteristic polyploid nucleus, podosome-mediated tethering to the extracellular matrix and proplatelet extensions into the bone marrow sinusoid are illustrated. Created with illustrative assistance by Jo Westmoreland (MRC Laboratory of Molecular Biology, Cambridge.) Reproduced with permission from the British Journal of Haematology [39].

(b) Proplatelet formation in a mouse MK demonstrating proplatelet formation following directed differentiation of bone marrow-derived $CD34^+$ cells. Stained with antibodies against alpha-tubulin (cytoskeleton), CD41 (membrane) and DAPI (nucleus) and mounted on fibrinogen. Image provided by Mr Cavan Bennett, University of Cambridge. Red bar indicates 50µm.

(c) Transmission electron microscopic image demonstrating normal platelet ultrastructure. δ – dense granule. α – alpha granule. OCS – open canalicular system. Mitochondria are not seen in these sections.

1.2.3 Platelet structure

Platelets comprise a plasma membrane that expresses numerous receptors that regulate platelet activation and signaling. Pro-coagulant phospholipids such as phosphatidylserine are concentrated in the inner leaflet of the membrane at rest. The cytoplasm is segregated by a cell surface-connected, invaginated membrane known as the open canalicular system that has roles that include the expulsion of granule contents and supply of extra membrane required during platelet spreading. They contain mitochondria and a residual endoplasmic reticulum (ER). The ER is the major intracellular store of calcium, a key signaling molecule during platelet activation. The platelet cytoskeleton consists of a spectrin-based support system for platelet membranes, an actin-based network throughout the cytoplasm and a marginal band of alpha-beta-tubulin-rich microtubules. This intricate network is responsible for the dramatic shape change that occurs during platelet activation and spreading [40].

There are three types of intracellular secretory granule in platelets, the alpha (α-) granules, dense (δ-) granules and lysosomes. α– granules are the largest and most abundant and usually appear as 200-500 nm diameter spherical or ovoid organelles. There are usually 50-80 α– granules per platelet and they contain hundreds of proteins that mediate haemostasis, inflammation, wound healing and angiogenesis. Overall it appears that agonists and antagonists of haemostasis co-localise within granules and are secreted simultaneously. It is likely that external factors regulate the activity of granule contents in the face of a haemostatic challenge [41, 42]. δ- granules are smaller, at approximately 250 nm in size and have characteristic electron-dense cores. There are usually around 10 per platelet and they contain a range of smaller metabolites including ADP, ATP, serotonin and calcium. Lysosomes are the least abundant, with at most 3 per platelet. Unlike the other granules which are platelet-specific, lysosomes are ubiquitous and contain catabolic enzymes such as proteases and glycosidases that degrade and recycle cellular components. Their function in mature platelets is less well-characterised but is now emerging, with putative roles in inflammation, arterial thrombosis and macrophage biology [43].

1.2.4 Platelet function

Platelets adhere via surface receptors to exposed sub-endothelial matrix at the site of injury, first loosely via interaction between the platelet GP1b receptor and von Willebrand Factor (VWF), then stably via interaction between the GPVI receptor and collagen. Adhesion stimulates a cascade of intracellular signaling that leads to platelet spreading, secretion of platelet granules containing autocrine and paracrine activating factors such as adenosine diphosphate (ADP) and thromboxane A_2 (TXA_2), and aggregation of platelets, mediated by platelet surface integrin GPIIb/IIIa, fibrinogen and vWF. Activation of platelets also stimulates exposure of phosphatidylserine, providing a procoagulant surface for thrombin generation. Thrombin in turn amplifies platelet activation by stimulating platelet thrombin receptors PAR1 and PAR4. It also cleaves circulating and platelet-secreted fibrinogen to form fibrin, which polymerises to form a mesh that stabilises the incumbent platelet plug. Subsequent integrin-driven contraction of the platelet cytoskeleton leads to clot retraction [44].

1.2.5 cAMP, a major regulator of platelet activation

Cyclic adenosine monophosphate (cAMP) is a central, powerful inhibitor of platelet activation. Figure 2 illustrates our incomplete knowledge of the regulation of this complex pathway.

Prostacyclin, adenosine, and pituitary adenylate cyclase-activating peptide act upon their respective G_s-protein-coupled receptors to activate the $G_s\alpha$ subunit, which in its GTP-bound state can activate adenylate cyclase (AC) to synthesise cAMP [45-47]. Several AC isoforms are present in platelets, although their individual roles remain uncharacterised [48]. The major target of cAMP in platelets is protein kinase A (PKA). PKA is an inactive tetramer at rest, comprising two regulatory units inhibiting two catalytic units. cAMP binds the regulatory units and relieves this inhibition, activating the kinase function of the protein to phosphorylate multiple downstream targets that inhibit platelet adhesion, activation and aggregation [49]. There are numerous known substrates of PKA in platelets, however the mechanisms of action remain incompletely understood [50]. Inositol 1,4,5-trisphosphate receptor is a target, and has been shown to inhibit calcium release from platelet intracellular stores [51]. Rap1GAP2 and GP1Bβ are phosphorylated and have known roles in cell adhesion, so may mediate cAMP-inhibition of platelet aggregation [52, 53]. Vasodilator-stimulated protein (VASP) is phosphorylated and is known to interact with actin, so may inhibit platelet spreading and granule release [54, 55]. Increased cAMP also reduces thrombin receptor binding and jthrombin-induced responses, although the pathway is poorly understood [56].

Naturally, opposing mechanisms exist to reduce intracellular cAMP to allow platelet activation. There is a constant turnover of cAMP due to degradation by phosphodiesterases PDE2A and PDE3A [48], and their activity is highly up-regulated during platelet activation [57]. For example, adenosine diphosphate (ADP), epinephrine and bind their respective G_i-protein-coupled receptors to activate their $G_i\alpha$ subunits, which bind and inhibit AC [50]. Regulator of G-protein signaling 2 also contributes by hydrolysing $G_s\alpha$-GTP to the inactive $G_s\alpha$-GDP [58]. Intra-platelet cAMP concentrations are therefore carefully controlled to minimise spontaneous platelet activation, whilst maintaining the capability to provide effective haemostasis on demand.

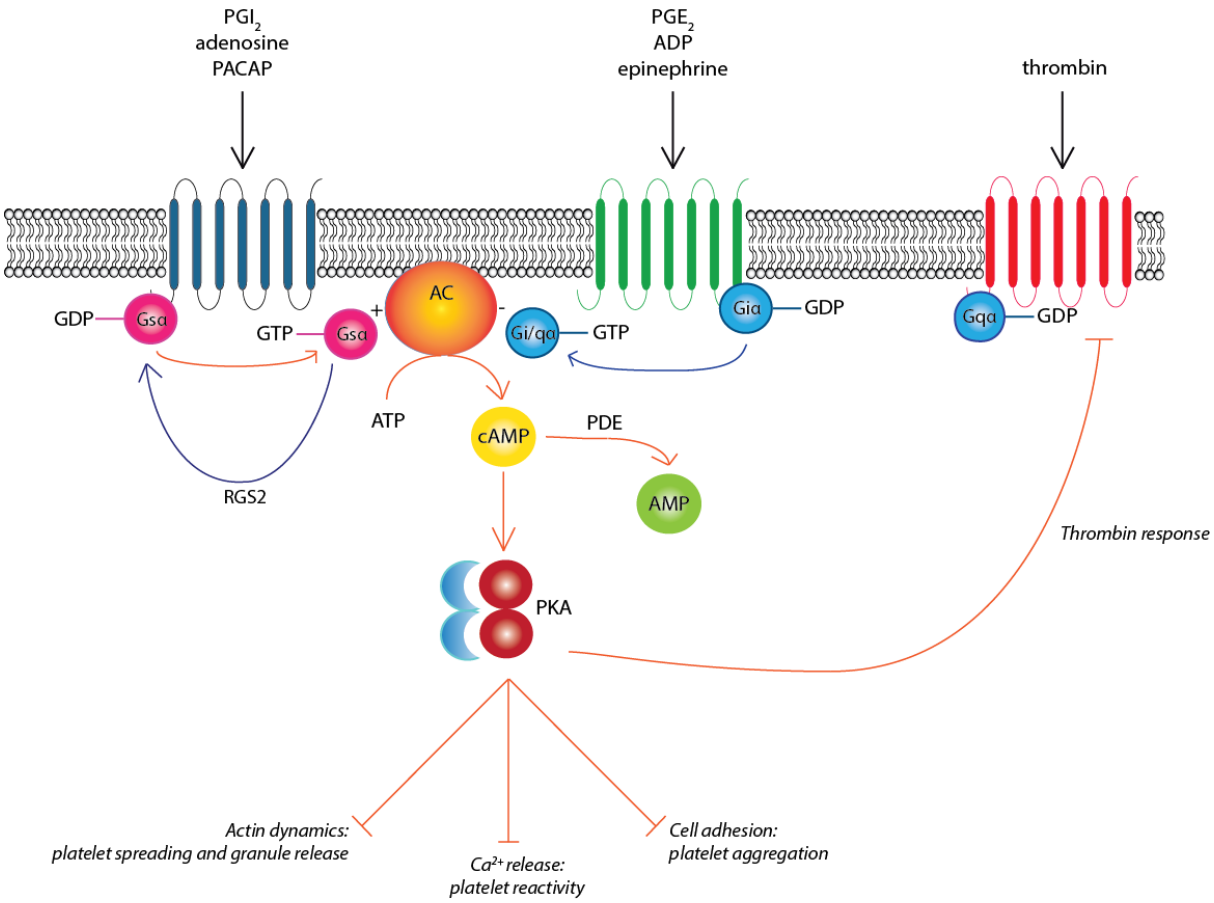


Figure 2 Platelet cAMP regulation.
Prostacyclin (PGI_2), adenosine, and pituitary adenylate cyclase-activating peptide (PACAP) act upon G_s-protein-coupled receptors to activate $G_s\alpha$ to synthesise GTP. $G_s\alpha$-GTP activates adenylate cyclase (AC) to synthesise cAMP [45-47]. cAMP activates the tetrameric protein kinase A (PKA) by inducing release of regulator subunits. PKA phosphorylates targets to inhibit platelet adhesion, activation and aggregation [49]. Increased cAMP also reduces thrombin receptor binding and thrombin-induced responses, although the pathway is poorly understood [56]. Pathways that reduce cAMP include phosphodiesterase (PDE) -mediated degradation, inhibition of AC by GPCRs coupled to $G_i\alpha$ and $G_q\alpha$ [50] and regulator of G-protein signaling 2 (RGS2) –mediated hydrolysis of $G_s\alpha$-GTP to $G_s\alpha$-GDP [58]. Adapted from Noé *et al*., 2010 [49].

1.2.6 Inherited disorders of platelet-dependent haemostasis

Inherited bleeding and platelet disorders (BPD) are clinically heterogeneous. Patients usually present with a history of bleeding or easy bruising, often with a family history of bleeding suggestive of a genetic aetiology. Bleeding is usually mucocutaneous (epistaxis, menorrhagia, easy bruising) or in response to haemostatic challenges, such as tooth extraction, surgery and childbirth. As a result, the bleeding history may be lifelong or only noticed in later life following a specific challenge [12]. Although classified as mild, in some cases bleeding can be life-threatening. They are often associated with varied multi-system phenotypes that include immune, neurological, ocular and skeletal features [59].

To date the UKHCDO has received 1402 registrations of platelet function disorders and 124 registrations of unclassified bleeding disorders, suggesting a total prevalence greater than that of Haemophilia B [60]. Inherited thrombocytopenias are less frequent, constituting less than 5% of patients presenting with persistent thrombocytopenia [61]. It is likely that the number of cases with inherited platelet disorders is underestimated. Presentation usually requires exposure to significant haemostatic challenges and diagnosis requires a low threshold of suspicion in referring clinicians and access to a specialist laboratory. In the case of inherited thrombocytopenia misdiagnosis as immune-mediated thrombocytopenia is disconcertingly common and results in inappropriate therapy including steroids, immunoglobulin, and even splenectomy. The lack of a genetic diagnosis limits the ability to tailor management, screen family members, aid with family planning, predict clinical progression and detect serious consequences such as myelofibrosis, lung fibrosis and malignancy in a timely manner.

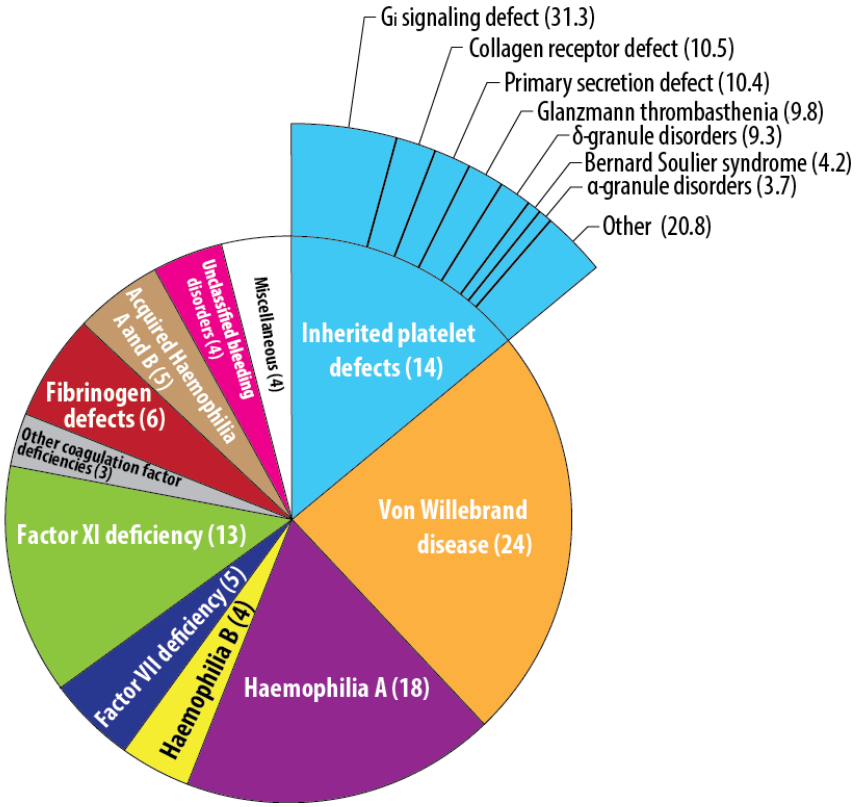


Figure 3 Recorded incidence of bleeding and platelet disorders**.**
Central circle represents annual registration data from UK Haemophilia Centres Doctors Organisation (UKHCDO) 2015 Bleeding Disorder Statistics Report (percentages are given in brackets). Haemophilia A and B data includes low level carriers and Von Willebrand disease (VWD) data includes 'probable' cases. Outer circle represents data from Gresele *et al.* [59], and is a breakdown of annual presentations of inherited platelet disorders. 'Other' platelet disorders each had a frequency of <3.5% and include platelet-type VWD, combined α- and δ- granule disorders and several platelet-related syndromes. Created with illustrative assistance by Jo Westmoreland (MRC Laboratory of Molecular Biology, Cambridge.) Reproduced with permission from the British Journal of Haematology [39].

1.3 The current diagnostic process

A key motivation for this study is the inadequacy of the diagnostic process for a patient presenting with a personal or family history of excessive mucocutaneous bleeding. The value of the bleeding history is limited, restricted by exposure to haemostatic challenges which usually accrue with age. The usual process of physical examination, blood count and film morphology, electron microscopy, platelet nucleotides and flow cytometry for surface glycoproteins is rarely diagnostic. PFA-100 and similar automated platelet function measurement devices lack adequate diagnostic specificity and sensitivity [62].

Light-transmission aggregometry remains the mainstay of current diagnostic testing, yet is expensive, labour intensive and of limited availability, as it requires a highly specialist laboratory. It is prone to false positives or negatives due to interference from a number of variables including a multitude of drugs, foodstuffs, platelet count and volume, the level of fibrinogen and inadvertent platelet activation. It is not able to simulate *in vivo* platelet activation under physiological shear and can only detect abnormalities within historically recognised signaling pathways. It cannot detect abnormalities in mechanisms such as inadequate vessel wall integrity or mutations in genes encoding functionally unidentified plasma proteins that may underlie the bleeding diathesis. Inter-laboratory standardisation remains suboptimal though attempts are being made to improve this [63, 64].

This cascade of highly specialised diagnostic tests in its entirety is only available in some tertiary referral centres and it can take several months with multiple takings of samples to arrive at a diagnosis. The main drawback of this approach is that it rarely leads to identification of the underlying biological abnormality with the result that clinically heterogeneous conditions are unhelpfully lumped together under generic terms such as 'platelet function defect.' In reality inconsistent results or normal laboratory markers despite clearly abnormal clinical bleeding mean only 40-60% of cases will receive a diagnosis indicating the pathway at fault [1, 2]. The lack of a genetic diagnosis limits the ability to tailor management, screen family members, aid with family planning, predict clinical progression and, importantly, detect serious consequences in a timely manner, for example myelofibrosis in GPS, lung fibrosis in HPS and haematological malignancy in *ANKRD26*, *ETV6* or *RUNX1*-related thrombocytopenia.

1.4 Diagnostic high throughput sequencing

The International Society on Thrombosis and Haemostasis currently recommend molecular studies as a 3rd line investigation [65]. This refers to targeted Sanger sequencing to confirm a suspected clinical genetic diagnosis, based upon the results of several previous lines of investigation, as detailed above. Recently the ThromboGenomics consortium (www.thrombogenomics.org.uk) reported the results of a high-throughput sequencing platform [66] that allows simultaneous molecular screening of all known genes implicated in inherited BPD on one blood sample. 65 BPD genes are currently included in the report and the platform detects single nucleotide variants (SNVs), small insertions and deletions (indels)

and copy number variants with good sensitivity and reproducibility. The results obtained with another high-throughput sequencing panel test focusing on a subset of 23 genes have also recently been reported [67]. Widespread uptake of such panels lends immediate clinical value to the search for even ultra-rare novel BPD genes, as discoveries can be applied to panels and screened for alongside more common variants.

Major role	Gene name	Gene symbol
Early Megakaryopoiesis	Thrombocytopenia and thrombocythemia 1	THPO
	Congenital amegakaryocytic thrombocytopenia (CAMT)	MPL
Regulation of transcription	Paris-Trousseau thrombocytopenia and Jacobsen syndrome	FLI1
	X-linked thrombocytopenia with dyserythropoiesis	GATA1
	Autosomal dominant thrombocytopenia 2	ANKRD26
	Familial platelet disorder with predisposition to AML	RUNX1
	Gray platelet-like syndrome	GFI1B
	Thrombocytopenia and susceptibility to cancer	ETV6
	Thrombocytopenia-absent radius syndrome (TAR)	RBM8A
	Radioulnar synostosis with amegakaryocytic thrombocytopenia 1	HOXA11
	Radioulnar synostosis with amegakaryocytic thrombocytopenia 2	MECOM
Granule biogenesis	Hermansky-Pudlak syndrome	AP3B1
		AP3D1
		BLOC1S3
		BLOC1S6
		DTNBP1
		HPS1
		HPS3
		HPS4
		HPS5
		HPS6
	Chediak-Higashi syndrome (CHS)	LYST
	ARC syndrome	VIPAS39
		VPS33B
	Familial hemophagocytic lymphohistiocytosis type 5	STXBP2
	Gray platelet syndrome	NBEAL2
	Bleeding disorder, platelet-type, 20	SLFN14
	Dense granule abnormality	NBEA
Cytoskeletal regulation	May-Hegglin and other MYH9 disorders	MYH9
	Wiskott-Aldrich syndrome	WAS
	Macrothrombocytopenia	ACTN1
		FLNA
		TPM4
	Macrothrombocytopenia, Beta-tubulin 1 related	TUBB1
	Thrombocytopenia & Immune Deficiency	ARPC1B
	Macrothrombocytopenia and sensorineural hearing loss	DIAPH1
Glycoprotein signalling	Autosomal dominant macrothrombocytopenia and Bernard Soulier syndrome	GP1BA
		GP1BB
		GP9
	Von Willebrand Disease	VWF
	ADP receptor defect	P2RY12
	Thromboxane A2 receptor defect	TBXA2R
	Ghosal syndrome	TBXAS1
	Deficiency of phospholipase A2, group IV A	PLA2G4A
	Glanzmann thrombasthenia	ITGA2B
		ITGB3
	Platelet-type bleeding disorder 18	RASGRP2
	Leukocyte integrin adhesion deficiency, type III	FERMT3
	Thrombocytopenia, anemia and myelofibrosis	MPIG6B
Other	Bleeding diathesis due to glycoprotein VI deficiency	GP6
	Ehlers-Danlos syndrome	COL1A1
		COL5A1
		COL5A2
		CHST14
		COL3A1
	Quebec platelet disorder	PLAU
	Stormorken syndrome	ORAI1
		STIM1
	Autosomal dominant thrombocytopenia 4	CYCS
	Myopathy associated with thrombocytopenia	GNE
	Noonan Syndrome	PTPN11
	Scott syndrome	ANO6
	Sitosterolemia & Macrothrombocytopenia	ABCG5
		ABCG8
	Thrombocytopenia 3	FYB

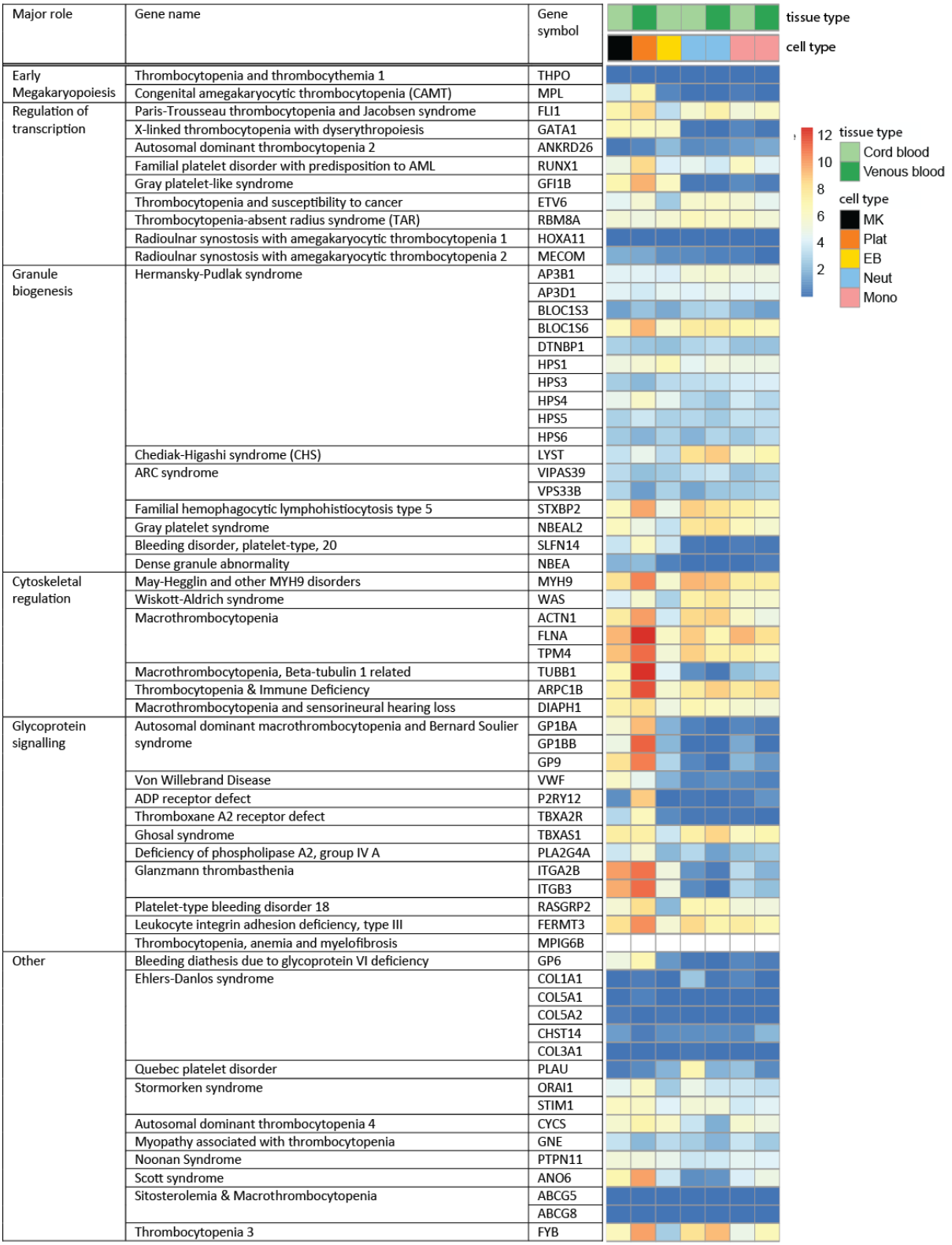


Table 1 65 genes involved in bleeding and platelet disorders, grouped by major function.
The ThromboGenomics consortium (www.thrombogenomics.org.uk) [66] allows simultaneous molecular screening of all 65 genes implicated in inherited BPD on one blood sample. Heatmap, provided courtesy of Dr. Luigi Grassi, University of Cambridge, shows expression levels determined by RNA sequencing (RNA-seq) expressed as $\log_2$ of fragments per kilobase of transcript per million reads mapped (FPKM). The FPKM values range from high in red to low in blue (inset key). RNA-seq transcript levels for blood cells were retrieved from the Blueprint RNA-seq data (http://dcc.blueprint-epigenome.eu/#/datasets) at the EMBL European Genome-phenome Archive. The RNA-seq data for all samples were processed through the same analysis pipeline using Trim Galore 0.3.7. Trimmed reads were aligned to the Ensembl v70 human transcriptome using Bowtie 1.0.1 [68], MMSEQ 1.0.8a [69, 70] was used to quantify gene expression and Mmdiff was used to summarise and normalize gene expression across cell types. Figure adapted from www.thrombogenomics.org.uk [66, 71].

1.5 Genomic discovery through integration of big data

The causes of numerous genetically determined BPD remain unknown. In a large collection of 519 patients with undiagnosed BPD with a high likelihood of a genetic cause based on family history, age of presentation and multi-system phenotypes, exome sequencing identified a variant in a known gene in less than 10% of cases, with many considered to be variants of unknown significance [72]. The clinical and laboratory variation between patients with inherited thrombocytopenias, platelet function disorders and unclassified bleeding disorders suggests that they are a genetically heterogeneous group, and the number of patients with a given genotype at any one centre is likely to be small. Multi-centre collaboration should improve the statistical power for identification of pathogenic mutations.

1.6 The NIHR BioResource- Rare Diseases project (formerly the BRIDGE Study Group)

This thesis is the result of work undertaken on the BPD arm of the NIHR BioResource-Rare Diseases Study (NIHR BR-RD; formerly known as the BRIDGE Consortium: Biomedical Research Centre Rare Inherited Diseases and Genetic Evaluation). The NIHR BioResource Rare Diseases project constitutes the pilot phase of the 100 000 Genomes Project, which aims to sequence the whole genomes of approximately 60,000 National Health Service patients with rare diseases and their healthy relatives, 20,000 patients with cancer and 20,000 genomes derived from the cancers these individuals suffer from [73]. The project commenced recruitment and exome sequencing in 2012, followed by whole genome sequencing from 2014.

This project is large and collaborative. My role within this project was to design methods and tools for consistent, accurate and detailed phenotyping, to use these tools to phenotype patients I had recruited, to apply algorithms generated by bioinformaticians to analyse potential gene discoveries and to functionally corroborate or refute these discoveries using patient-derived samples and *in vivo* models of megakaryopoiesis.

1.7 Phenotyping

1.7.1 The Human Phenotype Ontology

Isolating phenotypically similar groups that are likely to share variants in a single or a few interacting genes in large cohorts requires computational approaches, therefore the first step in gene discovery is the systematic coding of cases using a controlled, standardised vocabulary or ontology.

Several ontologies are available. The World Health Organisation International Classification of diseases (ICD) (http://www.who.int/classifications/icd) and International Health Terminology Standards Development Organisation Systematized Nomenclature of Medical Clinical Terminology (SNOMED-CT) (https://www.ihtsdo.org/snomed-ct) provide a post-diagnosis disease classification, rather than using constituent clinical or laboratory phenotypes. Similar post-genetic diagnosis ontologies such as the Online Mendelian Inheritance in Man (OMIM) (http://www.ncbi.nlm.nih.gov/omim) and Orphanet [74]

databases provide gene-associated prosaic descriptions rather than systematic labels. Gene Ontology (GO) terms (https://geneontology.org) provides systematic computational classification for biological processes, cellular functions and molecular functions attributable to genes but does not cater for clinical or pathological descriptions. Similarly the scope of the Bleeding History Phenotype Ontology (https://bioportal.bioontology.org/ontologies/BHO), as the name suggests, is restricted to bleeding symptoms and coding of BPD requires a wider remit, including extra-haematological features and laboratory abnormalities.

The Human Phenome Ontology (HPO) [20] is designed to systematically apply phenotypes to individuals with heritable disorders, in order to facilitate computational analysis and comparison of cases in the literature. It was originally based on the Online Mendelian Inheritance in Man (OMIM) database, and has been rewritten using a controlled vocabulary by geneticists and specialists in the relevant fields. HPO terms are constructed in 'trees', from general descriptive terms at the top to specific terms at the bottom, as illustrated in Figure 4 [75]. Operators can select a leaf from anywhere on the tree, allowing for situations when details are not known. Algorithms can be designed to account for differences in the level of detail specified, in order to provide lists of cases that share either broad phenotypic labels, e.g. cardiovascular anomaly, or more specific labels e.g. atrio-ventricular septal defect, in order to link cases and their associated variants. Coding patients allows clustering of patients with similar and distinct phenotypes, under the hypothesis that similar phenotypes are caused by variants in the same gene. Whilst useful for categorical coding of diagnoses for gene discovery, HPO does not code severity. In terms of bleeding disorders, the inability to code frequency of bleeding episodes and the frequency, amount and invasiveness of therapy administered is a significant limitation, particularly for clinical use.

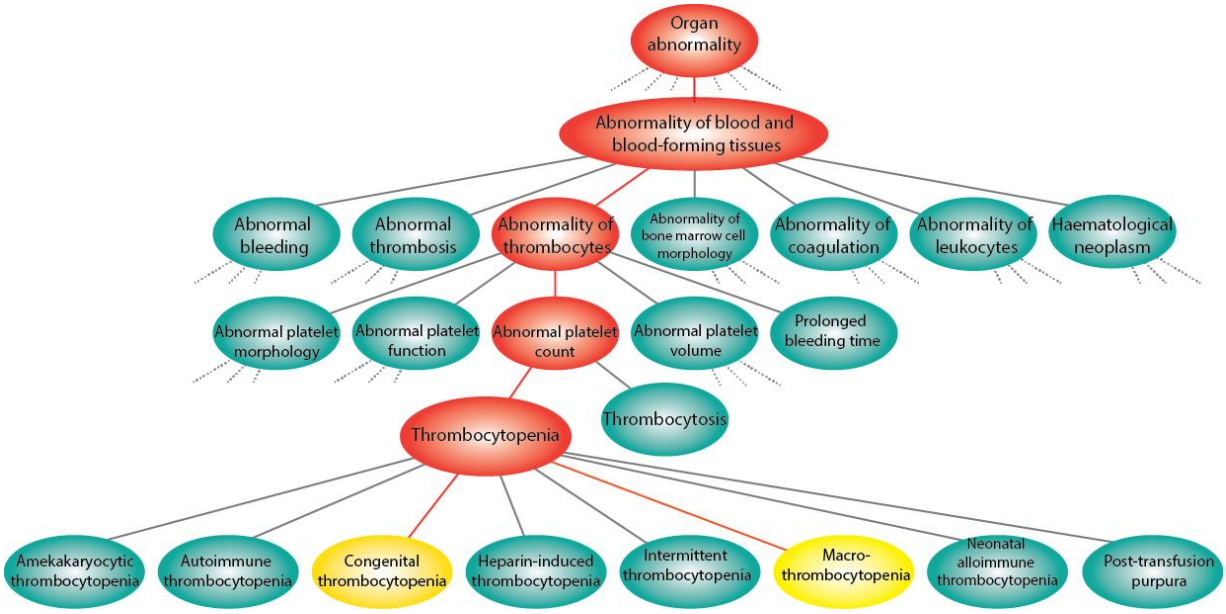


Figure 4 The Human Phenotype Ontology (HPO).
Tree structure showing parent terms (red) leading to target phenotypes (yellow) and alternative terms at each stage (blue). The restrictive coding of phenotypes facilitates automated analysis. Algorithms can account for shared parent terms to allow for variability in coding practices.

1.7.2 Bleeding scores

Coding phenotype severity is important to our study because BPD are often mild, and therefore present a greater challenge in differentiating a significant bleeding history from the bleeding symptoms commonly reported in normal individuals [76]. There are many factors that determine whether or not patients report bleeding symptoms [76]. Quantifying symptoms using bleeding assessment tools (BATs) may be one way of improving diagnostic accuracy since this approach provides a standardized measurement of the diversity and severity of bleeding symptoms expressed as a single numerical parameter termed the bleeding score.

Bleeding scores were successfully used by the Vicenza [77] and MCMDM-1 VWD [78] study groups to discriminate between patients with von Willebrand disease (VWD) and healthy controls. Bowman *et al.* designed a 'condensed,' more clinically applicable version of the MCMDM-1 VWD score that takes 5-10 minutes to complete for experienced users, as opposed to 40 minutes for the original score [79]. When applied prospectively, a condensed score ≥ 4 demonstrated a sensitivity of 100% and a specificity of 87% for type 1 VWD and correlated well with bleeding scores obtained using the full length BAT. Other studies have applied similar tools prospectively in paediatric VWD patients [80, 81] and to women with unexplained menorrhagia [82] with some success.

Since VWD and platelet function disorders (PFD) are both disorders of platelet-dependent haemostasis, it is likely that if the bleeding score is useful for diagnosis of VWD, then the same will hold true for PFD. To our knowledge there have been few studies of the role of bleeding scores in PFD. Tossetto *et al.* demonstrated the excellent negative predictive value of the condensed MCMDM-1 VWD score and a normal APTT when applied prospectively to adults referred for haemostatic evaluation of bleeding symptoms. 14 patients were found to have PFD, but specific scores for this group of patients were not reported [83]. Marcus *et al.* published data on the prospective application of a Vicenza-style score to the diagnosis of PFD in children, defining an abnormal score as >3 in males and >5 in females. The negative predictive value was 90.7% (95% CI 81.7-96.2) but this was no better than the simpler definition of more than two symptoms [81]. Most studies to date have been hampered by low patient numbers, partly due to the rarity of cases and because multicentre collaboration is limited by lack of standardisation of the diagnostic work-up, particularly platelet function testing. The largest study to date reported that amongst 79 adults referred prospectively for haemostatic evaluation at a central reference laboratory in the UK GAPP study, 52% were found to have a PFD on laboratory testing and the remainder, bleeding of unknown origin. The ISTH BAT did not predict the presence of a PFD [84].

1.8 Next generation sequencing

Traditional linkage mapping as an approach for gene discovery relies upon large, informative pedigrees which are rare in BPD, and this approach has been exhausted. Identifying candidate genes from the laboratory phenotype has been highly successful in the past, but is a slow process with diminishing

returns as the pathways in the remaining disorders may not be understood. A substantial proportion of BPD genes remain undiscovered, and we need a strategy for these.

Next generation sequencing technology (NGST) refers to a revolution in genomics that has vastly superseded conventional Sanger sequencing in terms of efficiency. We initially applied NGST in the form of whole exome sequencing for 18 months, and subsequently performed whole genome sequencing. Sequencing of a whole genome can occur in a few hours, rapidly generating large quantities of sequence data. Interpretation relies upon isolating clinically relevant variants using algorithms to relate phenotype to genotype.

1.8.1 Whole exome sequencing

Whole exome sequencing (WES) is the sequencing of all protein-coding regions in the human genome. This constitutes about 2% of the genome. Many WES libraries also capture splice site and 3' and 5' untranslated regions (UTR). WES became technically possible in 2009 [85] and can now be applied widely because it is affordable. It has already successfully identified causative loci for a wide range of inherited disorders, reflected by almost 2000 new entries in OMIM since 2008 [86]. For example, WES in a small number of BPD cases identified the *NBEAL2* and *RBM8A* genes as being causative of Gray Platelet Syndrome (GPS) [87] and Thrombocytopenia and Absent Radius (TAR) Syndrome [88], respectively. GPS was revealed as a classic example of homozygous or compound heterozygous mutations that led to a loss-of-function of the NBEAL2 protein. Morpholino silencing of *nbeal2* in *Danio rerio* resulted in severe tail bleeding. The TAR discovery showed that complex genetic mechanisms not yet reported in the human genetics literature may underlie platelet-related syndromes; in 2007 Klopocki *et al.* [89] observed that a group of 30 patients with TAR all had an obligatory large deletion on chromosome 1q21.1 which was necessary but not sufficient for development of TAR. The other genetic event remained elusive at that time. The WES array allowed reasonable coverage of the 3'UTR, allowing the group to identify a 3'UTR SNP in the *RBM8A* gene as being consistently associated with TAR. This SNP, which is present in 1 in 30 healthy individuals, was identified on the other haplotype to the deletion and introduces a binding site for the transcription factor EVI1, which at this position functions as a repressor of the transcription of *RBM8A*. The reason why they succeeded in discovering the causative genes for these two disorders, where classical genetic approaches had not, is because they observed the entire coding sequence in four patients with each condition.

Most disease-causing mutations to date have been shown to be located in the coding fraction, although this is likely to be because it is experimentally easier to find disease-causing mutations in the exome [90].

1.8.2 Whole genome sequencing

WES has the advantage of being less expensive and the analysis is less complex than whole genome sequencing (WGS). However the TAR discovery shows the importance of being able to interrogate the

non-coding space, and only a tiny fraction of this is covered by WES library capture. Therefore as costs reduced during the course of the project, the BRIDGE-BPD consortium switched to WGS, which currently costs approximately $1000 per genome. Advantages of WGS include significant coverage of regulatory regions. These regions are not well understood, and variants are more difficult to interpret, though significant recent advances have been made in the annotation of the regulatory regions of haematopoietic cells [91, 92]. WGS also offers improved coverage of the coding space because no enrichment of the exome is required, which can be inefficient if the area where probes hybridize contains a high number of SNVs [93]. WGS also offers improved recognition of copy number variations and other structural variants because these often arise in the non-coding space, because WGS read length is not limited by exon length, with longer reads allowing improved alignment to the genome, and because WGS read depth is more uniform, so that algorithms that use read depth to infer structural variants are more effective [94] .

1.9 Identification and confirmation of candidate genes

In the coming years high-throughput sequencing projects will discover novel genes in BPD. However sequencing alone will not provide answers; only by integrating big data from a wide variety of sources, followed by co-segregation combined with rigorous functional studies, can candidate genes be confirmed. In this section the sources of this data are discussed and a strategy for gene discovery employed in this study is discussed.

1.9.1 Identification of genes that carry novel, rare variants

After identifying a cluster of patients that share a rare phenotype, the next step is to obtain sequencing data on variants within their genomes. These can be prioritised depending on whether or not they occur in one of the 65 known BPD genes [66] but ultimately each variant has to be assessed for its level of pathogenicity. Knowing the frequency of a variant in the general population is an important part of this process and several databases have been developed recently to meet this need such as the Exome Aggregation Consortium (ExAC) [95], GnomAD [95], UK10K [96], 1000 Genomes [97], and UK Biobank [98].

Interpretation of individual variants in the context of population-based minor allele frequencies requires caution. Firstly, common variants such as the Factor V Leiden variant in *F5,* which modifies the risk of venous thrombosis but has a minor allele frequency of 0.012 (ExAC database, accessed July 2016; the frequency is likely to be higher in non-Caucasian populations), would typically be removed when reviewing high-throughput sequencing results. Secondly, samples from non-Caucasian populations are under-represented in existing control collections, limiting their value for the assessment of rare variants groups of patients of non-Caucasoid ethnicity. There is progress on this front from sequencing projects focusing on other ethnicities such as individuals from sub-Saharan Africa [99] and Pakistan [100]. By virtue of the multicultural population of the UK it is predicted that the 100 000 Genomes Project will

also help to rectify this issue of inequality in the variant catalogues. Thirdly, another limitation to the use of reference control cohorts of apparently healthy individuals occurs when the variant causing disease is variably penetrant and where the true prevalence of the disease is unknown. This is an important problem in the mild and moderate BPD where presentation and diagnosis depends upon haemostatic challenges. Consequently asymptomatic individuals who carry causative variants may be present in reference control collections. It is necessary to adapt the minor allele frequency threshold accordingly, to define the frequency below which a variant is 'rare enough' to consider further.

1.9.2 Isolation of variants that affect gene expression

In the search for variants in the coding fraction with consequences, synonymous variants are filtered out, leaving missense variants, insertion, deletions, frameshifts, variants affecting splicing and copy number and structural variations for further analysis. However it is important to note that apparently 'silent' variants have been shown to alter splicing or substrate specificity [101], therefore this category of variants may also merit further consideration.

Prioritising variants in the coding space for further assessment involves the application of specific knowledge of protein structure and function. Analysis of cross-species sequence conservation through evolution using comparative sequence alignment software is often used to identify critical residues. Various software algorithms for pathogenicity prediction have been developed but their Receiver Operating Characteristic curves are below par for their use in diagnostic tests. Combined Annotation Dependent Depletion (CADD) [102] scores appear more promising and integrate multiple datasets including measures of conservation, protein-structure impact and published expression levels in cell lines, as well as a number of regulatory annotations. These algorithms are inevitably only as good as the data used to train them, so will be less effective for less annotated genomic loci or less well understood protein domains.

A further validation tool is the assessment of candidate gene expression levels, specifically of transcripts containing the variant, in relevant cell types. Over the past decade catalogues of transcript level information for blood cells and their progenitors have been published. The more recent expression studies have used highly purified cells, including progenitors, and have analysed RNA by sequencing (RNA-seq), rather than by expression arrays that were previously used. Chen *et al.* [103] have published the results of transcriptome analysis by RNA-seq of HSCs and seven different types of progenitors including erythroblasts and megakaryocytes as part of the Blueprint epigenome of blood cells project. The RNA-seq data allows assessment of absolute and relative expression levels of transcripts harbouring candidate variants and to assess the extent of the use of alternatively spliced isoforms throughout haematopoiesis.

Of course variants in non-protein coding regions can also alter gene expression. The exome comprises just 2% (~64 Mb) of our 3.2 billion base pair genome. Most genetic discoveries to date have been in the

exonic fraction of our genomes because it has been accessible and interpretable in the context of our understanding of transcription, translation and protein structure-function relationships. It is increasingly clear that variants in regulatory regions also underpin BPD, such as 5'UTR variants in thrombocytopenia absent radius (TAR) syndrome [88] and thrombocytopenia secondary to *ANKRD26* variants [104].

Several projects such as Blueprint [103], Roadmap [105] and Encode [106] under the umbrella of the International Human Epigenome Consortium have made progress on the formidable task of assigning function to the non-coding space in hundreds of different types of cells. Approaches include mapping open chromatin regions and expression quantitative trait loci and using chromatin immunoprecipitation to identify histone modifications such as DNA methylation and binding sites for transcription factors [107]. The results of experiments applying these technologies are essential for the interpretation of variants in non-coding portion of the genome.

Genome-wide association studies (GWAS) for blood cell indices in large population samples provide a further resource for identifying regions of the genome harbouring genes implicated in haematopoiesis. Meta-analysis of GWAS has demonstrated 68 loci associated with platelet count and volume to date [108, 109]. Recently more powerful GWAS has been performed in 175,000 volunteers from the UK Biobank and the INTERVAL [110, 111] cohorts and has expanded this catalogue of regions containing SNVs associated with effects on platelet formation. The genomic regions signposted by GWAS as potential regulators of megakaryopoiesis and platelet formation are also candidate regions of harbouring variants for inherited and as yet unexplained platelet disorders.

1.10 Cellular models for functional validation of variants

A variety of haematopoietic cellular models have been employed to assess hypotheses resulting from sequencing discoveries. In this section the models used in this thesis are introduced.

1.10.1.1 Induced pluripotent stem cells

In 2006 Takahashi and Yamanaka discovered that mouse embryonic or adult fibroblasts could be reprogrammed back to a pluripotent state by transduction with transcription factors OCT4, SOX2, KLF4 and c-MYC [112] to generate induced pluripotent stem cells (iPSC). iPSC have since been widely used as an *in vivo* model of human stem cell behavior, in place of human embryonic stem cells (ESC).

Advantages of iPSC include their human origin, allowing the extrapolation of conclusions to human conditions. They provide potentially unlimited material and they do not suffer the ethical constraints of ESC or animal modelling [113]. They are easily accessible and can be genetically manipulated. Importantly for this study, iPSC can be differentiated to induced megakaryocytes (iMK), and thereafter to platelet-like particles (PLP) [114]. Large studies are underway to perform extensive characterisation of iPSC lines from hundreds of donors using standardised pipelines that include genotyping,

transcriptional, epigenetic and proteomic assays, providing detailed datasets for phenotypic comparison [115].

However several caveats must be considered when generalizing the results of iPSC studies to human disease. Several studies have shown that iPSC are similar but not identical to ESC, with the acquisition of genetic and epigenetic mutations during reprogramming and ongoing culture that may bias phenotypic studies [116-118]. Moreover, the local microenvironment, or niche, influences cell fate and cell behaviour *in vivo*, and these influences cannot be fully recapitulated in vitro [119]. Practical limitations include an inherent capacity for differentiation, meaning maintenance of undifferentiated cultures is more difficult than other mammalian cell lines. For example, antibiotic usage can stimulate differentiation, making cultures prone to loss through bacterial infection. Media and cytokines must be replenished daily. More gentle reagents are required for passage and manipulation and cultures are slower to grow than many other mammalian cell lines.

1.10.1.2 Megakaryocyte forward programming

Cells bearing the markers of mature MK have been generated by the directed differentiation of human stem cells including ESC and iPSC for some time. The sequential addition of cytokines stimulates haemato-endothelial mesoderm formation, mimicking embryogenesis, followed by differentiation to MK using stem cell factor (SCF), Fms-related tyrosine kinase 3 ligand (Flt3L) and thrombopoietin [120]. Alternative well-established protocols involve co-culture of stem cells with stromal OP9 cells [121], or $CD34^+$ cell sorting following embryoid body formation, albeit with low MK and platelet yield (around seven platelets per MK) [122]. Recently Moreau *et al.* [123] published a forward programming (FoP) protocol using lentiviral overexpression of key transcription factors GATA1, FLI1 and TAL1 that led to high yield, high purity, long-lived 'iMK' cultures. Cultures from one particular iPSC line remained viable and $CD41/CD42^+$ for over 90 days. One iPSC resulted in $2x10^5$ iMK, and $1x10^6$ PLP. The iMK yield was 26% higher and 70% purer than a directed differentiation protocol. Whilst this is certainly a breakthrough, MK in vivo generate 1000x as many platelets, and iMK are smaller, of lower ploidy and generate platelets that have a lower lifespan when transfused into mice when compared with $CD34^+$-derived MK. Therefore work is ongoing to enhance existing protocols, with a major focus on the three dimensional structure within which iMK are cultured, and using inducible systems to improve the reliability of reprogramming.

A further practical limitation of this protocol for the purposes of this project is that around 25% of iPSC lines will fail to reprogram to iMK. Among those that can there is high level variability in terms of iMK and yield, purity and lifespan although experience suggests the performance of an individual iPSC line is reproducible (*personal communication from Dr. Cedric Ghevaert*). This necessitates the careful selection of controls.

1.10.1.3 CRISPR Cas9 gene-editing

A key feature of the microbial adaptive immune system is the expression of clustered regularly interspaced short palindromic repeat (CRISPR) genes. CRISPR gene arrays encode Cas endonucleases that protect the host bacterium from the integration of viral DNA by recognizing target DNA and introducing a double stranded break (DSB) [124]. The Type II CRISPR system derived from *S.pyogenes* is a ribonucleoprotein complex consisting of the Cas9 endonuclease and a 20 nucleotide guide sequence (sgRNA) connected to a trans-activating CRISPR RNA (tracrRNA). The sgRNA directs the Cas9 to a complementary region on the target genome just upstream of a 5'NGG protospacer-adjacent motif (PAM) and binds the complementary sequence, 3' end first. The Cas9 creates a DSB between the 17th and 18th nucleotide of the complementary DNA sequence [125]. Following creation of a DSB the target may be repaired by error-prone non-homologous end-joining (NHEJ) which often leaves indels that can lead to frameshifts, premature stop codons and nonsense-mediated decay, knocking out the gene targeted. Alternatively, in the presence of a template DNA sequence, homology-directed repair (HDR) can mediate high fidelity repairs, though the efficiency of HDR is lower.

By altering the sequence of the sgRNA and expressing it alongside a CRISPR gene array, Cas9 can be targeted to any section of any genome with a PAM sequence, which occurs on average every 8 base pairs in mammalian DNA. As a result, permanent genome editing has become cheap, widely available and highly successful when compared with alternative genome engineering approaches including transcription activator-like effector nucleases (TALENs) and zinc finger nucleases [126]. It is clear that CRISPR Cas9 has revolutionized the field of molecular biology and is taking great strides towards having a similar impact on molecular medicine.

A key limitation of CRISPR Cas9 gene editing is the introduction of unwanted genome edits due to off target cleavage. Cas9 nickases (Cas9n) have been engineered to carry mutations in the catalytic domains (D10A in the RuvC domain or H840A in the HNH domain), that modify the endonuclease to make a single strand break. This can be exploited to reduce off target cleavage; two sgRNAs that each direct a nickase to make single strand breaks on opposite strands of target DNA require twice as much complementary sequence. This is particularly important in view of reports that mismatches between the sgRNA and target DNA may be tolerated, particularly towards the 5' end of the sgRNA which binds later [127]. Furthermore, the use of a Cas9n system with a repair template in the form of an oligonucleotide or plasmid encourages the use of the high fidelity HDR pathway to repair breaks. By introducing mutations in the template one can harness HDR to introduce small, precise edits to a genome [126] such as knock-in of a variant of interest.

1.11 Thesis objectives and summary

My objective is to improve the phenotypic and molecular characterisation of patients with disorders of platelet-dependent haemostasis, to improve the diagnosis of these conditions and to gain novel insights into MK and platelet biology.

The project is divided into four stages:

1. To recruit cases with platelet function disorders, congenital thrombocytopenias and unclassified bleeding disorders of unknown molecular aetiology from Royal Free NHS Trust
2. To investigate, design and use phenotyping tools to annotate the NIHR BR-RD BPD cohort, in order to facilitate computational variant calling
3. To use a range of statistical tools to interpret high throughput sequencing results to identify novel variants associated with BPD, and to confirm their validity by co-segregation within pedigrees
4. Functional characterisation of variants identified through empirical study of gene function *in vivo* and *in vitro*.

Chapter 2: Materials and Methods

2 Materials and Methods

2.1 Recipes

A list of suppliers is provided in Appendix 8.1.1.

AE6^{++} medium

- 500ml Dulbecco's modified eagle's medium (DMEM)/F12 (cat. 11330-032, Thermo Fisher Scientific)
- 3.6ml 7.5% sodium bicarbonate (cat. 25080094, Thermo Fisher Scientific)
- 5ml L-ascorbic acid 2-phosphate (cat. A8960, Sigma, Gillingham, UK)
- 10ml 50x insulin transferrin selenium (cat. 41400045, Thermo Fisher Scientific)
- 15ng/ml recombinant FGF2 (cat. 233-FB, R&D systems)
- 15ng/ml recombinant activin A (cat. 338-AC, R&D systems)

PBE flow buffer

Dulbecco's Phosphate-buffered saline (D-PBS, cat. D8537, Sigma)

0.5% bovine serum albumin (cat. A9576**,** Sigma)

2mM ethylene-diamine-tetra-acetic acid (CIMRCK)

HEK293T complete

500ml DMEM - high glucose (cat. D6429, Sigma)

55ml heat-inactivated fetal bovine serum (cat. F9665, Sigma)

5.5ml 10,000U/ml penicillin-streptomycin (cat. 15140-122, Thermo Fisher Scientific)

5.5ml minimal essential medium non-essential amino acids (cat. 11140-035, Thermo Fisher Scientific)

0.6ml 8mg/ml tylosin solution (cat. T3397, Sigma)

HEPES Buffered Saline (HBS)

NaCl, 150mM (cat. 10241AP, AnalaR BDH)

KCl, 5mM (cat. 101984L, AnalaR BDH)

$MgSO_4.7H_2O$ 1mM (cat. 101514Y, AnalaR BDH)

HEPES 10mM (cat. H4034, Sigma)

Made up to 1L with deionised water.

Formyl saline

5.4mL of 37% formaldehyde (cat. S22854-234, Sigma)

1L 0.85% NaCl (cat. S0817-1GA, Sigma)

Saponin-gelatin-PBS (SGPBS)

0.1% saponin (cat. S-7900, Sigma)

0.2% gelatin (cat. G-7745, Sigma)

0.02% azide (CIMRK)

NP-40 lysis buffer

0.5ml of 10mM Tris pH 8 (cat. T3038, Sigma)

1.5ml 150mM sodium chloride (cat. S5150, Sigma)

0.5ml 0.1% Tergitol (cat. NP40S, Sigma)

47.5ml Ultra-pure water (cat. TMS-011, Sigma)

1 cOmplete protease-inhibitor tablet (cat. S8820, Sigma)

2.2 Study name and ethical approval

Cases were recruited initially to the BPD arm of the BRIDGE study (UK REC10/H0304/66) then subsequently to the BPD arm of the NIHR BR-RD study (UK REC 13/EE/0325). The cases carrying *KDSR* variants (Chapter 4) were recruited to the BRIDGE study in Leuven, Belgium, and the study was approved by the Ethics Committee of the University Hospital Leuven, approval number ML3580. All study procedures were performed after the participants provided informed written consent and were in accordance with the Declaration of Helsinki.

2.3 Recruitment

2.3.1 Eligibility criteria

Eligibility criteria for the study were defined as follows:

Inclusion criteria:

1. A history of excessive bleeding with laboratory abnormalities consistent with a platelet function disorder (PFD) or
2. Diverse or severe lifelong bleeding symptoms or
3. A personal history of bleeding with a significant family history of bleeding or
4. Thrombocytopenia (platelet count <100x10^9), likely congenital or
5. Abnormal platelet morphology or size (MPV <6 or >12fl), likely congenital

Exclusion criteria:

1. Patient has a disorder of known molecular aetiology
2. Likely acquired aetiology
3. Patient under 16 years of age.
4. Patient not competent to consent

2.3.2 Recruitment and data collection

Patients with BPD registered at the Katherine Dormandy Haemostasis and Thrombosis Unit at the Royal Free London NHS Foundation Trust (RFH) were invited for a face-to-face interview. The patient record was reviewed to collect phenotypic information including historical bleeding scores where available, and

to ensure that outstanding clinical investigations pertinent to the eligibility criteria were completed and. The following clinical details were collected for each case recruited:

- Height, weight, gender, ethnicity,
- Bleeding history and severity (mild/severe)
- Condensed MCMDM-1 VWD bleeding score (see below)
- Past medical and surgical history
- Drug history (including retrospectively at time of platelet function testing that led to diagnosis)
- Family history

The consultant diagnosis, previous bleeding scores and blood results were reviewed following completion of the interview to avoid bias, particularly in terms of bleeding score calculations. The following laboratory tests were undertaken if not historically performed

- Full blood count
- Platelet morphology determined by light and/or electron microscopy
- Activated partial thromboplastin time, prothrombin time, Clauss fibrinogen assay
- Von Willebrand antigen, ristocetin cofactor, collagen binding and factor VIII coagulant activity assays
- PFA-100® (Siemens Diagnostics), Impact-R®(Matis Medical)
- Light transmission aggregometry (LTA) using a standard panel of agonists* [128] on the platelet aggregation profiler model PAP-8E® (Bio/Data Corporation).
- Bioluminescent assay of ATP release in response to agonists** on the Model 700 Whole Blood/Optical Lumi-Aggregometer® (Chrono-log Corporation) e.g. ADP if LTA suggested weak agonist defect
- Bioluminescent assay of nucleotide content in lysed platelets, read using the Junior LB 9509 Portable Tube Luminometer®(Berthold Technologies)
- 12ml of EDTA blood was sent to the Cambridge Translational Genomics laboratory (CATGO) at room temperature within 48 hours of venipuncture.
- Platelet morphology determined by light and/or electron microscopy
- Results of bone marrow analysis if available.

*Initial panel: ADP 2µM, 3µM; epinephrine 3µM, collagen 1µg/ml, ristocetin 0.5mg/ml, 1.5mg/ml, arachidonic acid 1.0mmol, U46619 1µM. Subsequently where appropriate: ADP 1µM, 5µM, 10µM; epinephrine 5µM, 10µM, collagen 2µg/ml, ristocetin 1.25mg/ml, 1.5mg/ml, arachidonic acid 0.5mmol/L]

** ADP 2µM, 3µM; collagen 1µg/ml, 2µg/ml

Where there were discrepancies between testing on different days, these tests were repeated to establish the diagnosis.

Dr. Veerle Labarque and Professor Kathleen Freson kindly provided all clinical data for the pedigree of IPD123.

2.3.3 Database design for data collection

A secure, online database was designed to collect the data specified in section 2.3.2. The majority of data was collected as quantitative or categorical variables, with a minority as free text to maintain detail. A minimum dataset for patient inclusion was defined and enforced through the database. Database images are available in Appendix 8.1.1.

I contributed to database design as part of a multidisciplinary team (MDT) with Dr. Claire Lentaigne, Clinical Research Fellow, Imperial College; Mr Tony Attwood, University of Cambridge, Professor Kathleen Freson, University of Leuven and Dr. Ernest Turro, Senior Analyst, University of Cambridge.

The MDT was responsible for the quality control of the dataset. Algorithms were designed to check numerical values entered were of the order of magnitude expected, that genetic gender estimates were compatible with the entered gender, that phenotypes attributed were gender-compatible (e.g. bleeding score > 0 on menstrual bleeding/ postpartum bleeding; gender-specific phenotypes were entered appropriately). The MDT approached individual recruitment centres for extra information to clean up the dataset, based on these algorithms.

2.3.4 Phenotyping

2.3.4.1 Application of the Human Phenotype Ontology

Most Human Phenotype Ontology (HPO) terms were assigned by searching by text or by escalation through the phenotype tree to find the most suitable term. Cases were also automatically assigned relevant HPO terms based on entries for bleeding symptoms or laboratory results. Manual confirmation of HPO terms was required to prevent inappropriate coding in case the reference range was not applicable to the case in question, e.g. if a child or pregnant.

2.3.4.2 Bleeding scores

2.3.4.2.1 Definition of PFD and UBD subgroups

For analysis, BPD were divided into platelet function disorders (PFD) and unclassified bleeding disorders (UBD). PFD was defined as a reproducible abnormality in the traces of at least one agonist on light transmission aggregometry (LTA), with the exception of epinephrine which was considered abnormal only if another agonist trace was also affected, in keeping with guidelines [128]. UBD was defined in the absence of evidence for PFD, with abnormal mucocutaneous bleeding that was severe, diverse or also present in family members, without a reproducible abnormality on LTA or bioluminescent assays of

nucleotide content or ATP release. Isolated abnormalities of Impact-R® or PFA-100® (Section 2.3.2) were therefore classified as UBD.

2.3.4.2.2 Recruitment for bleeding scores assessment

Cases were identified from the cohort of index cases in the NIHR BR-RD BPD study from four UK Haemophilia Centres: Royal Free London NHS Trust, Imperial College Healthcare NHS Trust, Addenbrooke's NHS Trust and University Hospitals Bristol NHS Trust. UK centres were chosen due to a high uptake of the BAT and consistency in application. Patients with UBD where the MCMDM-1 VWD bleeding score or any similar scoring system had been used as a diagnostic tool were also excluded. Bleeding scores were collected face-to-face at recruitment.

In a proportion of patients with PFD at RFH, the BAT had been used by a haemostasis expert not involved in collecting scores for the current study to calculate bleeding scores at the time of first patient referral for evaluation of excessive bleeding, before laboratory testing and diagnosis. In order to check the validity of the retrospective assessment of bleeding score used in the main study and to check inter-observer variability, the score collected for this study was compared to the historical scores.

2.3.4.2.3 Control recruitment

Controls comprised healthy volunteers and patients attending the local outpatient phlebotomy service for routine blood tests. Controls were eligible if they reported themselves to be in good health and had never been diagnosed with a bleeding disorder. Laboratory screening was not undertaken in controls. All patients and controls provided written, informed consent. The study was approved by the National Research Ethics Service.

2.3.4.2.4 The bleeding assessment tool

The BAT is based on the condensed MCMDM-1 VWD tool and is shown in Figure 5 Royal Free bleeding score, based on condensed MCMDM-1 VWD score [79]. The bleeding score was calculated using events up to the point of diagnosis to avoid bias resulting from interventions and to determine a score cut-off that could be applicable in a future prospective study. The range of the score differs by gender, from -3 (in the absence of bleeding symptoms despite several haemostatic challenges) to +45 (severe bleeding symptoms across all categories) in females and -2 to 37 in males. The same tool was applied to both patients and controls.

2.3.4.2.5 Statistical methods

Bleeding scores comprise ordinal values that are not normally distributed. Bleeding scores were compared using median values and the two-sample Wilcoxon rank-sum (Mann-Whitney U) test. Sensitivity, specificity, positive and negative predictive values were determined using score cut-offs that were previously shown to be of value in VWD [77]. The relationship between PFA-100/ Impact-R and bleeding scores was examined using a Spearman's rank correlation. Differences in PFA-100 and Impact-R results in study groups were assessed using the Student's t-test. Odds ratios and 95% confidence

intervals were calculated for a patient having a score of >0 in each symptom category of the bleeding score. The majority of symptom categories contained zero controls. However, the true population frequency of these symptoms amongst controls is greater than zero. We therefore added 0.1 to the number of controls in the following categories: cutaneous bleeding, oral cavity bleeding, post-tooth extraction bleeding, post-operative bleeding, intramuscular haemorrhage, haemarthrosis, CNS bleeding and gastrointestinal bleeding. In the absence of robust estimates this correction assumes that the true control population frequency for these six symptom categories is 0.2%. This assumption biases our results in favour of reduced odds ratios, particularly for rarer symptoms such as CNS bleeding and haemarthrosis. To examine phenotypic variation between patients with PFD and UBD, differences between the distribution of symptoms that contribute to the bleeding score were assessed using Fisher's exact test. Finally inter-observer variability was evaluated using a Wilcoxon signed rank test and a Bland-Altman plot. All bleeding score analyses were performed using the IBM SPSS® software package (version 20).

Score	-1	0	1	2	3	4
Epistaxis	-	No or trivial (<5 episodes per year)	>5 or more than 10 minutes duration	Consultation only	Packing or cauterization or antifibrinolytic	Blood transfusion or replacement therapy or desmopressin
Cutaneous	-	No or trivial (<1 cm)	>1 cm and no trauma	Consultation only		
Bleeding from minor wounds	-	No or trivial (<5 episodes per year)	>5 or more than 5 minutes duration	Consultation only	Surgical haemostasis or antifibrinolytic	Blood transfusion or replacement therapy or desmopressin
Oral cavity	-	No	Referred at least once	Consultation only	Surgical haemostasis or antifibrinolytic	Blood transfusion or replacement therapy or desmopressin
Tooth extraction	No bleeding in at least two extractions	None done or no bleeding in one extraction	Referred in <25% of all procedures	Referred in >25% of all procedures, no intervention	Resuturing or packing	Blood transfusion or replacement therapy or desmopressin
Surgery	No bleeding in at least two surgeries	None done or no bleeding in one surgery	Referred in <25% of all surgeries	Referred in ≥25% of all procedures, no intervention	Surgical haemostasis or antifibrinolytic	Blood transfusion or replacement therapy or desmopressin
Menorrhagia	-	No	Consultation only	Antifibrinolytic or pill use	Dilatation and currettage, iron therapy	Blood transfusion or replacement therapy or desmopressin or hysterectomy
Postpartum haemorrhage	No bleeding in at least two deliveries	No deliveries or no bleeding in one delivery	Consultation only	Dilatation and currettage, iron therapy, antifibrinolytic	Blood transfusion or replacement therapy or desmopressin	Hysterectomy
Muscle haematomas	-	Never	Post trauma no therapy	Spontaneous, no therapy	Spontaneous or traumatic, requiring desmopressin or replacement therapy	Spontaneous or traumatic, requiring surgical intervention or blood transfusion
Haemarthrosis	-	Never	Post trauma no therapy	Spontaneous, no therapy	Spontaneous or traumatic, requiring desmopressin/ replacement therapy	Spontaneous or traumatic, requiring surgical intervention or blood transfusion
CNS bleeding	-	Never	–	–	Subdural, any intervention	Intracerebral, any intervention
GI bleeding	-	No	Associated with GI pathology e.g. ulcer, varices, haemorrhoids	Spontaneous	Surgical haemostasis/ blood transfusion/ desmopressin/ replacement therapy/ antifibrinolytic	-

Figure 5 Royal Free bleeding score, based on condensed MCMDM-1 VWD score [79]

2.3.5 High throughput sequencing

2.3.5.1 DNA extraction

Following clinical interview 12ml of EDTA blood was sent to the Cambridge Translational Genomics laboratory (CATGO) at room temperature within 48 hours of venepuncture.

Genomic DNA was isolated from venous blood in the Cambridge Translational Genomics laboratory (CATGO) by Jennifer Jolley. In brief, 200ul blood was incubated in lysis buffer (ammonium chloride 7.7g, sodium bicarbonate 0.084g, water 1000ml) to lyse red cells and centrifuged (2500rpm for 15min) to yield a white cell pellet. The sample was resuspended in lysis buffer and repeatedly centrifuged until the pellet is no longer red. To extract DNA, the following were added: 3.5ml 6M guanidine hydrochloride, 260ul 7.5M ammonium acetate, 260ul 120% sodium sarcosyl, 50ul 10mg/ml Proteinase K. The tube was then vortexed to dissolve the pellet and incubated at 37^{o}C overnight. 2ml chloroform was added and the sample was vortexed until a white emulsion formed. The sample was centrifuged at 2500rpm for 5 min and the aqueous upper layer was collected and transferred to a tube containing 10ml 70% ethanol and incubated at -20^{o}C overnight. It was then rotated for 5 min to precipitate DNA and centrifuged (300rpm for 15 min) to pellet DNA. The ethanol was poured off and the ethanol wash was repeated twice. The sample was allowed to dry for 15 min, resuspended in 400ul of TE buffer on a rolling platform and refrigerated overnight. DNA concentration was assessed by Qubit (Life Technologies) and quality was assessed using gel electrophoresis and a Trinean spectrophotometer. A minimum of 100µl of DNA at 30ng/µl, with optical density between 1.75 and 2.04 was required to proceed to WES or WGS. DNA for co-segregation by polymerase chain reaction was extracted from saliva using the Origene Saliva kit protocol, following the manufacturer's protocol.

2.3.5.2 Sequencing pipelines

Whole exome sequencing (WES) library capture was performed following manufacturers protocols by the Welcome Trust Sanger Institute using the SeqCap EZ 64Mb Human Exome Library version 3.0 (Roche NimbleGen) and the TruSeq DNA LT Sample Prep kit (Illumina) on the Beckman Biomek FX automated workstation (Beckman Coulter). Whole genome sequencing (WGS) library capture was performed following manufacturers protocols by the Welcome Trust Sanger Institute using a TruSeq DNA PCR-free protocol. WGS was performed in 125 bp paired-end reads such that 95% of the GRCh37 reference genome had a read coverage of at least 15X. Sequencing was performed by Illumina (Chesterford, UK).

2.3.5.3 Sanger sequencing

All variants taken forward were confirmed by conventional Sanger sequencing by Mr. Jonathan Stephens of the CATGO Laboratory. Primers were designed to amplify a 300-500bp segment containing the chromosomal position of interest by polymerase chain reaction. The Sanger sequencing service was provided by Source Bioscience (Cambridge, UK). Primers used to confirm the presence of *KDSR* and *ABCC4* variants are shown in Table 8.

2.3.5.4 Quality control of sequencing data and data filtering

A large core analysis team led by Dr. Kathy Stirrups and Dr. Ernest Turro processed the sequencing data and performed quality control. Adapter sequence was removed. Estimates of yield, error rate in bases called, quality scores, flow cell cluster density, purity and synchronicity were used to ensure data quality. Additional quality control measures included checks of genetic gender, ethnicity and relatedness. Reads were aligned to the GRCh37 build of the human reference genome. Variants unlikely to be pathogenic and technical artefacts were excluded from further analysis if they did not co-segregate in other recruited affected family members or if there were <3 reads supporting the variant allele. Analyses also excluded variants with a frequency >0.1% in the GnomAD database [95] of 61,568 individuals' exome sequences and 7748 individuals' whole genome sequences and in the in-house NIHR BR-RD collection of 9472 individuals across 15 rare disorder categories at minimum 30X coverage.

2.3.5.5 Variant calling

In order to facilitate variant calling a searchable database was created by statistical genetics team at the University of Cambridge, led by Dr. Ernest Turro and Mr Daniel Greene. The database consolidated information about variant impact and frequency, phenotypes and pedigree information, the aforementioned CADD pathogenicity scores and expression in haematopoietic lineages. I was responsible, with the phenotyping MDT, for the accuracy of phenotype information, for searching for novel genes on a case-by-case basis, by gene name, or by phenotype, and for clinical evaluation of candidate genes identified by novel statistical approaches [129].

This evaluation of multiple candidate genes comprised a large part of the study. Following clinical evaluation of phenotypes, existing biological information about genes was used to annotate and prioritize variants:

- Interrogation of genome-wide association studies of platelet parameters [108]
- Assessment of conservation at variant loci to establish phylogenetic importance
- Manual literature searching to identify whether gene is known to be associated with platelet number, structure or function, and if so, estimation of impact of variant on protein function
- Review of crystal structure of protein, if known, to determine likely impact of variant [130]
- Comparison with mouse knockout phenotypes [131]
- Assessment of connectivity with platelet-associated genes in protein-protein interaction networks [132]

2.3.6 Co-segregation studies

Potential causative variants were initially investigated by inviting affected and unaffected members of the family for interview and blood sampling where required, to ensure the variant co-segregated with the phenotype in question. Co-segregation was performed on DNA extracted from blood or saliva (details of extraction protocol are provided in Section 2.3.5.1).

2.4 Cell culture and manipulation

2.4.1 Induced pluripotent stem cells (iPSC)

2.4.1.1 Origin of cell lines

Stock iPSC lines (A1ATD-1, QOLG, PODX, BIMA) were provided as a gift by Dr. Cedric Ghevaert, University of Cambridge.

IPD123, 15G and 14I patient-derived iPSC lines were prepared by the laboratory of Professor Ludovic Vallier of the Cambridge Biomedical Research Centre (BRC) from fibroblasts isolated from a 2mm skin biopsy of the proband' upper arm. Fibroblasts were reprogrammed following the protocol of Yamanaka and colleagues [112, 133] by overexpression of transcription factors OCT4, SOX2, KLF4 and c-MYC using the CytoTune™-iPS 2.0 Sendai Reprogramming Kit (cat no A16517, Invitrogen) followed by culture on mouse embryonic fibroblast feeder cells in chemically defined media. Following establishment of iPSC colonies approximately one month after reprogramming clones 1, 2 and 4 were picked and passaged to feeder-free conditions. Clones were validated and cryopreserved prior to acquisition for this study.

2.4.1.2 Cell maintenance

S4 iPSC were cultured feeder-free in $AE6^{++}$ media (see 'Recipes'). Problems with differentiation and recurrent infection led to a change of medium for subsequent culture. IPD123 patient-derived lines, A1ATD-1 and derived lines, QOLG, PODX and BIMA stock lines were cultured feeder-free in Essential 8 medium (E8, cat A1517001, Thermo Fisher Scientific). All media were incubated at 37^{o}C prior to contact with iPSC.

Medium did not contain antibiotics. A variety of sizes of Nunc tissue culture plates were used, according to need (Thermo Fisher Scientific). Plates were coated with 10µl/ml vitronectin VTN-N (cat. A14700, Thermo Fisher Scientific) in D-PBS (cat. D8537, Sigma). Medium was changed approximately every 24 hours. Strict sterile technique was observed and plates were only opened inside a sterile laminar flow cabinet in a dedicated stem cell facility. Cells were incubated between manipulations at 37^{o}C, 5% CO_2.

2.4.1.3 Passage

Cells required passage approximately every 3-5 days, depending on seeding density, usually when cells reached ~80% confluence. Adherent cells were washed once with 2ml D-PBS. Standard passage for cell maintenance involved incubating cells in 1ml of 50µM ethylene-diamine-tetra-acetic acid (EDTA, CIMRCK) in D-PBS (PBS-EDTA) for 5 minutes. 1ml D-PBS was added to the well and pipetted over the dish to gently lift small cell clumps into suspension. This was added to 5ml D-PBS and centrifuged (200g, 5 minutes). The pellet was resuspended gently in 1ml E8 and split between 1:5 and 1:20 depending on the starting cell density and the requirements of upcoming experiments.

For certain experiments, e.g. MK reprogramming, nucleofection, cells were split to single cells using TrypLE (cat. 12563029, Thermo Fisher Scientific). iPSC culture medium was removed and cells were

washed with 1ml D-PBS. 1ml room temperature TrypLE was added for 3 minutes (37^{o}C) then aspirated without disturbing cells. Cells were collected in 10ml basal E8 media and centrifuged (300g/ 5 minutes). The pellet was resuspended in 1ml media containing 10µM ROCK inhibitor and seeded at the required density in vitronectin-coated tissue culture plates.

2.4.1.4 Management of differentiation and picking of colonies

Undifferentiated iPSC form distinct colonies of homogenous, small cells with a high nuclear-to-cytoplasmic ratio. Prior to passage differentiated colonies were removed by marking the plate under the microscope and scraping in a sterile laminar flow cabinet. If differentiation was marked but considered salvageable, then passage was performed using Dispase and Collagenase IV (cat. COLLDISP-RO, Sigma), diluted in D-PBS to 1mg/ml each. Cells were incubated for 5-30 minutes at 37^{o}C, checking for lifting of whole, undifferentiated colonies every 5 minutes. Once lifted colonies were gently collected and washed twice in 10ml D-PBS prior to reseeding in fresh medium.

Individual colonies were picked by marking the plate under the microscope then scraping and aspirating concurrently inside the laminar flow cabinet.

2.4.1.5 Thawing cells

To initiate iPSC culture from frozen stock, frozen vials containing approximately 1x10^{6} cells were thawed in a water bath at 37^{o}C until a small piece of ice remained. They were sprayed with 70% ethanol, transferred to a 15ml falcon and 10ml of E8 basal medium was added drop wise. The sample was centrifuged (200g, 3 minutes). The pellet was resuspended gently in 1ml E8 containing 10µM final concentration of Y-27632 (ROCK inhibitor, cat. Y0503, Sigma) and 5x10^{5} cells were seeded per well in a 6 well plate. The contents of each well were topped up to 1.5ml with E8 medium. 24 hours later the medium was replaced with E8.

2.4.1.6 Cryopreservation

To cryopreserve cells, cells were lifted and washed using PBS-EDTA as described above, and resuspended in freezing media consisting of 0.1ml DMSO (cat. D8418, Sigma) and 0.9ml knockout serum replacement medium (cat. 10828028, Thermo Fisher Scientific). The cell mix was transferred to a cryovial and placed in a Mr Frosty freezing container containing isopropyl alcohol (cat. I9516, Sigma). This was placed in a -80^{o}C freezer for up to six months or -150^{o}C freezer for longer term storage. These steps were undertaken quickly because DMSO is toxic to cells at room temperature.

2.4.1.7 TUNEL assay

Apoptosis was induced in iPSC lines at ~50% confluence by incubating cells for 12 hours in 0-500µM hydrogen peroxide (H_2O_2, cat. 88597, Sigma), diluted in E8 medium. H_2O_2 dilutions were made fresh on each occasion and the stock solution was stored in darkness. Cells were collected and processed using the APO-BrdU TUNEL Assay kit (cat. A23210, Thermo Fisher Scientific) following the manufacturer's instructions. For the majority of experiments the anti-BrdU mouse monoclonal antibody PRB-1, Alexa

Fluor 488 conjugate in the kit was used. For experiments involving an eGFP tagged line an anti-BrdU mouse monoclonal antibody (MoBU-1), Alexa Fluor 647 conjugate (cat. B35140, Thermo Fisher Scientific) was used at a 1:100 dilution, to avoid overlapping spectra.

2.4.1.8 Growth curve

Cells were passaged to single cells with TryPLE. $1x10^4$ cells were plated in each well of a vitronectin-coated 6-well plate. One well of cells was collected as single cells every 24 hours, washed, stained with trypan blue and viable cells were counted using a haemocytometer.

2.4.1.9 Immunostaining

iPSC were passaged using PBS-EDTA and plated in E8 medium on sterile glass coverslips coated in 10µl/ml vitronectin in D-PBS. At 60-80% confluence cells were fixed in 1% paraformaldehyde (10µl/ml) in D-PBS for ten minutes at RT. Cells were gently washed three times with D-PBS and the plate was kept at 4^oC until staining.

D-PBS was aspirated and cells were quenched with 50µM ammonium chloride (CIMRK) for 5 minutes at RT. This was aspirated and cells were permeabilized with saponin-gelatin-PBS (SGPBS, see 'Recipes') for 5 minutes, three times. Cells were incubated with primary antibody, followed by 3x 5 minute washes with SGPBS. Then cells were incubated with 1µg/ml DAPI and secondary antibody followed by 3x 5 minute washes with SGPBS. Cells were mounted with Hydromount mounting medium (cat. HS106100ML, Thermo Fisher Scientific) overnight in the dark. Cells were examined within 48 hours using a Leica DMi8 fluorescent microscope or the Leica Sp5 confocal microscope. Images captured were analysed using FIJI software [134].

Antibodies were all applied for 1 hour at RT, and secondary antibodies were applied in low light and incubated in the dark. Secondary-only stains and where relevant, un-transduced control stains were performed as controls. Antibody combinations were as follows:

- Rabbit anti-human DYKDDDDK (FLAG) (1:200, cat. 2368S, Cell Signalling); donkey anti-rabbit Alexa Fluor 555 (1:1000, cat. ab150074, Abcam)
- Mouse anti-human calnexin (1:250, cat. ab31290, Abcam); goat anti-mouse Alexa Fluor 633 (1:1000, cat. A21050, Thermo Fisher Scientific)
- Chicken anti-human calreticulin (1:1000, cat. ab2908, Abcam); goat anti-chicken Alexa Fluor 647 (1:2000, A21121, Thermo Fisher Scientific)
- Mouse anti-human myc '9E10' (1:250, cat. MA1-980, Thermo Fisher Scientific); goat anti-mouse Alexa Fluor 488 (1:1000, cat. A21121, Thermo Fisher Scientific)

2.4.2 Induced megakaryocytes (iMK)

2.4.2.1 Forward programming (FoP) to generate iMK

2.4.2.1.1 Overview

Moreau *et al.* published the base protocol used in this project during the course of the project [123]. Alterations have been made by the Ghevaert lab to the published protocol to improve yield, chiefly the single cell seeding of iPSC, culture in standard tissue culture flasks rather than an Aggrewell dish, and the removal of Ly-294002 (PI3 kinase inhibitor) from the protocol. The protocol used in this study is as follows.

2.4.2.1.2 Reagents

- 1X TrypLE Select (cat. 12563029, Thermo Fisher Scientific)
- E8 media (cat A1517001, Thermo Fisher Scientific)
- Vitronectin VTN-N (cat. A14700, Thermo Fisher Scientific)
- Nunc 6- and 12-well uncoated tissue culture plates (Thermo Fisher Scientific)
- Recombinant FGF2 (cat. 233-FB, R&D Systems)
- BMP4 (cat. 314-BP-010, R&D Systems)
- Rock inhibitor Y-27632 (cat. Y0503, Sigma)
- rhTPO (cat. 01417-050, CellGenix)
- rhSCF (cat. PHC2116, Thermo Fisher Scientific)
- Protamine Sulphate (cat. P4505, Sigma)
- D-PBS (cat. D8537, Sigma)
- TrypLE (cat. 12563029, Thermo Fisher Scientific)
- CellGro-SCGM (cat. 0020802-0500, CellGenix,)
- Viral Vectors (Vectalys), see Figure 6 Vectalys custom vector maps for forward programming to iMK

  pTRIPU3-TAL1 (batch pV.2.3.107_p16_01_1, titre $5.5x10^9$TU/ml)

  pWPT-hFLI1 (batch pV.2.3.993_p15_11_2, titre $1.5x10^9$TU/ml)

  pWPT-hGATA (batch pV.2.3.1073_p15_10_1, titre $2.6x10^9$TU/ml)
- Standard iMK flow cytometry:
  - FITC mouse anti-human CD235a (cat. 559943, BD Biosciences)
  - APC mouse anti-human CD41a (cat. 559777, BD Biosciences
  - PE mouse anti-human CD42b (cat. 555473, BD Biosciences)
- GFP^+ iMK flow cytometry
  - APC-H7 mouse anti-human CD41a (cat. 561422, BD Biosciences)
  - PE mouse anti-human CD235a (cat. 555570, BD Biosciences)
  - APC mouse anti-human CD42b (cat. 551601, BD Biosciences)
- Isotype controls:
  - PE mouse $IgG_1\kappa$ (cat. 555749, BD Biosciences)
  - APC mouse $IgG_1\kappa$ (cat. 555751, BD Biosciences)
  - FITC mouse $IgG_2\beta$ (cat. 555742, BD Biosciences)

- Flow-Count™ Fluorospheres (cat. 7547053, Beckman Coulter)
- PBE flow buffer (see ‘Recipes’)
- Paraformaldehyde (cat. 30525-89-4, Santa Cruz Biotechnology)
- DAPI in tissue culture-grade water (cat. D9542, Sigma)**.**

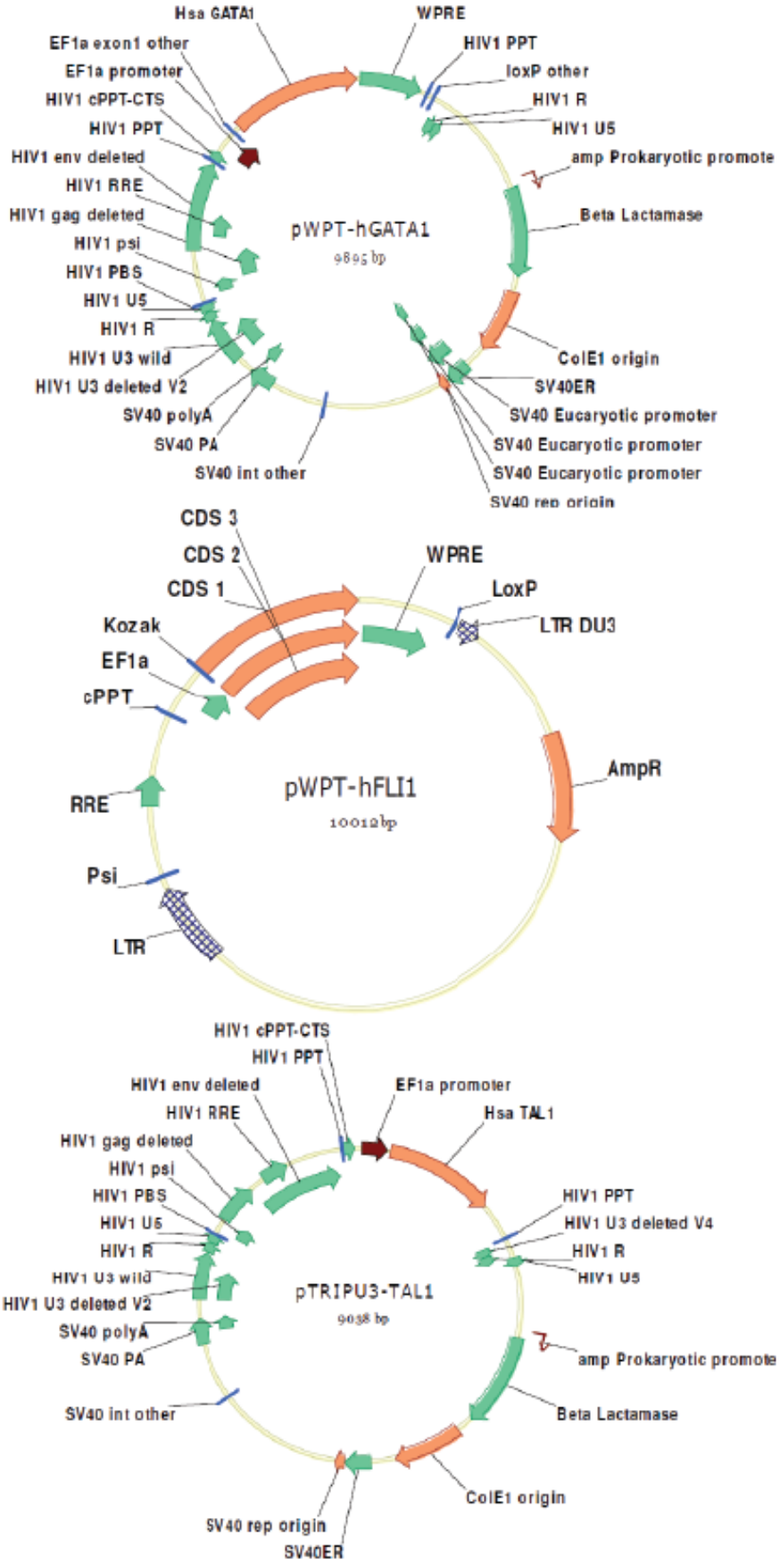


Figure 6 Vectalys custom vector maps for forward programming to iMK

2.4.2.1.3 Protocol

All media was warmed to at 37°C prior to contact with cells. All incubations were at 37°C, 5% CO_2. Strict sterile technique was observed and cells were transduced and manipulated up to day 10 in a dedicated CL2 environment, in a laminar flow cabinet.

Day -1

iPSC were passaged using TrypLE as described above and the pellet was resuspended in 1ml E8 media containing 10µM ROCK inhibitor. Cells were counted using a haemocytometer and seeded at $4x10^5$ cells per well in vitronectin-coated 12 well tissue culture plates. Medium was topped up to 1ml per well with E8 supplemented with 10µM ROCK inhibitor. Cells were cross-mixed and incubated overnight.

Day 0

Medium was aspirated and cells were washed with 1ml D-PBS. Fresh transduction medium was added, containing E8 basal medium (no supplement) with FGF2 20ng/ml, BMP4 10ng/ml and protamine sulphate 10µg/ml. A viral mastermix was added to transduce the cells at MOI 20 for each vector, and gently swirled to mix. Cells were incubated overnight.

Day +1

Medium was aspirated and cells washed with 1ml of D-PBS. 500µl of fresh transduction medium was added, omitting only protamine sulphate. Cells were incubated overnight.

Day +2

Medium was aspirated and cells washed with 1ml of D-PBS. 500µl MK medium (Cellgro, TPO 20ng/ml, SCF 25ng/ml) was added. Cells were incubated until day 5.

Day +5

500µl 2X MK medium (Cellgro, TPO 40ng/ml, SCF 50ng/ml) was added to obtain 1ml final volume in each well. Cells were incubated until day 5.

Day +8

400µl of media was removed from each well, leaving the cellular layer undisturbed. 500µl 2XMK media was added to each well.

Day +10

Supernatant was collected form each well. Cells were collected by passage with 300µl TrypLE (37°C, 10 minutes) and combined with the supernatant. The mix was quenched with 10ml room temperature (RT) D-PBS and centrifuged (300g, 5 minutes). The pellet was resuspended in 200µl MK medium and cells were counted with a haemocytometer. 40µl* was removed for flow cytometry (see below) and the remainder was added to a single well of an uncoated 6 well tissue culture plate and topped up with MK medium to 1.5ml per well. Cells were maintained in suspension at $<2x10^6$/ml thereafter.

*20µl for multi-colour staining, 5µl for unstained mix and 5µl for each single stained mix.

Day +13/16/19

0.8ml of media was carefully removed to minimize cell disturbance and 1ml of 2X MK medium was added per well.

Day +20

The contents of the well were collected and centrifuged (120g/ 8 minutes/ acceleration 3, brake 3). The pellet was gently resuspended in 2ml MK medium and cells were counted with a haemocytometer. 70µl* was removed for flow cytometry. The remainder was returned to an uncoated 6-well tissue culture plate, adjusting cell density to $0.5x10^6$/ml.

*50µl for multi-colour staining, 5µl for unstained mix and 5µl for each single stained mix. Cells were harvested for use at or just after day 20 so no long-term culture was undertaken.

Flow cytometry analysis at day 10 and day 20

Cells were routinely directly stained with a mastermix of CD235a-FITC (1:1000), CD41a-APC (1:10) and CD42b-PE (1:1000) for 20 minutes in the dark at RT. The reaction was washed and collected with PBE flow buffer (see 'Recipes') supplemented with 0.5% paraformaldehyde, 1µg/ml DAPI to gate viable cells and, at day 20, 5,000 Flow Count fluorosphere counting beads (Beckman Coulter). An unstained mix and isotype controls (specify) were used to visualise cells and set gates. Single stained mixes were used to perform compensation. Flow cytometry was performed on the Beckman Coulter Gallios Cytometer and analysed using Kaluza Analysis v.1.5a (Beckman Coulter).

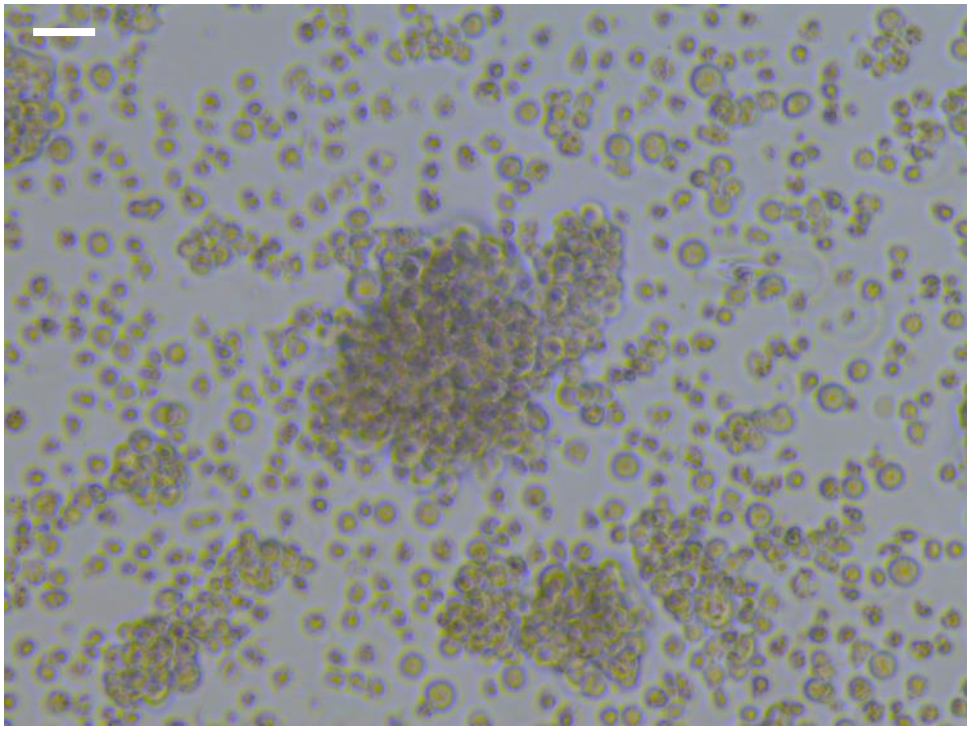


Figure 7 iPSC were forward programmed to iMK. Large grape-like clusters of bright, spherical cells in suspension at day 20 was indicative of successful reprogramming and was confirmed by dual expression of CD41 and CD42, and lack of CD235 expression or DAPI uptake . Scale bar indicates 50µm

Replicates

Each iPSC line was forward programmed in three independent, parallel experiments. iPSC lines of the same passage number were split into separate wells and cultured separately for 5-10 days prior to commencement of FoP.

2.4.2.2 Proplatelet assay

Sterile glass coverslips were coated in 200µg/ml fibrinogen (cat. F3879, Sigma) in D-PBS overnight at 4^{o}C then washed with D-PBS. iMK were diluted to a concentration of 0.5x10^{6}/ml in Cellgro medium with no cytokines. 300µl of cell suspension was seeded per well of a 24-well plate. Cells were examined regularly and when proplatelet formation was seen they were fixed in 1% paraformaldehyde in D-PBS for ten minutes at RT. Cells were gently washed three times with D-PBS and the plate was kept at 4^{o}C until staining.

D-PBS was aspirated and cells were quenched with 50µM ammonium chloride (CIMRK) for 5 minutes at RT. This was aspirated and cells were permeabilized with saponin-gelatin-PBS (SGPBS) for 5 minutes, three times and incubated with mouse anti-human α-tubulin antibody (1:250, 1hour, RT, cat.T5168, Sigma) followed by 3x 5 minute washes with SGPBS. Then cells were incubated with 1µg/ml DAPI and goat anti-mouse Alexa Fluor 488 antibody (1:1000, 1hour, RT, dark, cat. A21121) followed by 3x 5 minute washes with SGPBS. Cells were mounted with Hydromount mounting medium (cat. HS106100ML, Thermo Fisher Scientific) overnight in the dark. Cells were examined within 48 hours using a Leica DMi8 fluorescent microscope. Images captured were analysed using FIJI software [134]. Proplatelets were

defined as a clear projection from the main cell body with approximately parallel sides to the shaft (conical suggests simply spread on fibrinogen) that are longer than they are wide. Distinct branches of a single origin at the cell body were counted individually.

$CD41^+$ immunostaining of iMK was undertaken with mouse anti-human CD41 antibody (1:100, 1 hour, RT, cat. CD41UL-100, Alpha Diagnostic) followed again by goat anti-mouse Alexa Fluor 488 antibody.

2.4.2.3 cAMP ELISA

cAMP ELISA was performed using the Amersham cAMP Biotrak EIA System (cat. RPN225, GE Healthcare) using the acetylation procedure in accordance with the manufacturer's instructions. iMK pellets containing $1x10^5$ iMK were resuspended in lysis buffer 1B containing 400µM IBMX (cat. I5879, Sigma) and incubated for ten minutes (RT). These were kept at -20°C until analysis. Acetylation was performed in a fume hood. During the assay samples were incubated on a BioshakeIQ microplate shaker and end-point determinations was performed using a Tecan Infinite F200 Pro plate reader. Each sample was tested in triplicate on each ELISA plate. iMK from each of the three replicate FoP experiments were tested on three separate plates.

2.4.2.4 Cytospin

$25x10^5$ iMK were plated on a cytospin cassettes using a Shandon Cytospin 4 machine by Ms. Samantha Farrow and stained with a Romanowsky stain and coverslips were applied.

2.4.3 HEK293T (ATCC CRL-11268)

2.4.3.1 Cell maintenance

Cells were maintained in uncoated $10cm^2$ tissue culture plates in 10ml HEK293T Complete medium (see 'Recipes'). Medium was warmed to 37°C prior to contact with cells and was changed three times per week. Cells were incubated between manipulations at 37°C, 5% CO_2.

2.4.3.2 Cell passage

When cells were >50% confluent, they were gently washed with RT D-PBS to remove serum then incubated in 2ml trypsin for 3-5 minutes at 37°C, until cells were seen to be lifting. HEK293T Complete medium was added to the dish to neutralise the trypsin and collect the cells. Cells were homogenised, counted with a haemocytometer and reseeded in uncoated dishes at the required density.

2.4.3.3 Transfection for colocalisation studies

$4x10^6$ HEK293T cells were plated in a $10cm^2$ dish in HEK293T complete medium (see 'Recipes') and allowed to attach for 4 hours. 10µg of endoplasmic-reticulum marker plasmid mCh-Sec61β carrying the mCherry fluorescent tag (Addgene, #49155) and 10µg of eGFP-tagged KDSR expression plasmid pLenti-EF1a-KDSR-mGFP (cat. CW302081, Origene, Figure 15) were combined with 1ml room temperature DMEM (cat. D6429, Sigma) and 75µl 1mg/ml PEI (cat. 408727, Sigma) and incubated at RT for 20 minutes.

The mixture was spread drop wise over the HEK293T cells. Cells were incubated at 37°C, 5%CO_2. 24 hours post-transfection the medium was refreshed.

Figure 8 mCh-Sec61 beta plasmid map, used for colocalisation studies in HEK293T cells

2.4.4 Platelet studies

2.4.4.1 Blood collection

Whole blood was drawn from the antecubital fossa via a 21-gauge butterfly needle and vacutainer tubes (Becton Dickinson). The first sample was taken into an EDTA tube and used to measure the full blood count on a Sysmex XN1000. The cuff was released to collect samples for platelet assays into 3.2% trisodium citrate tubes, to minimise platelet activation. Blood was processed within 15 minutes of venesection.

2.4.4.2 Whole blood flow cytometry for platelet function

Whole blood platelet function by flow cytometry was performed by Ms. Joana Batista and Mrs. Harriet McKinney of the Cell Phenotyping team in the Ouwehand laboratory at the University of Cambridge, as described in the protocol by Jones *et al.* [135]. In brief, 5µL of citrated whole blood were added to 50µL of HBS (see 'Recipes') containing 100µM aspirin (cat. A5376, Sigma) to inhibit the thromboxane A_2 pathway, 10U/ml hirudin (cat. 94581, Sigma) to inhibit thrombin and, for CRP-XL stimulated samples, 4U/ml apyrase (cat. A6535, Sigma) to inhibit the ADP pathway.

Samples were incubated (20 minutes, RT) with either ADP (cat. A2754 Sigma) or CRP-XL (gift from Dr. Richard Farndale, University of Cambridge), and either FITC–anti-fibrinogen (1:50, cat. F0111, Agilent),

PE–anti-P-selectin (1:2500, clone Throm6 IBGRL) or a PE-IgG1 isotype control (cat. 9E10, IBGRL) in the dark for 20 minutes at RT.

The samples were added 1:100 to formyl saline (see 'Recipes') to stop the reaction. Negative controls for the P-selectin antibodies were set using the isotype control and for the anti-fibrinogen using samples incubated with the antibody in the presence of 6 mm EDTA (cat. E7889, Sigma), that reduces binding of fibrinogen to GPIIb/IIIa. AbC Total Compensation Capture beads were used for compensation (cat. A10497, Thermo Fisher Scientific). Flow cytometry was performed on the FC-500 (Beckman-Coulter).

2.4.4.3 Washing and pelleting platelets

Blood in 3.2% trisodium citrate was centrifuged (150g, 15 minutes, no brake, RT) to obtain platelet-rich plasma (PRP). Two volumes ACD (cat**.** C3821, Sigma)**,** 1µM final concentration PGE1 (cat. P5515, Sigma) and 0.2U/ml apyrase (cat/ A6535, Sigma) were added and gently inverted to mix. PRP was centrifuged (100g, 15 minutes, no brake, RT) to pellet contaminating red and white cells. The PRP platelet count of the supernatant was measured by Sysmex XN1000 and centrifuged again (800g, 15 minutes, brake 1, RT) to pellet platelets.

2.4.4.4 Flow cytometry for GPVI expression

Flow cytometry for platelet GPVI expression was performed by Ms. Joana Batista of the Cell Phenotyping team of the University of Cambridge. PRP was diluted in PBE buffer (see 'Recipes') to 5 x 10^9/L and 100µl of this was incubated with anti-GPVI antibody (20µg/ml, cat. HY101, IBGRL) for 20 minutes. Platelets were then washed in 4ml PBE buffer (1946G for 4 minutes). For monomeric GPVI staining, samples were incubated with FITC Fab2 secondary antibody, (50µg/ml, cat. 115-096-072, Jackson Lab) for 20 minutes in the dark. For dimeric GPVI binding, samples were incubated with Fab N74A antibody (gift from Dr. Stephanie Jung, University of Cambridge). Samples were washed again and resuspended in PBE prior to flow cytometry on the FC-500 (Beckman Coulter).

2.4.4.5 cAMP ELISA

cAMP ELISA was performed on platelets as detailed in section 2.4.2.3 for iMK, using the Amersham cAMP Biotrak EIA System.

2.4.4.5.1 Unstimulated cAMP measurements

Platelet pellets were resuspended at a concentration of 1 x 10^5 platelets/µl in lysis buffer 1B containing 400µM IBMX (cat. I5879, Sigma) and incubated for ten minutes (RT). These were kept at -20^{o}C until analysis. The platelet concentration for ELISA was determined by titration using control platelets obtained from apparently healthy volunteers. 100µl of thawed sample was used per test, equivalent to 1x10^7 platelets. Each sample was assessed in triplicate on each ELISA plate.

2.4.4.5.2 Iloprost stimulation studies

Assistance during sample collection and reagents were provided by Ms. Chantal Thys of the University of Leuven. PRP was diluted in platelet-poor plasma to $2.5x10^5$ plt/µl and IBMX was added to a final concentration of 400µM. 20µl of PRP/IBMX mix was added to 160µl D-PBS in a 37°C waterbath with stirring. Iloprost (cat. 2038, Tocris) diluted in D-PBS was added to a final concentration of 1.25ng/ml. The reaction was stopped by adding 12% trichloroacetic acid (cat T6399, Sigma) at 10, 30, 60, 180 and 300 seconds, which precipitates the platelet proteins and leaves cAMP in the supernatant. The sample was centrifuged (7000g, 10 min, RT, for 10minutes) and 500µl 100% diethylether (cat. 344362, Sigma) was added to neutralise and remove trichloroacetic acid. Samples were mixed and stored at -20°C until analysis. For analysis the upper layer was poured off and the sample was lyophilised in a Speedvac concentrator (Savant) for 2 hours at 30°C. Samples were used undiluted in the cAMP ELISA as described for unstimulated cAMP measurements.

2.4.4.5.3 ADP stimulation following Iloprost administration

Samples were incubated with Iloprost for 60 seconds as described then 10µl ADP was added at 2, 5 and 10µM final concentrations. The reaction was stopped and subsequently processed as described for Iloprost stimulation studies.

2.4.4.6 Light transmission aggregometry

Assistance during sample collection and reagents were provided by Ms. Chantal Thys of the University of Leuven. PRP was diluted in platelet-poor plasma to $2.5x10^5$ plt/µl and incubated with horm collagen (cat. 10500, Takeda) at a final concentration of 0.25, 0.5, 1 and 2µM. Samples were analysed on the platelet aggregation profiler model PAP-8E® (Bio/Data Corporation) using platelet-poor plasma to set the baseline and following aggregations for 6 minutes.

For Iloprost-induced inhibition of aggregation, PRP was pre-incubated with D-PBS as a control or Iloprost at 0.625, 1.25 and 2.5ng/ml for 1 minute followed by 2µg/ml horm collagen before analysis on the PAP-8E® as described.

2.4.4.7 Transmission electron microscopy

Samples for transmission electron microscopy from the *ABCC4* pedigree were processed with the assistance of Ms. Harriet McKinney of the Cell Phenotyping team at the University of Cambridge. PRP was added to an equal volume of 0.4% glutaraldehyde (made in 0.05 M sodium cacodylate pH 7.4, both reagents provided by the Cambridge Advanced Imaging Centre (CAIC)), and incubated at RT for 4 hours. The sample was made up to 10ml with 0.05 M sodium cacodylate and centrifuged (1500g, 10 minutes, RT). The pellet was washed twice in 1ml 0.05M sodium cacodylate (2000g, 5 minutes, RT) It was resuspended in 1ml 0.05M sodium cacodylate and stored at 4°C until analysis, 3 days later.

The samples were fixed, embedded in resin and sectioned on a diamond knife by Dr. Jeremy Skeppers of the CAIC, who also provided assistance in image acquisition. Processing for samples from the *KDSR*

pedigree was performed in the laboratory of Professor Kathleen Freson, KU Leuven, who also assisted in the interpretation of all images.

2.4.4.8 Leucodepletion for platelet proteomics and trizol suspension

PRP was leucodepleted by Ms. Frances Burden of the University of Cambridge. PRP was centrifuged (150g, 10mins, acceleration 9, brake 4). The supernatant was retained and 1ml 0.1M EDTA/sodium chloride was added, independent of the PRP volume. The sample was centrifuged again (150g, 10mins, acceleration 9, brake 4). The resulting supernatant was further leucodepleted with CD45 magnetic beads following the manufacturer's protocol. The volume of leucodepleted PRP that was obtained from 10ml of whole blood was snap frozen for proteomics and stored at -80^{o}C until analysis.

The remainder was centrifuged (1500g, 10 minutes, acceleration 9, brake 9). The resulting pellet was resuspended vigorously in 500µl Trizol (cat. 15596026, Thermo Fisher Scientific), snap frozen and stored at -80^{o}C until RNA extraction.

2.4.4.9 Platelet proteomics

Platelet pellets obtained were sent frozen to the Lamond laboratory at the University of Dundee where tandem mass tag proteomics was performed using high pH reverse phase liquid chromatography followed by analysis on a Fusion Orbitrap mass spectrometer. The process tags each protein with at least one chemical reporter. Reporter intensities were $\log_2$ transformed prior to reporting.

2.4.4.10 Platelet immunoblotting

Proteins were extracted from pelleted iPSC using an NP-40 lysis buffer (see 'Recipes') and quantified by Bradford assay. 30µg protein was loaded onto each well. The NuPage electrophoresis system (Thermo Fisher Scientific) was used, using the materials, equipment and methods specified in the manufacturer's protocol, including a 10% Novex Bis-Tris gel, PVDF membrane and MOPS running buffer (cat. NP0301BOX, 88518, NP0001 respectively, Thermo Fisher Scientific). Semi-wet transfer was performed using the XCell II Blot module.

5% milk powder was used for blocking and antibody dilutions and all stages were shaken at 70rpm. Primary antibodies were incubated overnight at 4^{o}C, followed by 3x 10 minute TBS-T washes. Secondary antibodies were incubated for 1 hour at RT, followed by 3x 10 minute TBS-T washes.

Primary antibodies used on platelet extracts were rabbit anti-human SNG antibody (1:1000), a gift from Prof. Andreas Greinacher of the University of Griefswald and mouse and rat anti-human ABCC4 antibodies (1:500, cat. ab56675 and ab15602, respectively, Abcam). The loading control was mouse ant-human β-actin (1:1000, cat. ab6276, Abcam). Secondary staining was with goat anti-rabbit, rabbit anti-mouse and goat anti-rat horse-radish peroxidase-conjugated antibodies (1:10,000, cat. ab97051,

ab97046, ab97057, respectively, Abcam). Secondary antibodies were all a gift from Dr. Jose Guererro, University of Cambridge. The membrane was developed using an Amersham hypercassette and Amersham ECL detection reagent (cat. RPN2106, GE Healthcare).

2.4.4.11 Clinical platelet/ coagulation tests prior to the commencement of this study.

Assays listed in section 2.3.2 'Data collection' were performed on samples from the *ABCC4* pedigree by the Katherine Dormandy Haemostasis and Thrombosis laboratory and the Haematology laboratory at the Royal Free NHS Trust, and for the *KDSR* pedigree by the Haematology Laboratory at KU Leuven.

2.5 Molecular techniques

2.5.1 DNA manipulation

2.5.1.1 DNA visualisation

Snapgene software (https://www.snapgene.com/) was used to visualise DNA manipulation including primer positions, cloning, Sanger sequencing alignments, restriction digestion and CRISPR Cas9 gene editing strategies. Sequences were obtained from the NCBI Reference Sequence Database (https://www.ncbi.nlm.nih.gov/refseq/) or manufacturers of plasmids and oligonucleotides.

2.5.1.2 DNA extraction

Genomic DNA (gDNA) extraction from cell cultures was performed using the Wizard Genomic DNA Purification Kit (cat. A1125, Promega) following the manufacturer's instructions.

2.5.1.3 Polymerase chain reaction (PCR)

All standard PCR was undertaken using the KAPA Hifi Hotstart Readymix (cat. KK2601, Roche) following the manufacturer's instructions on a GeneAmp PCR System 9700 thermal cycler. All products were annealed at 65^{o}C. Primers used for confirmation of *KDSR* and *ABCC4* mutations in samples are given in Table 2. These were purchased as lyophilised desalted oligonucleotides (Sigma).

Gene	Primer name	Sequence 5' to 3'	5' nucleotide position (RefSeq accession no.)	Amplicon size
ABCC4	ACCV1F	TGTTCACAGACAGTTTCGGG	143,033 (NG_050651.1)	384bp
	ACCV1R	CTGAATGACTCCTGGAAGCC	143,417 (NG_050651.1)	
KDSR	KF154F	ATTCTCCATGTGCACAGCCTT	21,135 (NG_028249.1)	312bp
	KF154R	AAGCGGAATGTATATTGCTGA	21,447 (NG_028249.1)	
	KF236*F	TCCAGATGGACCTTATTCTAG	33,245 (NG_028249.1)	359bp
	KF236*R	GATATGACAAATCTGACAGAC	33,604 (NG_028249.1)	

Table 2 Primers for Sanger sequencing confirmation of variants identified by WGS in cases with *ABCC4* and *KDSR* variants, and in iPSC derived from those cases.

1.5% Agarose gels were used for all purposes in this study. 1.5g of agarose gel was dissolved by heating in in 100ml 1x TBE buffer (CIMRCK). When cool, 0.1µl/ml of SYBR Safe DNA gel stain (cat. SS3102, Thermo Fisher Scientific) was added. PCR products were mixed at 1:4 ratio with 6X Orange loading dye (cat. R0631, Thermo Fisher Scientific). Samples were run alongside a 1kb Generuler DNA ladder (cat. SM0311, Thermo Fisher Scientific). Gels were run at 120mV for 30 minutes and DNA was imaged with a Syngene Chemi Genius or InGenius^3 UV light box. Following standard PCR DNA was purified using the Qiaquick PCR purification kit (cat. 28104, Qiagen) and quantified by spectrophotometry using the Labtech Nanodrop Spectrophotometer at 260nm. Sequencing was performed by Source Bioscience, Cambridge, UK. Material was supplied as specified by the service provider.

2.5.1.4 Transformation and purification of plasmids

Vectors for expansion were transformed into NEB 5α competent *E.coli* (cat. C2987, NEB) following the manufacturer's instructions. The Qiaprep miniprep kit (cat. 27104, Qiagen) was used to purify plasmids from <5ml bacterial cultures. The Promega Pureyield Maxiprep (cat. A2393, Promega) was used to purify plasmids from larger bacterial cultures. All plasmids were eluted in nuclease-free water.

2.5.1.5 Restriction endonuclease digestion

Restriction digestion was performed following protocols recommended on the NEB online tool (https://www.neb.com/tools-and-resources). All enzymes were purchased from NEB.

2.5.2 RNA and cDNA manipulation

2.5.2.1 RNA extraction and validation

Approximately 2x10^9 platelets following leucodepletion, 1x10^5 iPSC and 0.1 - 1 x 10^5 iMK were collected as described in previous sections, pelleted by centrifugation and resuspended in 500µl Trizol reagent (cat. 15596026, Thermo Fisher Scientific) in a fume hood. Samples were kept at -80^{o}C until further processing.

RNA was manipulated in a fume hood, limiting environmental nuclease contamination with RNAaseZap (cat. AM9780, Thermo Fisher Scientific). RNA was extracted using the Direct-zol RNA MiniPrep kit (cat. R2050, Zymo Research) which includes a DNAase digestion step. Platelet and iMK RNA were quantified using the Qubit RNA HS Assay kit (cat. Q32852, Thermo Fisher Scientific). iPSC RNA was quantified using the Qubit RNA BR Assay kit (cat. Q10210, Thermo Fisher Scientific). Both were used with the Qubit 2.0 Fluorometer. RNA quality was assessed with the Agilent 2100 Bioanalyser, using the Agilent RNA 6000 Pico kit (cat. 5067-1513, Agilent).

2.5.2.2 Reverse transcription and RTqPCR

RNA was reverse transcribed using the Superscript III First-Strand Synthesis SuperMix kit (cat. 18080-400, Thermo Fisher Scientific). RTqPCR was performed using the Brilliant II SYBR Green qPCR Mastermix with Low Rox (cat. 600830, Agilent) on the Stratagene MX3000P using Stratagene MxPro software. Ct values obtained for *ABCC4* and *KDSR* were normalised relative to expression of housekeeping gene

GUSB. Primers were designed to cross exon-exon junctions and are detailed in Table 3. These were purchased as lyophilised desalted oligonucleotides (Sigma). All annealing temperatures were set to 64^{o}C.

Target	Primer name	Sequence 5' to 3'	5' nucleotide position (Refseq accession no.)	Amplicon size
GUSB ORF	GUSB_F	ACGTGGTTGGAGAGCTCATT	22,942 (NG_016197.1)	120bp
	GUSB_R	CTCTGCCGAGTGAAGATCCC	26,321 (NG_016197.1)	
KDSR ORF	KQ3_F	GCAGATGGAGGTGAAGCCATA	755 (NM_002035.2)	228bp
	KQ3_R	GCCGAGAGCATGTACCCATC	983 (NM_002035.2)	
ABCC4 ORF	QPCR3_F	CCAAGATGTTGCCTATGTGCT	2160 (NM_001301829.1)	260bp
	QPCR3_R	ACCGGAGCTTTCAGAATTGAC	2420 (NM_001301829.1)	
KDSR 5'UTR	K5UTR1_F	TGGGCCTTTTCCGCACTATT	1045 (NM_002035.2)	235bp
	K5UTR1_R	AGCACTGGTCCAATCTGACG	1280 (NM_002035.2)	
dCas9 ORF	dCas9.2F	GCACGTGGCCCAAATTCTCG [136]	13,542 (phage TRE dCas9-KRAB, Addgene #50917)	316bp
	dCas9.2R	AGCGGTGGCCTTGCCTATTT [137]	13,660 (phage TRE dCas9-KRAB, Addgene #50917)	

Table 3 RTqPCR primers. ORF, open reading frame. UTR, untranslated region.
GUSB primers were provided as a gift by Dr J Lambourne of the University of Cambridge. dCas9 ORF primer sequences were obtained from Zhou *et al.* [137]

2.5.2.3 cDNA library preparation

I prepared the iPSC libraries. iMK libraries were prepared with the assistance of Ms. Frances Burden of the University of Cambridge. Mr. John Lambourne prepared the platelet libraries. Libraries were synthesized from RNA using the KAPA Stranded RNA-Seq Kit with Riboerase (cat. 07962304001, Roche), using adapters listed in and Agencourt AMPure XP beads (cat. A63880, Beckman Coulter) for purification. Libraries were quantified by RTqPCR using the KAPA Library Quantification Kit (cat. 07960140001, Roche) using a 1:160 final dilution of sample. RNA sequencing (RNA-seq) was performed by the Welcome Trust Sanger Institute on the HiSeq4000 using 150bp read length paired-end sequencing. All lines on which comparisons were made were run on the same lane. Sequences were trimmed and aligned by Dr. Luigi Grassi and Dr. Denis Seyres of the University of Cambridge. Assistance was provided by Dr. Grassi, Dr. Seyres, Dr. Frontini and Dr. Ernest Turro in the analysis.

A001	ATCACG
A002	CGATGT
A003	TTAGGC
A004	TGACCA
A005	ACAGTG
A006	GCCAAT
A007	CAGATC
A008	ACTTGA

Table 4 Adapters for RNA-seq

2.5.3 CRISPR Cas9n knock-in strategy

2.5.3.1 sgRNA and ssODN design

Figure 9 and Table 5 summarise the strategy of this experiment. Three pairs of single guide RNA (sgRNA) were designed to target the region of the pArg154Trp variant using the Sanger CRISPR finder (http://www.sanger.ac.uk/htgt/wge/find_crisprs) (Figure 10). The software identified and ranked target sites within the specified genomic region and predicted off target effects. sgRNA pairs were designed to be 17-20 nucleotides in length. The highest ranked sgRNAs were analysed using the NCBI nucleotide BLAST tool https://blast.ncbi.nlm.nih.gov/Blast.cgi and the highest ranked sgRNAs with at least three mismatches to unintended targets were chosen to go forward.

The optimal distance of the intended Cas9n-mediated breaks from the mutation was not published, therefore a range of distances was specified over the different conditions. Overhangs were added to the sgRNA sequence to allow ligation into the *BbsI* restriction endonuclease sites in the Cas9 plasmid. G-C base pairs were also added at the 5' end of the guide sequence where there were none to facilitate transcription initiation from the U6 promoter. The sgRNAs were purchased as lyophilised desalted oligonucleotides (Sigma) (Table 6).

Three single stranded oligonucleotide (ssODN) templates were designed, one for each sgRNA pair. The p.Arg154Trp variant and at least three silent mutations were introduced to each template to reduce the likelihood of sgRNA-binding and therefore Cas9n cleavage of the template. One of these mutations was in the site corresponding to the proto-spacer adjacent motif (PAM). The first templates were designed to be 170bp long and were purchased as desalted, lyophilised oligonucleotides (Sigma), suffix 'short' in Table 7. 200bp ssODNs were subsequently purchased as desalted, lyophilised 'Ultramer' oligonucleotides (Integrated DNA Technologies), suffix 'long' Table 7.

The targeted S4 iPSC line was shown by Sanger sequencing to carry the reference sequence at the variant position using the S4 primer pair in Table 8.

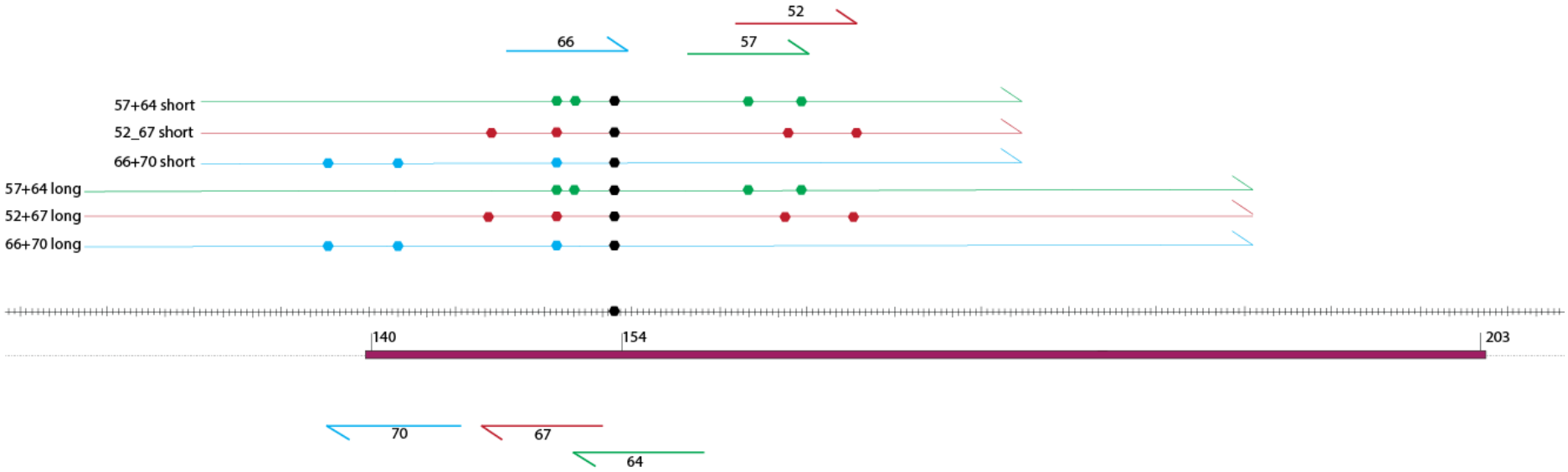


Figure 9 Schematic to illustrate CRISPR sgRNA and ssODN design strategy.
The maroon bar corresponds to the exon containing p.Arg154 (labelled). The central bar represents genomic DNA, and each gradation mark corresponds to a nucleotide. ssODNs and sgRNAs are denoted by coloured bars with each colour representing a separate condition, and are to scale. Polygons in the ssODNs represent variation from the reference genome, with p.Arg154Trp in black and silent mutations in colour.

Condition	sgRNAs	ssODN
1	sgRNA 66 and sgRNA 70	ssODN 66+70
2	sgRNA 52 and sgRNA 67	ssODN 52+67
3	sgRNA 57 and sgRNA64	ssODN 57+64

Table 5 Experimental conditions

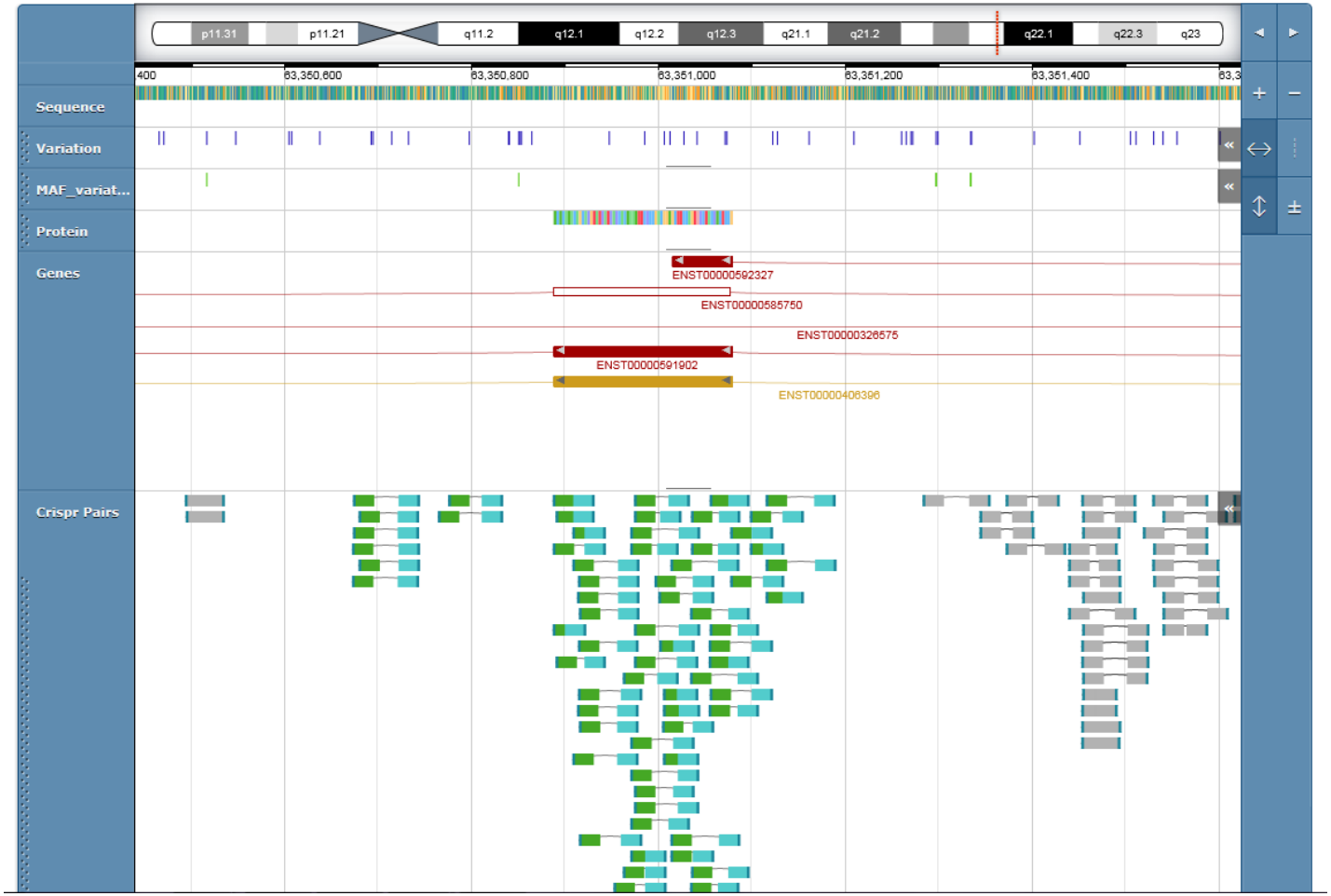


Figure 10 Screenshot of genome browser at http://www.sanger.ac.uk/htgt/wge/find_crisprs used to identify sgRNA candidates for BLAST screening.

Name	5’ to 3’ sgRNA sequence including PAM	Oligos without PAM, including overhangs and G-C pair addition	5’ nucleotide position of sgRNA (NG_028249.1)
KDSR_sgRNA_66	GGCAGCGTGTACCCCAGC-CGG	5’ **CACC**GGCAGCGTGTACCCCAGC	21,219
		5’ **AAAC**GCTGGGGTACACGCTGCC	
KDSR_sgRNA _70	TGATGCTCATTAACCTCTGC-AGG	5’ **CACCG**TGATGCTCATTAACCTCTGC	21,210
		5’ **AAAC**GCAGAGGTTAATGAGCATCA**C**	
KDSR_sgRNA _52	AAGGAGCGCCGGGTGGGC-AGG	5’ **CACCG**AAGGAGCGCCGGGTGGGC	21,258
		5’ **AAAC**GCCCACCCGGCGCTCCTT**C**	
KDSR_sgRNA _67	GGGGTACACGCTGCCC-AGG	5’ **CACC**GGGGTACACGCTGCCC	21,233
		5’ **AAAC**GGGCAGCGTGTACCCC	
KDSR_sgRNA _57	CCACCATGAAGGAGCGCC-GGG	5’ **CACCG**CCACCATGAAGGAGCGCC	21,250
		5’ **AAAC**GGCGCTCCTTCATGGTGG**C**	
KDSR_sgRNA _64	TGGTGATCACGGCCCGGCTG-GGG	5’ **CACCG**TGGTGATCACGGCCCGGCTG	21,252
		5’ **AAAC**CAGCCGGGCCGTGATCACCA**C**	

Table 6 sgRNA sequences including PAM and overhangs for BbsI-mediated ligation (bold) and G-C pair addition (bold) to facilitate U6 transcription.

57+64 short	CCTCATGTGGCCTCTTTGTTCCCTGCAGAGGTTAATGAGCATCAATTACCTGGGCAGCGTATATCCCAGCTGGGCCG TGATCACCACCATGAAAGAGCGCCGTGTGGGCAGGATCGTGTTTGTGTCCTCCCAGGCAGGAC
66+70 short	CCTCATGTGGCCTCTTTGTTCACTGCAGAGGTTGATGAGCATCAATTACCTGGGCAGCGTATACCCCAGCTGGGCCG TGATCACCACCATGAAGGAGCGCCGGGTGGGCAGGATCGTGTTTGTGTCCTCCCAGGCAGGAC
52+67 short	CCTCATGTGGCCTCTTTGTTCCCTGCAGAGGTTAATGAGCATCAATTATCTGGGCAGCGTATACCCCAGCTGGGCCG TGATCACCACCATGAAGGAGCGACGGGTGGGCAGAATCGTGTTTGTGTCCTCCCAGGCAGGAC
57+64 long	CACAGCCTTGAGGCTTTTGACCTCATGTGGCCTCTTTGTTCCCTGCAGAGGTTAATGAGCATCAATTACCTGGGCAG CGTATATCCCAGCTGGGCCGTGATCACCACCATGAAAGAGCGCCGTGTGGGCAGGATCGTGTTTGTGTCCTCCCAG GCAGGACAGTTGGGATTATTCGGTTTCACAGCCTACTCTGCATCCAA
52+67 long	CACAGCCTTGAGGCTTTTGACCTCATGTGGCCTCTTTGTTCCCTGCAGAGGTTAATGAGCATCAATTATCTGGGCAGC GTATACCCCAGCTGGGCCGTGATCACCACCATGAAGGAGCGACGGGTGGGCAGAATCGTGTTTGTGTCCTCCCAGG CAGGACAGTTGGGATTATTCGGTTTCACAGCCTACTCTGCATCCAA
66+70 long	CACAGCCTTGAGGCTTTTGACCTCATGTGGCCTCTTTGTTCACTGCAGAGGTTGATGAGCATCAATTACCTGGGCAG CGTATACCCCAGCTGGGCCGTGATCACCACCATGAAGGAGCGCCGGGTGGGCAGGATCGTGTTTGTGTCCTCCCAG GCAGGACAGTTGGGATTATTCGGTTTCACAGCCTACTCTGCATCCAA

Table 7 ssODN sequences

Primer name	5’ to 3’ sequence	5’ nucleotide position (NG_028249.1)	Amplicon size
S4 fwd	GGGAGACCCAAGGCTTGAAA	20,743	689bp
S4 rev	TGCTGAGCGGAGAGGAATAAA	21,431	
KF154F	ATTCTCCATGTGCACAGCCTT	21,135	312bp
KF154R	AAGCGGAATGTATATTGCTGA	21,446	

Table 8 Primers for confirmation of reference sequence at target region in S4 iPSC (S4fwd/rev), for assessment of variant knock-in following nucleofection (K154F/R)

2.5.3.2 Episomal plasmid preparation

A published protocol for CRISPR/Cas9 genome editing was followed [126]. In brief, sgRNAs were phosphorylated, annealed, diluted at 1:200 and ligated into a linearised pSpCas9n (BB)-2A GFP plasmid (pX461, Addgene #48140). The plasmid was purified using PlasmidSafe exonuclease to digest residual linearised DNA, then transformed within 24 hours into NEB 5α competent *E.coli* (cat. C2987, NEB) following the manufacturer's instructions. The bacteria were spread on agar/ ampicillin plates and incubated at 37^{o}C overnight.

Plentiful colonies were obtained, with no growth on a no-insert negative control plate. Three colonies were picked per sgRNA pair and purified by miniprep then maxiprep, as described above, in LB medium (CIMRCK) containing 100µg/ml ampicillin (cat. A9518**,** Sigma).

A restriction digest using *PvuI* and *BbsI* enzymes in NEB 3.1 buffer (cat. R0150S, R0539S, B7203S respectively, NEB) supported insertion of all sgRNAs (the *BbsI* site is destroyed by the insert). This was confirmed on Sanger sequencing (Source Bioscience, Cambridge, UK) using the U6-Fwd primer (5' GACTATCATATGCTTACCGT).

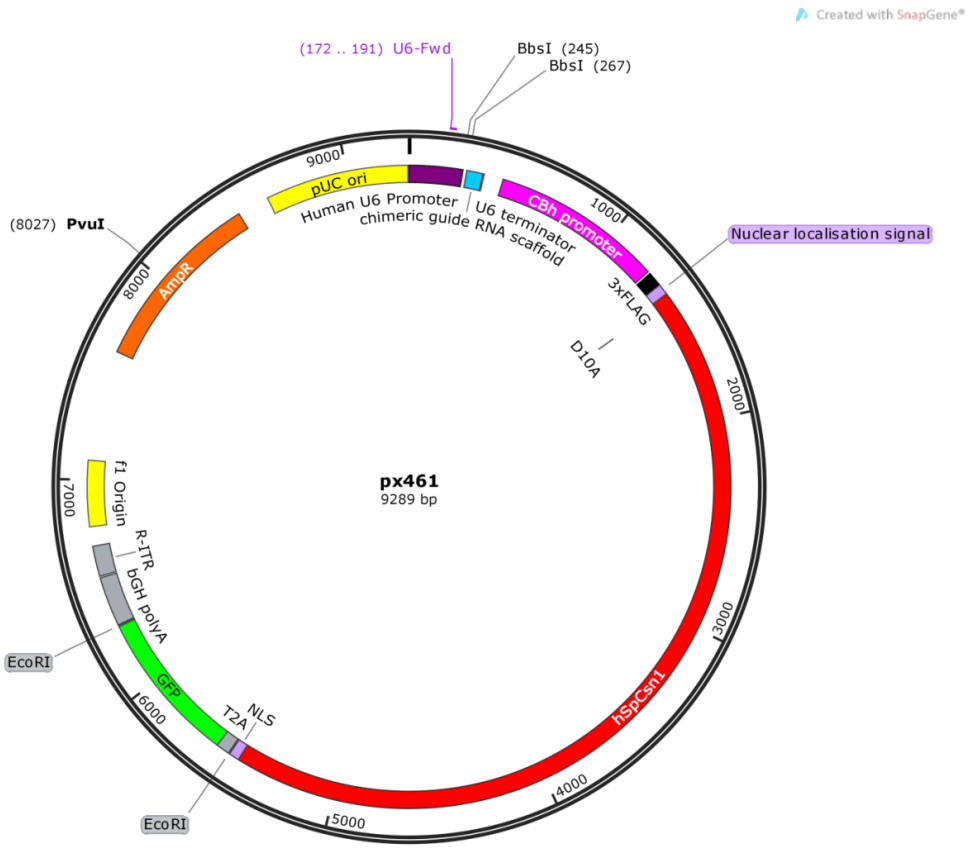


Figure 11 pSpCas9n (BB)-2A GFP plasmid (pX461, Addgene plasmid ID 48140).

2.5.3.3 Nucleofection

Cells were passaged using TryPLE as described above, resuspended in 1ml $AE6^{++}$ medium and counted by haemocytometer. $8x10^5$ cells were aliquotted into Eppendorf tubes, including control tubes (no treatment and electroporation only, without plasmid/ ssODN). Cells were centrifuged (3minutes, 200g). The pellets were resuspended in Amaxa Human Stem Cell Nucleofector Kit 2 DNA mix (cat. VPH-5022, Lonza) ssODN was added to a final concentration of 10µM. 2.5µg of each plasmid was added to the appropriate condition. The mix was transferred to an electroporation cuvette and electroporated using the B-016 cycle on the Amaxa Nucleofector II machine. 500ul $AE6^{++}$ containing 10µM ROCK inhibitor was added directly to each cuvette. Cells were seeded into plates and incubated. After 24 hours cells medium was replaced with $AE6^{++}$. GFP^+ expression was transient so cells were sorted at 48 hours.

2.5.3.4 Cell sorting

Cells were passaged using TrypLE. GFP^+ cells were sorted on the BD FACSDiva 8.0.1 by the NIHR Cambridge BRC Cell Phenotyping Hub. Viable single cells were plated in single wells of a 96 well plate with ROCK inhibitor for 24 hours. Three 96 well plates were seeded, with one plate per condition. A mixed pool was also seeded in a 6 well plate, and once established, cryopreserved. Medium was changed at 24 hours. Following the second nucleofection penicillin-streptomycin at 10ml/l (cat. P4333, Sigma) was added to the culture for 5 days to prevent contamination following cell sorting.

2.5.3.5 PCR for knock-in

3 wells were seeded for each clone following nucleofection, for cryopreservation, gDNA extraction and ongoing culture. Approximately $5x10^4$ cells were collected and DNA extracted as described above. PCR was performed as described above using the primers listed in Table 8.

2.5.3.6 Cloning into pGEM-T Easy

Cloning was required to separate alleles in clones that were heterozygous for indels, to allow Sanger sequencing. Cloning was performed into the pGEM-T Easy Vector System II (cat. A1380, Promega) following the manufacturer's instructions.

In brief, PCR amplicons were purified as described above and A-tailed to allow ligation using the Amplitaq Gold DNA polymerase kit and dATP (cat. N8080241 and R0141 respectively, Thermo Fisher Scientific). The amplicons were purified again and ligated into PGEMT-Easy. The plasmids were transformed into JM109 competent cells (cat. L2001, Promega) and plated on Fast-Media Amp XGal agar plates (cat. fas-am-x, Invivogen). Successful ligation of amplicons into the pGEM-T Easy vector interrupts the *LacZ* gene which alters the colour of the resulting *E.coli* colonies from blue to white, allowing selection for sequencing. Between five and ten colonies were picked, and plasmids were purified by miniprep and sent for Sanger sequencing using the T7F primer TAATACGACTCACTATAGGG.

2.5.4 CRISPR dCas9 knock-down strategy

2.5.4.1 sgRNA design

The promoter of *ABCC4* was identified by Dr. Romina Petersen of the University of Cambridge as the region at the 5' end of ABCC4 with the highest H3K27 acetylation and H3K4 trimethylation ChIP-seq signature in MK ([92], Figure 12). The location was defined at Chr 13:95,949,400-95,955,400 (Ensembl v.70, GRCh37build). The promoter was split into 250bp segments, and within each segment sgRNA targets were identified using the Zhang lab CRISPR design tool (http://crispr.mit.edu/) which ranks sgRNAs by a quality score that accounts for the degree of mismatch at other genomic loci. High-ranking sgRNAs were checked to ensure unique binding at the target locus using the NCBI Nucleotide BLAST tool (https://blast.ncbi.nlm.nih.gov/). 12 sgRNAs were chosen (Figure 13). *BsmBI* restriction enzyme-compatible overhangs were added to the sgRNA sequences and ordered as standard desalted oligonucleotides (Sigma).

2.5.4.2 sgRNA cloning into LentiGuide Puro plasmid p52963

The Zhang lab 'Target guide sequence cloning protocol' (http://genome-engineering.org/gecko/, [138, 139]) was followed. In brief, the LentiGuide Puro plasmid (Addgene #52963) was digested with *BsmBI*, dephosphorylated, then then gel-purified using the Qiaquick Gel Extraction kit (cat. 28704, Qiagen), selecting only the larger, 8kb band. sgRNA oligos were phosphorylated and annealed then ligated into the digested vector in twelve separate reactions. The products were transformed in Stbl3 competent E.coli (cat. C737303, Thermo Fisher Scientific). Plentiful colonies were obtained, with a single false positive colony on a no-insert negative control plate. Colonies were picked for each sgRNA and purified by miniprep, as described above, in LB medium (CIMRCK) containing 100µg/ml ampicillin (cat. A9518**,** Sigma). A restriction digest using *NcoI* and *BsmBI* enzymes in NEB 3.1 buffer (cat. R0193s and R0580S, respectively, NEB) supported insertion of all sgRNAs except P8 (the insertion of the sgRNA removes the *BsmBI* recognition site). These clones were confirmed on Sanger sequencing (Source Bioscience, Cambridge, UK) using the U6-Fwd primer (5' GACTATCATATGCTTACCGT).

2.5.4.3 Assessment of sgRNA efficiency in HEK293T cells by Surveyor assay

Plasmids carrying sgRNAs were separately co-transfected into HEK293T cells with Cas9-GFP (cat. PX458, received as a gift from Dr. Mattia Frontini of the University of Cambridge) as described in section 2.4.3.3, including a no-sgRNA control. 48 hours following transfection cells were 50-70% GFP positive and genomic DNA was extracted as described in section 2.5.1.2. PCR was performed using primers listed in Figure 13, ensuring coverage of the sgRNA target site in each instance, and cleaned up using a Zymo DNA Clean and Concentrator kit (cat. D4013, Zymo). PCR amplicons were digested using the T7 Endonuclease (cat. M0302, NEB), following the manufacturer's protocol [140]. This endonuclease detects and cleaves mismatches between amplicons, in this case when Cas9-cleaved strands have been re-annealed with un-cleaved strands. Cleavage was confirmed and all sgRNAs were taken forward for viral particle production.

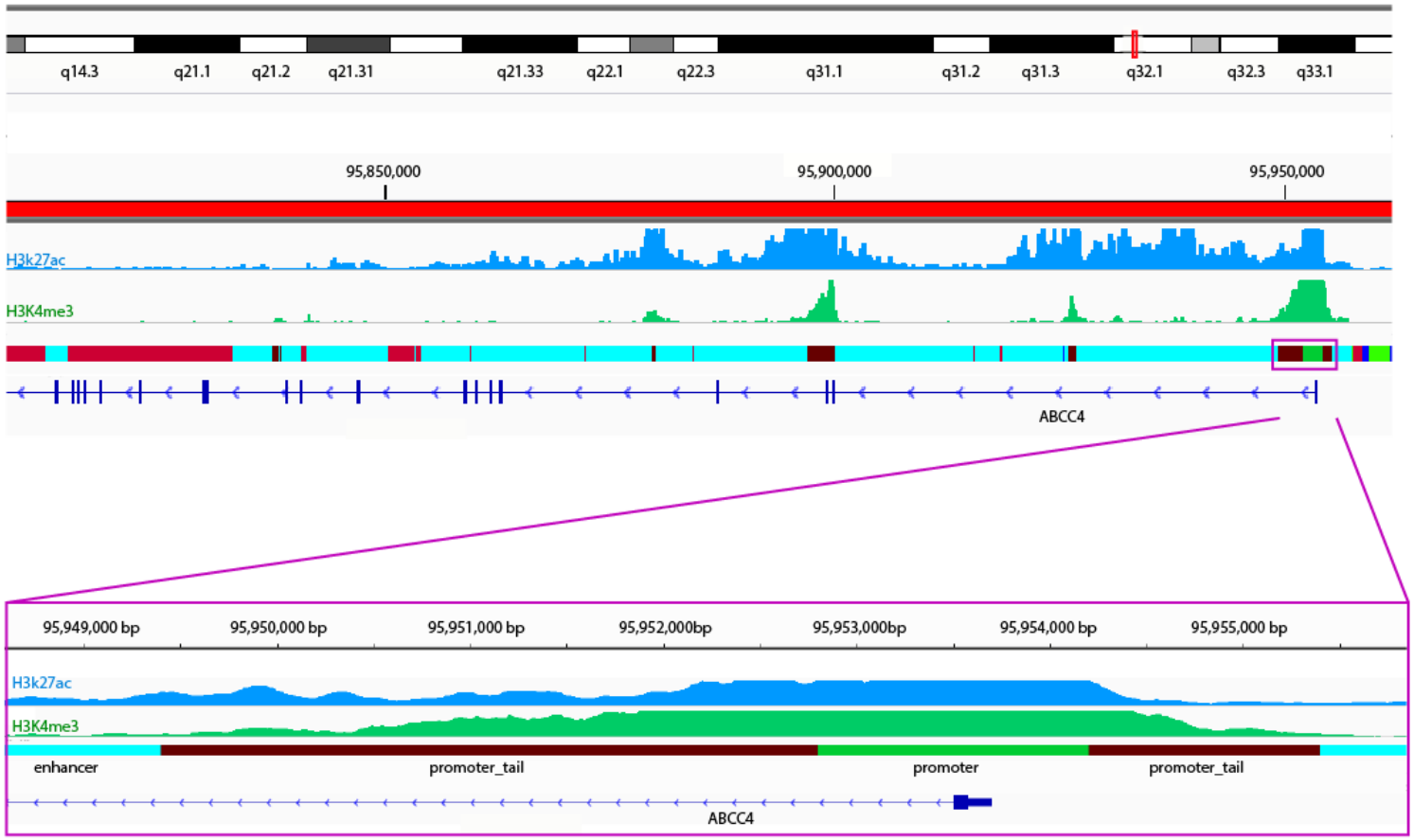


Figure 12 Schematic to show the *ABCC4* promoter location

The location was identified by high H3K27 acetylation and H3K4 trimethylation ChIP-seq signatures in MK [92], at Chr 13:95,949,400-95,955,400 (Ensembl v.70, GRCh37 build). Original data courtesy of Dr. Romina Petersen, University of Cambridge. Mapping performed using Integrated Genome Viewer software.

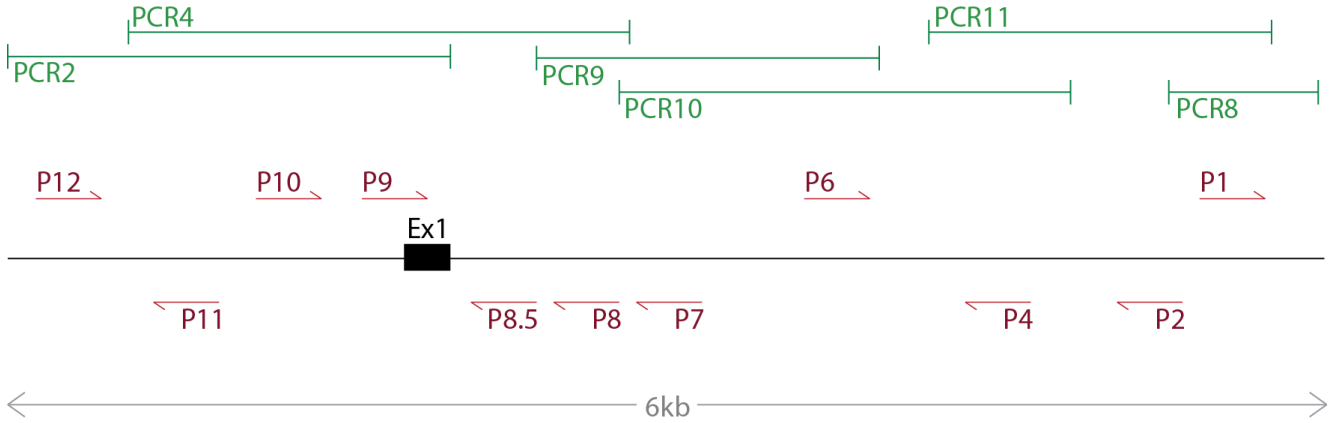


Figure 13 Schematic of ABCC4 promoter region (black) shows sgRNA location (red) relative to PCR primers (green).

sgRNA	5' to 3' sequence	5' nucleotide position (NG_050651.1)
P1	GAGTCATGCATTATGGCTGTTGG	9,090
P2	TCAAGGGGGTGGGAACTACGAGG	8,587
P4	ATTTCCGACAGCACCTGAACTGG	7,761
P6	GGTGGGGGTCTCCAATTTTTTGG	6,779
P7	CTCTGAAGGTTAGCGATTCCAGG	6,144
P8	TTGGATCGAGCTGTCTCGTAAGG	5,792
P8.5	AGGGACCACGCGGCCGGCGTGGG	5,398
P9	CTCCAGGCGGCGGCGCGTAGCGG	4,901
P10	CCTCCCATGGCACCGTCGTTTGG	4,477
P11	CACACACAGTGTCACGATTATGG	4,072
P12	TGGGCACTTCTCATGCACACTGG	3,481

Table 9 sgRNAs selected for CRISPR knockdown strategy.

Primer	5' to 3' sequence	5' nucleotide position (NG_050651.1)	Amplicon size
PCR2F	AGAGGTGGGAGTACACAAGC	3,409	1753bp
PCR2R	GGCTTCACCTCCTGGTACAC	5,161	
PCR4F	AAGCTCAGAGGTCCACCGTA	4,270	1803bp
PCR4R	CAGGGGAAAGAAGCTGGTCA	6,072	
PCR8F	TTGTGTACCCTTCCCACTGC	8,816	835bp
PCR8R	CCAGCTGGCCACAGTTCATA	9,650	
PCR9F	CGTCCCGTTGTACCTCTCTG	5,798	1745bp
PCR9R	ATCAATGGGAGCGCATCACT	7,542	
PCR10F	ACGGAGAAAGACTCGGGTGA	6,624	1438bp
PCR10R	GGAAGCAGTAAGCGCCTCAA	8,061	
PCR11F	AAGTGATGCGCTCCCATTGA	7,522	1600bp
PCR11R	CCACAATTGCCAACAGCCAT	9,121	

Table 10 PCR primers for Surveyor assay.

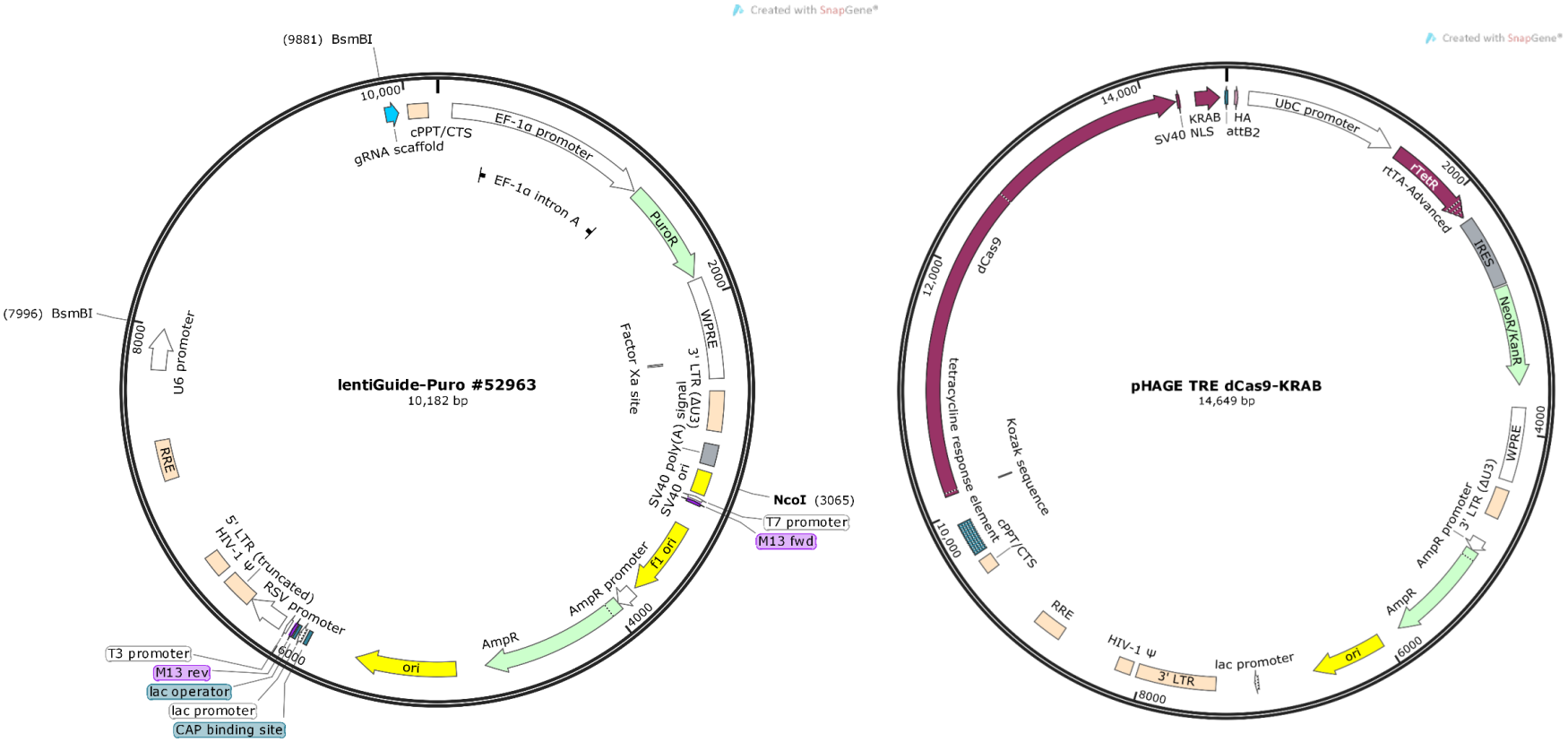


Figure 14 Left, LentiGuide-Puro (Addgene #52963). Right, phage TRE dCas9-KRAB (Addgene #50917)

2.5.4.4 Transduction into dCas9-KRAB-integrated iPSC

Lentiviral particles containing the plasmids carrying the sgRNAs were synthesised as described in section 2.5.5.2 and 3.

The iPSC line A1ATD-1 containing the lentivirally-integrated Phage TRE dCas9-KRAB (Addgene #50917, Figure 14) was a gift from Dr. Mattia Frontini of the University of Cambridge. In brief, Dr. John Lambourne and Ms. Isabel Rosa of the Frontini laboratory used a stock iPSC line (A1ATD-1) known to forward program reliably and with high yield of viable $CD41^+$ iMK at day 20. They stably integrated a lentiviral vector carrying an inducible dCas9-KRAB complex under the control of a tetracycline-inducible (tet-on) element [141].

The cell line was maintained under geneticin-G418 selection (cat. 10131035, Thermo Fisher Scientific) from the time of thawing to maintain the integrated line. Cells were transduced with lentiviral particles carrying the sgRNAs targeting the ABCC4 promoter as described in section 2.5.5.3. 72 hours following transduction puromycin hydrochloride (cat. P8833, Sigma) was added at 1µg/ml to select cells in which the sgRNA plasmid had been stably integrated. Puromycin was maintained in iPSC cultures, and stopped at the point of forward programming to iMK.

2.5.4.5 dCas9 expression in control line necessitated an alternative negative control

The theoretical advantage of the Tet-on system is that addition or omission of doxycyline provides the ideal test or control conditions, respectively. Doxycycline (cat. D9891 Sigma) was added at 1µg/ml to transduced dCas9/sgRNA-integrated iPSC in the culture for one week prior to immunoblotting and qPCR. Immunoblotting of iPSC was performed as described for platelets in section 2.4.4.10. Primary antibody staining was with mouse anti-Cas9 antibody (1:2000, cat. C15200223, Diagenode) in 5% milk, overnight at 4^oC at 70rpm. RTqPCR was performed as described in section 2.5.2.2.

The detection of dCas9 expression in the negative control by qPCR and immunoblotting may be attributable to the use of cytokines by previous users made up in bovine serum albumin, as tetracycline use in cattle is widespread. Therefore the dCas9-integrated iPSC were transduced with an empty Lentiguide-Puro vector and the test/ control conditions were altered to sgRNA/no sgRNA, respectively, with both iPSC conditions maintained in geneticin, puromycin and doxycycline.

2.5.5 Lentiviral vector expression

Lentiviral transduction and vectors for CRISPR Cas9 gene editing and iMK production are described above. In this section I discuss the use of constructs for expression of the normal open reading frame (ORF) of *KDSR* to rescue the phenotype in patient-derived iPSC.

2.5.5.1 Vectors

Lentiviral vectors overexpressing the ORF of *KDSR* were purchased from Origene. pLenti-EF1a-KDSR-mGFP (cat. CW302081) and pLenti-EF1a-KDSR-myc-DDK-IRES-Puro (cat. CW103991) were derived from

cloning TrueClone cDNA RC201153 into pLenti destination vectors PS100084 and PS100085, respectively (Figure 15). Importantly, the ORF used in these vectors is identical to transcript ENST00000591902 (Refseq accession no. NM_002035, Origene cat. RC201153), which is the transcript with the highest expression in MK (Figure 31). The mGFP and MYC-DDK (FLAG) tags were fused to the *KDSR* ORF. The EF1a promoter was chosen to optimise expression in iPSC [142]. The uncloned destination vector PS10085 (Origene) was used to generate an empty vector control line.

Vectors were replicated by transformation into *E.coli* and purification by maxiprep as described above. Following replication a diagnostic restriction digest was performed using PvuI and XhoI (cat. R01050S and R0146S, respectively, NEB) to confirm the structure of the plasmids were as expected.

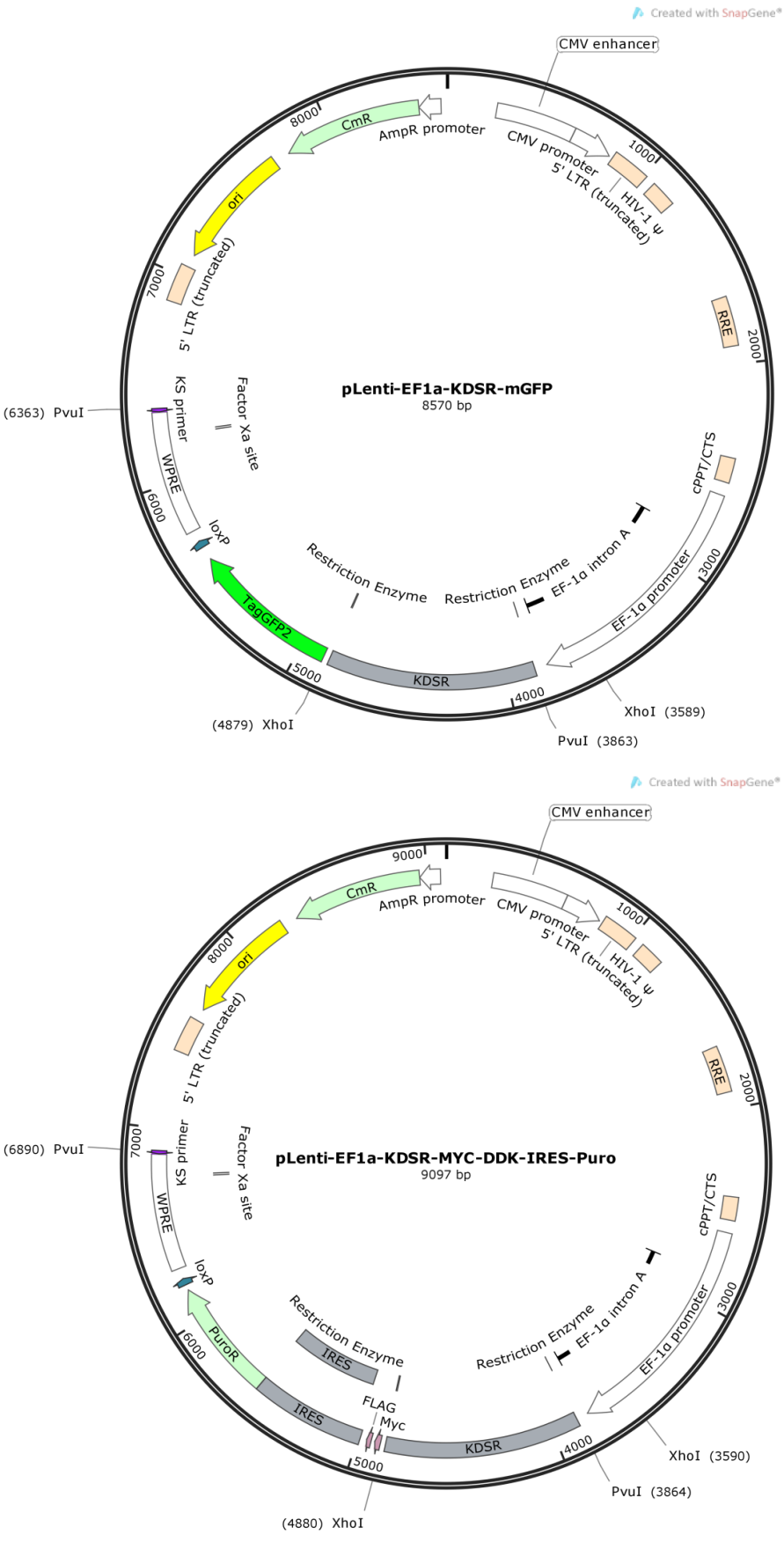


Figure 15 Vectors pLenti-EF1a-KDSR-mGFP (top) and pLenti-EF1a-KDSR-myc-DDK-IRES-Puro (bottom) were used to express the normal ORF in a patient-derived iPSC line.

2.5.5.2 Lentiviral particle production

Replication deficient lentiviral vector particles were produced by transient co-transfection of HEK 293T cells with a 2nd generation packaging system and a vector carrying the cDNA of interest. This was performed using strict sterile technique in a dedicated CL2 environment, in a laminar flow cabinet.

$4x10^6$ HEK293T cells were plated in a 10cm dish in HEK293T complete medium (see 'Recipes') and allowed to attach for 4 hours. 20µg of the vector carrying the cDNA of interest, 8µg psPAX2 and 2.5µg pCMV-VSV-G (Addgene #12260, #8454) were combined in a 5ml sterilin tube. 1ml RT DMEM (cat. D6429, Sigma) and 75µl 1mg/ml PEI (cat. 408727, Sigma) were added and the tube was incubated at RT for 20 minutes. The mixture was spread drop wise over the HEK293T cells. Cells were incubated at 37°C, 5%CO_2. 24 hours post-transfection the medium was refreshed. At 48 hours the crude supernatant containing viral particles was harvested, filtered using a sterile 45µm mesh filter and concentrated by polyethylene glycol-based precipitation (LentiX-concentrator, Clontech).

Viral particles containing pLenti-EF1a-KDSR-mGFP were titred using the eGFP marker by flow cytometry in HEK293T cells at $3.6x10^8$ TU/ml. Particles containing pLenti-EF1a-KDSR-myc-DDK-IRES-Puro were titred using mouse-anti-human anti-FLAG antibody (1:200) and secondary anti-mouse Alexa Fluor 633 (cat.) at $2.4x10^7$ TU/ml.

2.5.5.3 Lentiviral transduction

Cells were passaged with PBS-EDTA as described above, counted and $1x10^5$ cells were seeded per well of a vitronectin-coated six-well plate. Transduction medium was prepared, containing E8 medium supplemented with protamine sulphate 10µg/ml and the viral particles at the desired MOI.

Transduction of the IPD123 line with all vectors was complicated by recurrent cell death, either following transduction or following addition of selective antibiotics. Multiple conditions were used to optimise the protocol:

- MOI of 0.2, 2, 20, 40
- spinfection (2500rpm, 45 minutes at 37°C following transduction) vs. no spinfection
- transduction with polybrene 8µg/ml (cat. TR-1003, Sigma) vs. protamine sulphate 10µg/ml (cat. P4505, Sigma)
- initial seeding density of 5E5, 1E5, 5E4 cells
- transduction following TrypLE or PBS-EDTA passage
- transduction 1 day post-passage or once colonies re-established
- exposure to virus of 2 hours, 4 hours or 12 hours before first media change
- use of $AE6^{++}$ or E8 media for cell culture

The optimal conditions for this experiment were specific: 1E5 cells that had been cultured in E8 media were seeded following passage with PBS-EDTA and allowed to form small colonies over two days. These were transduced at MOI 2, in the presence of protamine sulphate, followed by spinfection and with the first media change at four hours. Few conditions yielded surviving cells.

Prior to selection with puromycin (cat. P8833, Sigma) at 72 hours post transduction a kill curve was performed on the patient-derived iPSC line targeted by pLenti-EF1a-KDSR-MYC-DDK-IRES-Puro. The optimal puromycin concentration was 0.1µg/ml, which caused cell death in 4 days. Both the KDSR-rescued line and the empty vector line were maintained in puromycin- supplemented medium.

2.5.6 Metabolic profiling

Global metabolic profiling of plasma from the proband, parents and unaffected sibling was performed by Metabolon, Inc. (Durham, North Carolina, US) using the DiscoveryHD4 platform. This is a liquid chromatography tandem mass spectrometry (LC/MS/MS) platform that has been previously described [143]. Missing values in the raw counts were imputed and treated as the limit of detection to calculate z-scores for each metabolite.

Confirmation of the sphingolipid profile lipid analysis was performed in collaboration with Dr. Albert Koulman and Dr. Ben Jenkins of the NIHR Biomedical Research Centre Core Metabolomics and Lipidomics Laboratory, University of Cambridge. We used a LC/MS method as described previously [144, 145] with the only exception that for the identification of KDS a Thermo Scientific Q-Exactive Orbitrap was used (Thermo Fisher Scientific, Hemel Hempstead, United Kingdom), set up to fragment the 300.3 ion (± 2m/z) using a relative collision energy of 30. The KDS standard was unavailable on the existing platform and was purchased from Matreya LLC, cat. no. 1876.

For metabolomic studies of iPSC, cells were treated with PBS-EDTA to lift cells as described in section 'iPSC, Passage,' counted using a haemocytometer and centrifuged (200g, 3 minutes). Pellets and 1ml aliquots of spent media were snap frozen in dry ice and stored at -80^{o}C until analysis.

For metabolomic studies of iMK, cell density was known from day 20 flow cytometry- based cell counting using fluorosphere count beads, as described in section 'Forward programming to iMK, Protocol.' A fixed volume of cell suspension containing the desired number of cells was collected and centrifuged (120g, 8 minutes, acceleration 3, brake 3). Pellets were snap frozen in dry ice and stored at -80^{o}C until analysis.

2.5.7 Zebrafish studies

Zebrafish studies were performed in collaboration, by Ms. Mara De Reys and Professor Kathleen Freson of KU Leuven. Tg(cd41:EGF) [146] embryos were injected with a *kdsr* ATG morpholino (5' ctcagaggacatgggtcaacctgat or kdsr-MO) from Gene Tools LLC (Philomath) or with buffer (control). Zebrafish *kdsr* has accession number ZDB-GENE-040426-853. Thrombocyte formation was analysed as described previously [147, 148]. Immunoblots were developed with goat anti-GFP (Rockland) and anti-

FVT1/KDSR (Clone H-149; Santa Cruz). Animal protocols were approved by the Ethical Committee of the University of Leuven.

2.5.8 Stem cell differentiation assays

Stem cell differentiation assays were performed in collaboration, by Ms. Chantal Thys and Professor Kathleen Freson of KU Leuven. CD34+ hematopoietic stem cells (HSC) were isolated from bone marrow from the propositus (at 5 years of age) and a healthy gender matched donor and were stored in liquid nitrogen. The recovered (differentiation day 0) CD34+ stem cells were cultured in StemSpan SFEM medium with StemSpan CC100 ensuring strong expansion of undifferentiated HSC for 3 days (Stem Cell Technologies, Vancouver, CA). HSC at day 3 were used for liquid MK cultures where differentiation was initiated by adding 50 ng/ml thrombopoietin, 25 ng/ml stem cell factor and 10 ng/ml interleukin 1β (Peprotech, Rocky Hill, New Jersey, US). MK were analysed using cytometry (total differentiation day 7 with FITC-labelled anti-CD41 and anti-CD42; BD Biosciences), proplatelet formation counts (total differentiation day 11), and immunostaining (phalloidin-rhodamine, anti-tubulin, anti-VWF and anti-CD63 from Sigma, Thermo Fisher, Dako, and Santa Cruz, respectively) and were performed as described previously [5, 47]. For immunostaining experiments, MK were placed for 4 hours on fibrinogen-coated coverslips and images were photographed at 63x magnification with a Zeiss Axiovert confocal microscope.

Chapter 3: The NIHR BR-RD BPD Cohort

3 The NIHR BR-RD BPD Cohort

The main objective of the NIHR BR-RD BPD study was to identify novel genes responsible for rare diseases. This was performed under the hypothesis that similar phenotypes and phenotype severity are explained by variants in a single gene. In view of the vast number of variants provided by high throughput sequencing, computational algorithms are required to assign probabilities of pathogenicity to genes, based on the interpretation of phenotypes. To allow this, the phenotypes and severity that are so variably described in clinical practice must be converted into a standard set of codes, and validated for use.

This chapter consists of:

- A description of the recruitment performance of the study
- A summary of the novel variants identified during the course of the study
- The results of two interim in-silico validation experiments in this cohort:
    - Assessment of the ability of the HPO to capture the richness of phenotypes in the cohort
    - Evaluation of the use of bleeding scores to capture the severity of BPD-specific phenotypes in the cohort

3.1 Recruitment and sequence assessment

3.1.1 Recruitment at the Royal Free NHS Trust

At the time of recruitment, there were 1,402 UK registrations of PFD [60], of which 240 had been registered at the Katherine Dormandy Haemostasis and Thrombosis Unit at the Royal Free London NHS Foundation Trust (RFH) over the previous twenty years. 42 cases with a UBD were also identified on the RFH registry. 282 clinical records were screened for suitability. 145 cases were contactable. Additional cases were identified by screening new referrals to the department and day unit attendance.

Potential participants were invited for a face-to-face interview. 116 cases were recruited at RFH, of whom 103 were index cases, 28 were affected family members and 3 were unaffected family members. Of the 103 index cases, 78 were female and 25 were male. The individual diagnoses in the index cases are given in Table 11. The majority had a prior diagnosis of unclassified bleeding (16%), storage pool disorder (27%) or a combined ADP and epinephrine agonist defect (40%). All cases either had clinical bleeding symptoms, or were identified by family screening with a first degree relative with clinical bleeding and a similar laboratory phenotype to the case recruited. One case had severe bleeding symptoms with a history of spontaneous bleeds. The severity in the remainder was mild or moderate.

3.1.2 Local assessment of sequence data in Royal Free cases

Sequence data returned was screened for variants in all known BPD genes. The full list of genes is given in Table 1 and was based on genes identified from the Human Genome Mutation Database (HGMD) [149] and the gene list at www.thrombogenomics.org.uk [66]. Coagulation factor genes were not

considered in these analyses as RFH cases were screened for coagulopathy and were not enrolled if testing indicated a deficiency.

Variants were assessed for pathogenicity if they were non-synonymous missense, frameshift, structural or affected splice sites, the 5'UTR or 3'UTR and if they had a minor allele frequency <1/1000 in GnomAD [95] and a CADD score >10 [102]. Variants were assessed in the context of the phenotype, pedigree data, likely inheritance pattern, predicted protein effects and literature searches.

3.1.3 Assigning causality to rare variants in known BPD genes in Royal Free cases

Potentially pathogenic variants that could explain the clinical phenotypes were identified in *ANKRD26, ACTN1, GATA-1* and *TBXA2R* (Table 12). These were confirmed by Sanger sequencing in the Cambridge Clinical Genetics laboratory. Likelihood of pathogenicity was assigned to variants following referral to the ThromboGenomics multidisciplinary team (MDT) [66], following EuroGentest guidelines [150]:

1. *Pathogenic variant*: a variant present in HGMD with a concordant phenotype to the case OR four previous cases in any cohort with a matched phenotype and genotype
2. *Likely pathogenic variant*: a non-HGMD variant in a gene for which less than four previously reported cases have a similar phenotype
3. *Variant of unknown significance*: variants that did not fulfil criteria for (1) or (2)

The *ANKRD26* variant (10:27389382 T>C, 5' UTR variant) was determined to be pathogenic, because this variant has been described in the literature in multiple cases with a matching phenotype. Co-segregation was undertaken and the variant was also identified in three family members. *ANKRD26* variants are associated with a risk of transformation to acute myeloid leukaemia [104], and clinical teams were informed. The *ACTN1* variant (14:69347630 C>T, p.Ser677Asn) was novel, and therefore determined to be a variant of unknown significance. The *GATA1* variant (X:48650811 C>T, p.Ala227Val) was assigned likely pathogenic despite its novelty, due to the location of the variant in the context of the known structure of GATA-1. Following MDT review a clinically-accredited report was issued and returned to the clinical team.

The *TBXAR2* case (RFH_8E) is yet to be discussed at MDT. RFH_8E is a 46 year old woman of Arabic origin. She has a history of menorrhagia and two caesarean section deliveries were complicated by life-threatening bleeding, requiring transfusion of 8 units of blood on one occasion. Her sister and mother suffered menorrhagia and postpartum haemorrhages. There was no known consanguinity in recent generations. Her light transmission aggregometry tracings showed a poor response to U46619, the thromboxane A2 receptor agonist. Sequencing showed heterozygosity for variant 19:3600434 G>T, p. Leu67Ile, in *TBXA2R*. The variant is novel and absent from the GnomAD database [95]. Cases with a similar phenotype and heterozygous variants at p.Asp42Ser, p.Trp29Lys, p.Asp304Asn and p.Arg60Leu have been published [151-154] alongside functional evidence of loss of receptor function. Three other

non-RFH cases in the cohort also had novel *TBXAR2* variants and were annotated with a U46619 defect, and three more had variants in common with the U46619-annotated cases but were only annotated with platelet function defect as a phenotype. Further enquiry showed that testing with agonists that would detect a TBXA2R had not been performed. These will be performed at the next scheduled clinic visits at their local centres. RFH_8E has also been shown to have a variant of unknown significance in *PRKAR1A* [155], which has been associated with the Carney complex of multiple premature neoplasias. This case presented with renal carcinoma and uterine leiomyomas at the 43 years of age. Her father has metastatic liver cancer.

One index case was withdrawn following investigations that found an acquired cause of bleeding. Two index cases were identified following WES with variants in a novel BPD gene (*ABCC4*) that is the subject of Chapter 5. 97 remaining unexplained index cases were subsequently sent for WGS. No further cases have yet been explained following WGS in the RFH cohort, and analysis is ongoing.

Diagnosis	Number of cases at RFH
Macrothrombocytopenia	4
Storage pool disorder	28
Storage pool disorder and thrombocytopenia	1
Weak agonist response defect (ADP and epinephrine)	41
Weak agonist response defect and thrombocytopenia	3
Other multi-agonist platelet defect	3
Thromboxane A_2 receptor defect	1
Isolated platelet nucleotide release defect	1
Isolated abnormal Impact R	2
Platelet function disorder suspected but inconsistent results	2
Unclassified bleeding disorder	16
Unclassified bleeding disorder with borderline Factor XI levels	1

Table 11 Diagnoses in index cases recruited at Royal Free NHS Trust. All cases recruited had clinical bleeding or a first degree relative with clinical bleeding and a concordant laboratory phenotype.

Patient_ID and phenotype	Status	Genes and zygosity	Var	CADD	GnomAD	MDT result
RFH_1166 Macro-thrombocytopenia	Explained	*ANKRD26* Hemizygous	10:27389382 T>C 5′ UTR variant	12.28	0	Pathogenic: described in multiple previous cases in the literature and this study.
RFH_79L Macro-thrombocytopenia	Possibly explained	*ACTN1* Hemizygous	14:69347630 C>T missense_variant p.Ser677 Asn	17.45	0.000014	Variant of unknown significance: novel variant, frequency of 1:120,000 in ExAC.
RAL_296V X-linked thrombocytopenia with dyserythropoiesis	Possibly explained	*GATA1* Hemizygous male	X:48650811 C>T missense_variant p.Ala227 Val	29.70	0	Likely pathogenic: not previously described in the literature or this study. Not present in the listed control datasets.
RFH_8E Bleeding and thromboxane A2 receptor defect. Also history of multiple neoplasia at ~40yrs	Possibly explained	*TBXA2R* Hemizygous	19:3600434 G>T missense variant p. Leu67Ile	22.7	0	Not processed by MDT yet.
	Possibly explained	*PRKAR1A* Hemizygous	17: 66525123 Coding sequence variant and splice-acceptor variant TAT[31]TGA>T	NA	0	

Table 12 Royal Free cases in whom a variant in a known BPD gene was identified.

3.1.4 Recruitment across the NIHR BR-RD BPD cohort

Recruitment figures by site are shown in Figure 16 as of November 2017. 13,021 cases had been recruited to the NIHR BR-RD project, and of these, 1,213 cases had been recruited to the BPD arm. 687 samples had been sent for WES and 1118 for WGS. For 914 cases detailed phenotypic information had been provided by recruiting clinicians. 33% of the cohort was male and 66% female. Broad phenotypic groups within the cohort are shown in Figure 17, based on HPO coding of cases at recruitment. The largest groups contained cases with thrombocytopenia (27%), unclassified bleeding (24%) and storage pool disorder (14%). A large cohort with BPD and congenital neurological phenotypes was recruited (11%), reflecting the specialist practice at the University of Leuven.

3.1.5 Pan-cohort variant identification and gene discovery

Variants were identified in known BPD genes either at a local level as described above for the RFH cohort, or by central application of statistical genetics [72, 129]. Variants were subsequently labelled as pathogenic or likely pathogenic at the ThromboGenomics MDT in 177 cases (Figure 18). The largest groups of variants were found in *ACTN1* (21%), *ANKRD26* (10%) and *MYH9* (13.9%).

Parallel processing of variants for novel gene discovery was also applied at a local and central level. At a local level I assessed variants in genes known or suspected to influence platelet function but not known to cause pathology, including lists from platelet proteomic analyses and interactors of NBEAL2, which is responsible for the majority of cases of Gray Platelet Syndrome [156]. At a central level a clinical working group was formed to facilitate variant processing including myself, Drs. Claire Lentaigne, Sarah Westbury, Janine Collins and Suthesh Sivapalaratnam at a junior level and Drs. Keith Gomez, Kate Downes, Sofia Papadia and Professors Mike Laffan, Andrew Mumford, Kathleen Freson and Willem Ouwehand at a senior level. The working group identified informative phenotypic groups in whom statistical analyses could be applied by Drs. Ernest Turro and Daniel Greene, using strategies such as manual and automated phenotype analysis [129] and comparison with GWAS studies of platelet traits [111]. The working group subsequently reviewed numerous candidate genes that arose from these analyses on a case-by-case basis to check for concordant phenotypes and supporting literature. Novel discoveries were made using this strategy for genes *SRC* [5], *DIAPH1* [6] and *TPM4* [7], and a novel mode of inheritance was described for *GP1BB [8].* Chapters 4 and 5 of this thesis relate to work following identification of novel candidate genes *KDSR* and *ABCC4* within this multidisciplinary framework.

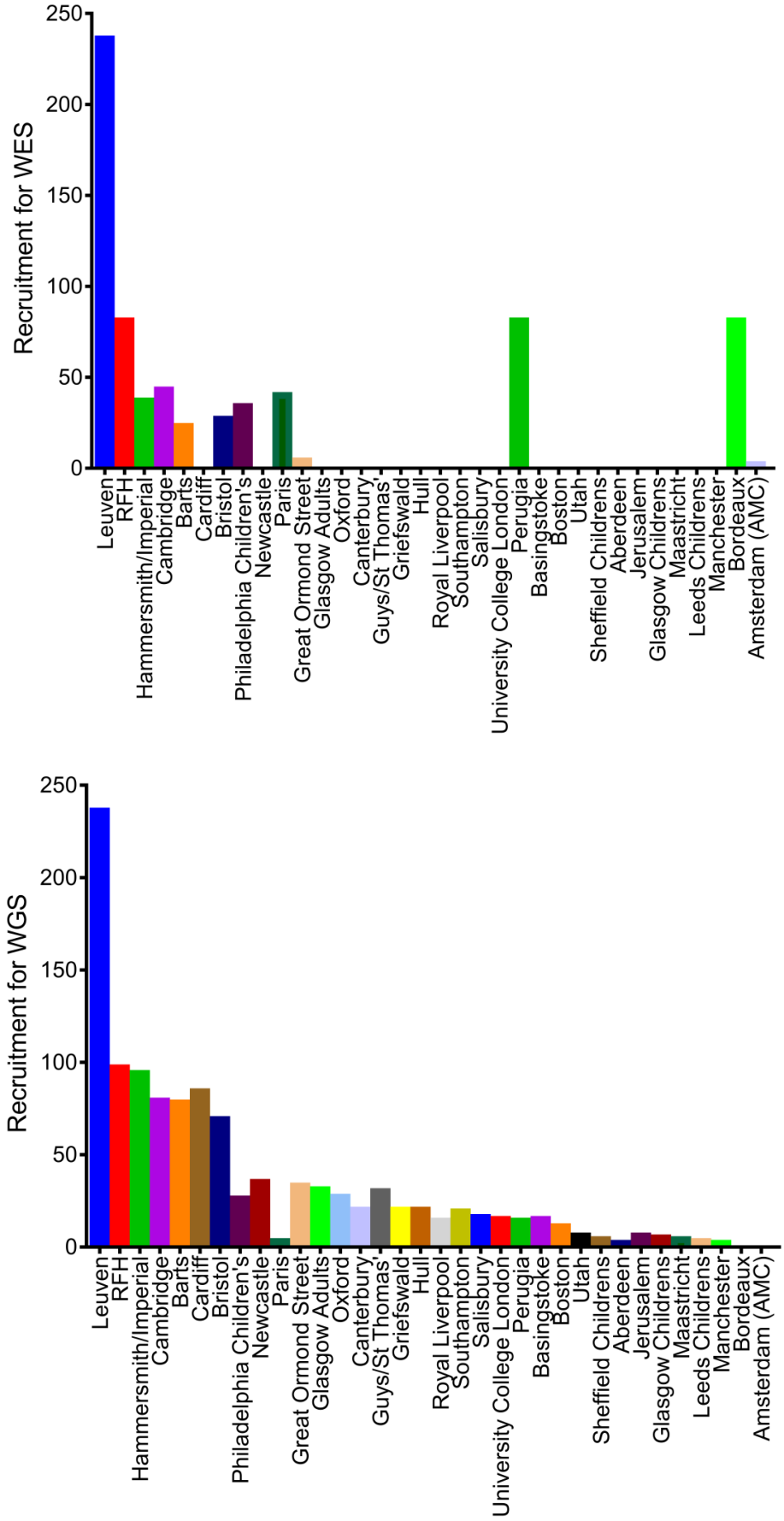


Figure 16 Summary of samples sequenced by WGS and WES by centre for the NIR-RD BPD project as of November 2017.
Data provided by Dr. Sofia Papadia, NIHR BioResource.

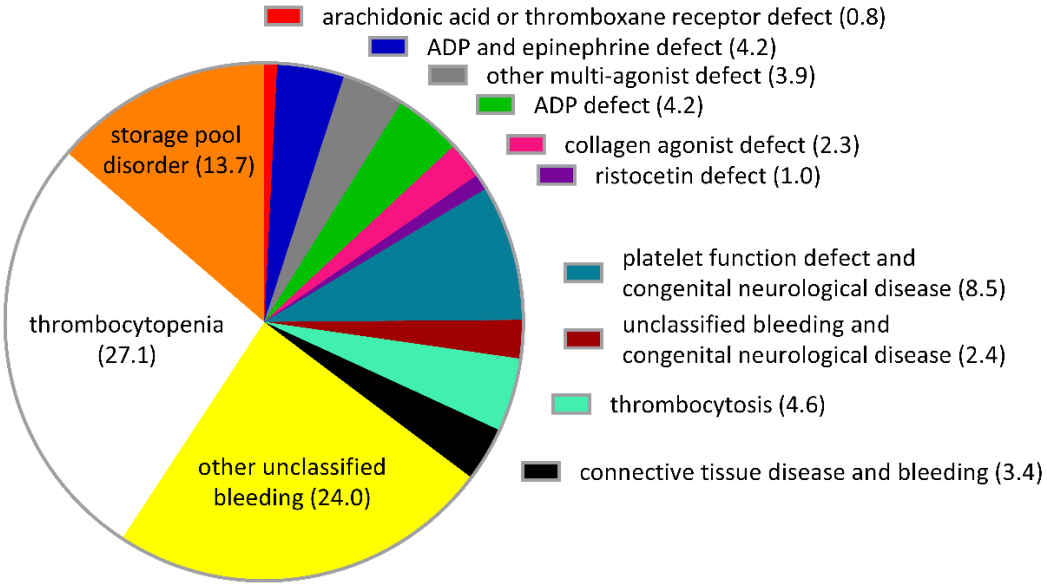


Figure 17 Distribution of diagnoses of 914 cases with platelet disorders or unclassified bleeding in the NIHR BR-RD BPD cohort.
Percentage of cohort given in brackets. Data provided by Dr. Ernest Turro, University of Cambridge and all enrolling centres.

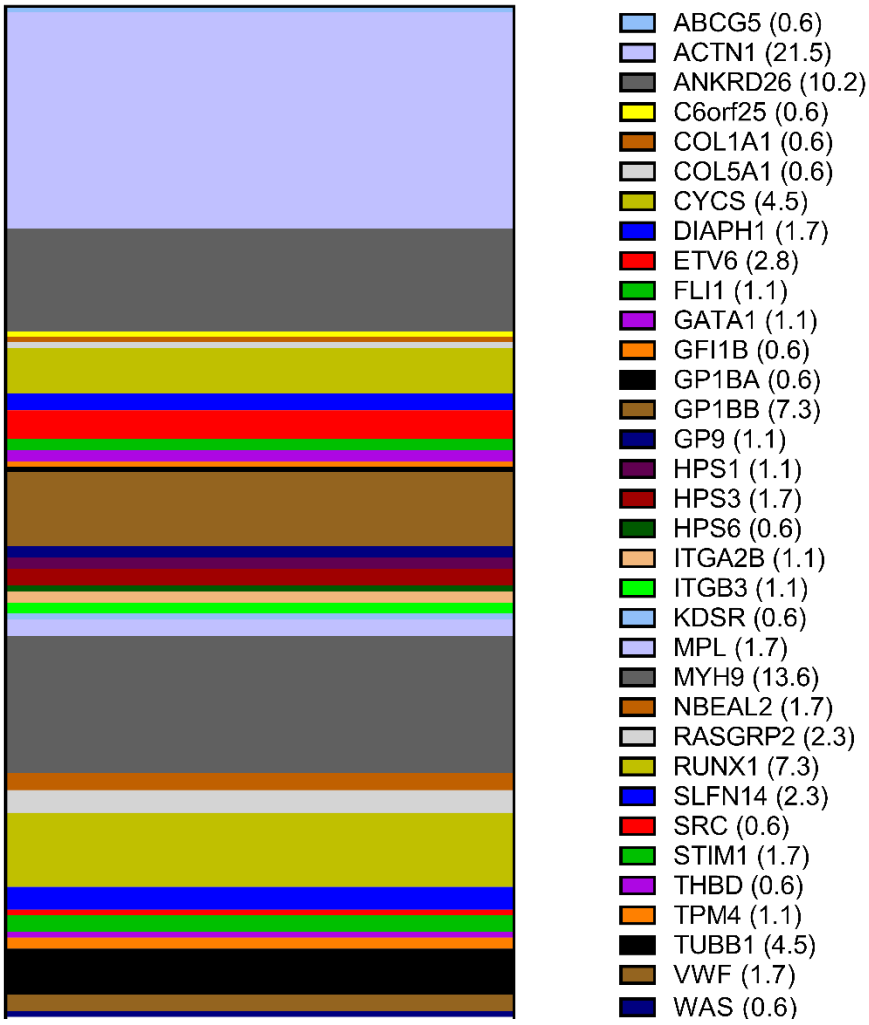


Figure 18 Distribution of variants in known BPD genes in 177 cases with clearly pathogenic or likely pathogenic variants at MDT.

3.2 Percentage of cohort shown in brackets. Data provided by Dr. Karyn Megy, University of Cambridge and the ThromboGenomics MDT [66, 157]. **HPO characterisation of the NIHR BR-RD BPD cohort**

3.2.1 Collaboration note

The definition of new HPO terms related to bleeding and platelet disorders was performed in a working group comprising myself, Dr. Claire Lentaigne, Dr. Sarah Westbury, Dr. Suthesh Sivapalaratnam, Dr. Anne Kelly, Dr. Minke Vries, Dr. Sol Schulman, Prof. Willem Ouwehand, Prof. Kathleen Freson and Prof. Andrew Mumford. The interim assessment of the cohort was performed in a working group comprising myself, Dr. Sarah Westbury, Dr. Ernest Turro, Dr. Daniel Greene, Dr. Claire Lentaigne, Dr. Anne M. Kelly, Prof. Willem Ouwehand, Prof. Kathleen Freson and Prof. Andrew Mumford [72].

3.2.2 Customisation of HPO for BPD

The phenotyping MDT identified 78 terms and associated relationships that were added to the HPO term library [158] in order to maximise the accuracy of HPO coding in our cohort. These codes were predominantly within the abnormality of blood and blood-forming tissue class (Appendix 8.2.1).

3.2.3 Experimental overview

The interim evaluation was performed at the time when 648 index cases and 59 affected pedigree members had been recruited and HPO coded, including use of the newly identified terms. Colleagues in the statistical genetics team of the University of Cambridge, Dr. Daniel Greene and Dr. Ernest Turro, designed an algorithm to test whether cases with rare coding variants in a disease gene would cluster on the basis of their assigned HPO terms. The similarity of HPO terms annotated to individual cases within a defined cluster was assigned a *p* value, indicating the likelihood that these terms would have been this closely located in the HPO 'tree' by chance.

3.2.4 Distribution of HPO terms in the cohort

The median number of HPO terms attached to each case was 7.5 (range 1-23) (Figure 19a). 387 (59.7%) index cases were annotated with HPO terms involving non BPD-specific terms. The commonest non-BPD terms were neurological (149 index cases), immunological (109 index cases) and skeletal (102 index cases) (Figure 19c). The analysis was repeated at the end of recruitment with 2,946 cases and the findings were similar (Figure 19d). These findings suggest that a large proportion of BPD in our cohort are syndromic disorders, particularly involving neurological phenotypes. This may in part reflect the specialist interest in syndromic disorders of recruiting physicians at the University of Leuven, which limits the extrapolation of conclusions about phenotypic heterogeneity to undiagnosed BPD in general.

3.2.5 Cases from the same pedigree cluster by phenotype

50 pedigree groups among the interim cohort were analysed by the clustering algorithm, under the hypothesis that affected members of a pedigree should share causal variants. 40 pedigrees had a phenotypic similarity *p* value of <0.05 (Figure 19c). Under the null hypothesis, of no similarity between pedigrees by HPO phenotype, only 2.5 pedigrees would be expected to have $p<0.05$. Four pedigrees

showed clustering *p* values >0.2. This was explained in one pedigree by detailed phenotyping in the index and limited coding in the affected family members. In another case the similarity scores were biased by gender-specific terms including menorrhagia and post-partum haemorrhage. These issues were accounted for in subsequent versions of the algorithm, prior to application of clustering algorithms to the whole cohort.

3.2.6 Cases enrolled by provisional syndromic diagnosis cluster by phenotype

Several centres recruited cases with provisional syndromic diagnoses, defined by phenotypic similarity and suspected to have a common genetic basis, including Gorham-Stout syndrome, Hermansky Pudlak syndrome, pseudohypoparathyroidism type Ib, Roifman syndrome and Wiskott-Aldrich syndrome. Unrelated index cases in each of these groups clustered together with $p<0.05$ for each group (Figure 19c). However, once again a large proportion of these cases came from the University of Leuven, where enrolling clinicians may have a tendency to use similar coding terms, limiting the applicability of these findings to undiagnosed BPD in general.

3.2.7 Cases with rare variants in a disease-causing gene (*ACTN1)* cluster by phenotype

During the course of this study variants in *ACTN1* were reported to be associated with macrothrombocytopenia [27, 159]. Review of our cohort identified 21 index cases and three family members with rare *ACTN1* variants. 22/24 cases had macrothrombocytopenia. In this cohort the clustering algorithm provided a phenotypic similarity value of $p = 3.4 \times 10^{-8}$ (Figure 19c).

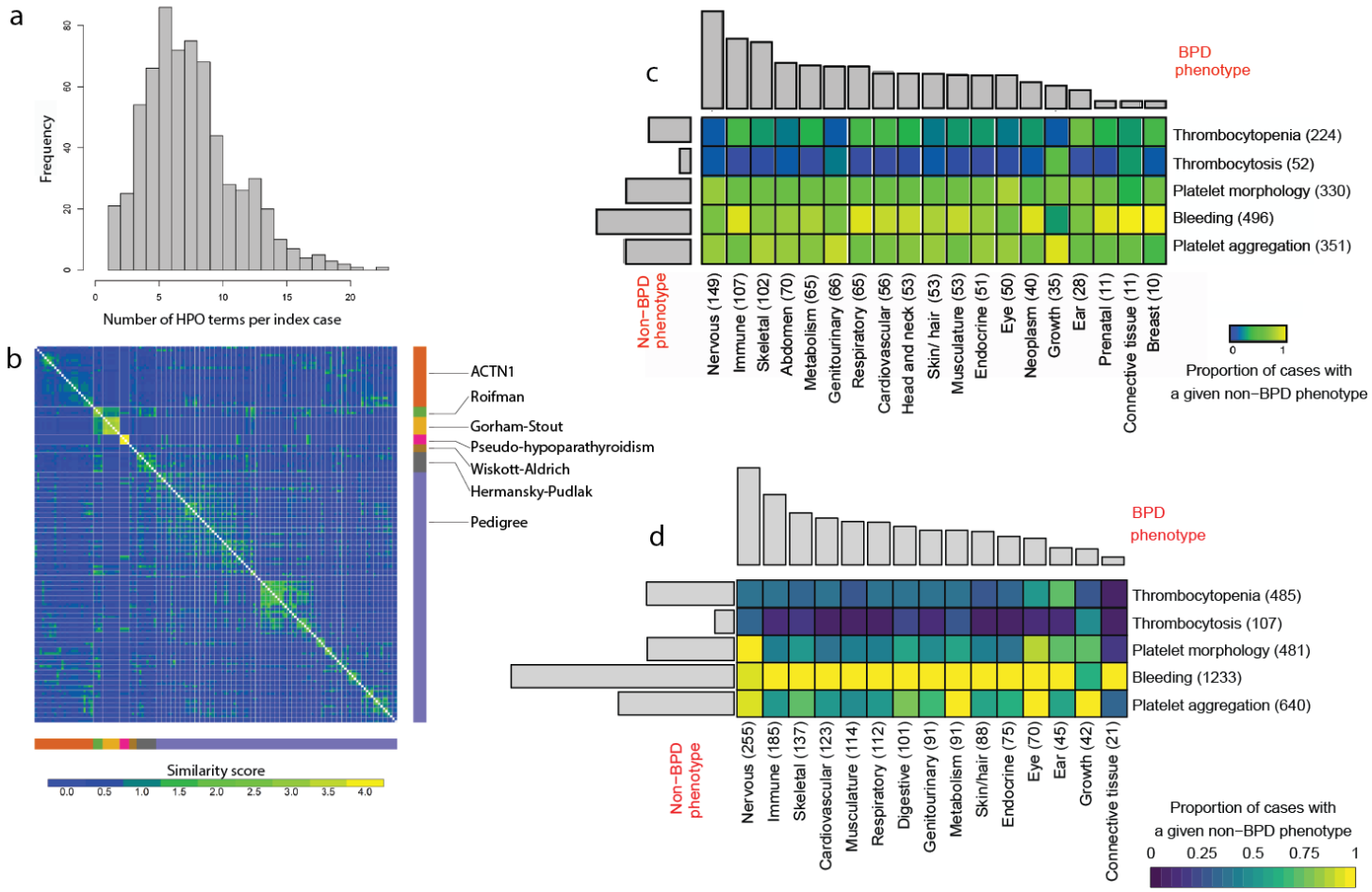


Figure 19 HPO application to an interim cohort for validation (a-c) and the final cohort at the time of writing (d).
(a) Number of HPO terms per index case shows a median number of terms of 7.5. (b) Clustering of cases using a similarity algorithm by pedigree, provisional syndromic diagnosis and ACTN1 variants. (c) Heat map showing non-BPD phenotypes mapped against BPD phenotypes. The phenotypic diversity was rich in the cohort. All figures reproduced with permission from Westbury *et al.* [72] and originally produced by Dr. Ernest Turro and Mr. Daniel Greene.

3.3 Interim review of bleeding scores in the NIHR BR-RD BPD UK cohort

3.3.1 Aims

A major limitation of the HPO is that it does not code severity of phenotype. Bleeding scores have been developed to improve accuracy and objectivity in the assessment of bleeding type and severity. This is important to our study because inherited BPD are usually mild, and therefore present a greater challenge in differentiating a significant bleeding history from the bleeding symptoms commonly reported in normal individuals [76]. A recent study had shown that bleeding scores are not predictive of the presence of a platelet defect on lumiaggregometry in patients with suspected inherited PFD [84], and therefore cannot be used as a screening test. We sought to evaluate the use of bleeding scores in our cohort including:

1. The score cut-off to maximize the recruitment of affected individuals and minimize recruitment of unaffected individuals
2. Differences in the bleeding phenotype between PFD and UBD
3. The most discriminant elements of the bleeding score
4. Correlation between the bleeding score and global measures of platelet-dependent haemostasis
5. The level of inter-observer variability in scoring

3.3.2 Experimental overview

Cases were identified from the cohort of index cases in the NIHR BR-RD BPD study from four UK Haemophilia Centres: Royal Free London NHS Foundation Trust, Imperial College Healthcare NHS Foundation Trust, Addenbrooke's NHS Foundation Trust and University Hospitals Bristol NHS Foundation Trust. The bleeding score was calculated using the MCMDM-VWD BAT tool at face-to-face interview in cases and controls. Detailed methods are described in section 2.5.6.

3.3.3 Cohort description

192 participants were available for the final analysis: 145 patients (107 PFD, 38 UBD; M:F ratio 25:120, median age at diagnosis 34 years) and 47 controls (M:F ratio 15:32, median age at recruitment 40). Of patients with PFD, 39/107 (36%) had low nucleotide levels, 45/107 (42%) had defects involving ADP +/- epinephrine, 17/107 (16%) had defects involving multiple agonists other than ADP, 4/107 (4%) had defects of a single agonist other than ADP, and 2/107 (2%) had an isolated defect of ATP release in response to ADP on lumi-aggregometry with normal light transmission aggregometry results.

3.3.4 Bleeding scores are higher in UBD than PFD

As expected, the bleeding score was significantly higher in cases compared with reportedly healthy controls, including when sub-stratified by diagnosis (PFD or UBD), gender (male or female) and age (<50 years or ≥50 years), (p<0.001 for all conditions). In cases with PFD the median score was 8 (range 2-24). In cases with UBD the median score was 10 (range 4-34). Cases with UBD had a significantly higher

bleeding score than cases with PFD (p =0.015). One female control had a bleeding score of 8 although she reported being in good health and confirmed that she had no known bleeding disorder. Subsequent investigations showed that this case was being investigated for multiple organ dysfunction, suggesting a possible acquired bleeding disorder, demonstrating a key limitations of the BAT in its inability to distinguish acquired and congenital bleeding.

	No. of recruits	Median score	p-value
Total cases	145	9	p<0.001 for all conditions
PFD cases	107	8	
UBD cases	38	10	
Female cases	120	9	
Female PFD	87	8	
Female UBD	33	10	
Male cases	25	9	
Male PFD	20	9	
Male UBD	5	6	
<50years cases	115	8	
<50 years PFD	91	8	
<50 years UBD	24	10	
50+ years cases	30	10	
50+ years PFD	16	9	
50+ years UBD	14	11.5	
Total controls	47	0	-
Female controls	32	0	-
Male controls	15	0	-
<50 years controls	32	0	-
50+ years controls	15	-1.0	-

Table 13 Summary of study population demographics, median bleeding scores and Wilcoxon rank-sum test of differences between cases and age/sex-matched controls.

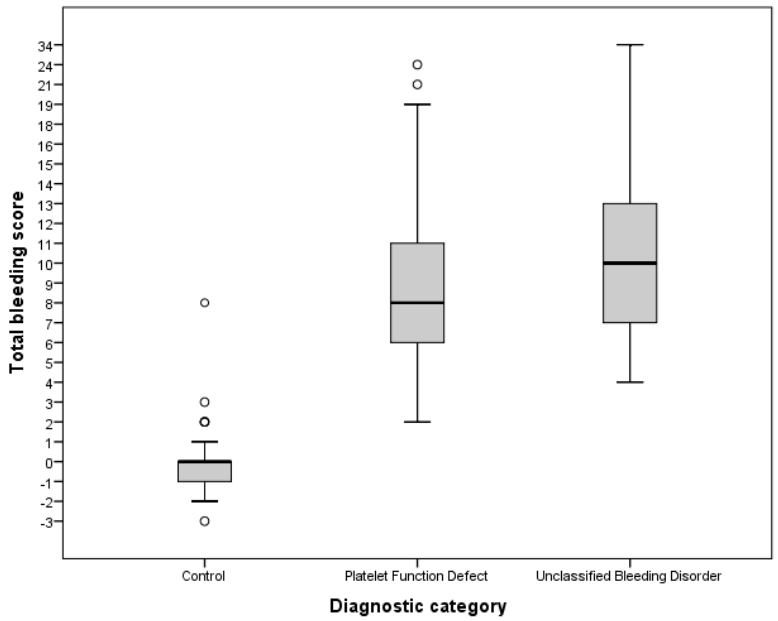


Figure 20 Bleeding scores in study subgroups.
Shaded area represents interquartile range. Thick horizontal lines represent median scores. Circles represent outliers.

3.3.5 The optimal score cut-off is ≥ 3 for PFD

A score of ≥4 was shown in a previous study [79] to have a negative predictive value (NPV) of 100% for type 1 VWD. In the PFD cohort a score of ≥4 had a NPV of 86.8% and a positive predictive value (PPV) of 99%. Figure 21 shows that the optimal score cut-off to maximise the PPV whilst minimizing the NPV is ≥3, with an associated PPV of 99% and NPV of 90%.

3.3.6 The bleeding score does not correlate with PFA-100 or Impact-R results

Results of PFA-100® and Impact-R® screening tests were evaluated in cases with PFD and UBD (Figure 22). The majority of cases' results were in the normal range for both tests (70.2% and 83.9% with PFA-100® epinephrine and ADP respectively; 50% and 89.8% of Impact-R®SC and AS respectively). There was no correlation between bleeding score and the PFA-100® closure time (CT) using ADP or epinephrine cartridges, or the Impact-R® surface coverage (SC) or aggregate size (AS) in either PFD or UBD. Cases with PFD had higher PFA-100 epinephrine cartridge closure times than cases with UBD (mean closure time 171 seconds for PFD vs. 143 seconds for UBD; p =0.04, Figure 22). There were no significant differences between the groups' results on PFA-100 ADP cartridge, Impact-R SC or AS.

3.3.7 Bleeding symptoms are similar in UBD and PFD cases

Amongst the PFD and UBD patient group, the most common symptoms captured by the BAT were cutaneous bleeding, bleeding from minor wounds, bleeding after tooth extraction and surgery, and menorrhagia (Figure 23). Using Fisher's exact test there was no significant difference between the distribution of bleeding symtoms in cases with PFD compared to UBD (p-values for all symptom categories >0.01).

3.3.8 Cutaneous, dental and surgical bleeding symptoms are discriminant for PFD

To assess the association between specific symptoms (defined as a bleeding score >0 in a symptom category) with the presence of PFD , odds ratios and 95% confidence intervals of an underlying PFD for each bleeding symptom captured on the BAT were calculated (Figure 24). Cutaneous bleeding, tooth extraction and post-surgical bleeding gave the highest odds ratios, suggesting these are highly discriminant features of the bleeding score. However, because an aforementioned correction was applied to compensate for zero cases in some categories, the discriminant power of symptoms that are very rare in the control population will be underestimated, specifically CNS symptoms, haemarthrosis and muscle haematomas.

3.3.9 Inter-observer variability in scoring is low

For 15 cases at one centre, the MCMDM-1 VWD bleeding score was calculated at the time of the initial referral for evaluation of excessive bleeding by a haemostasis expert not involved in collecting scores for the current study. A paired-sample Wilcoxon signed-rank was applied to test if the mean difference between the historical scores and the scores calculated retrospectively in the main study was zero. There was some tendency for the historical scores to be less than those collected in the main study, but the difference between the scores was not clinically or statistically significant (mean difference 0.6 points,

p =0.084). A Bland-Altman plot (Figure 25) demonstrates moderate agreement with a maximum difference of 3 points. The difference appears to be independent of the mean, suggesting that the difference is not due to a systematic error such as one operator consistently scoring each symptom higher than the others.

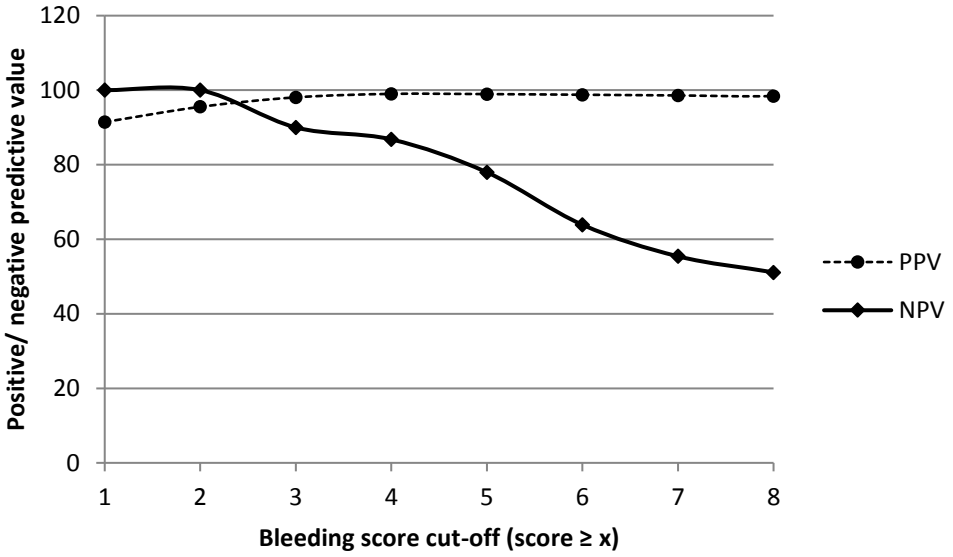


Figure 21 Negative and positive predictive values (NPV and PPV) associated with a range of bleeding score cut-offs in cases with platelet function disorders (PFD).

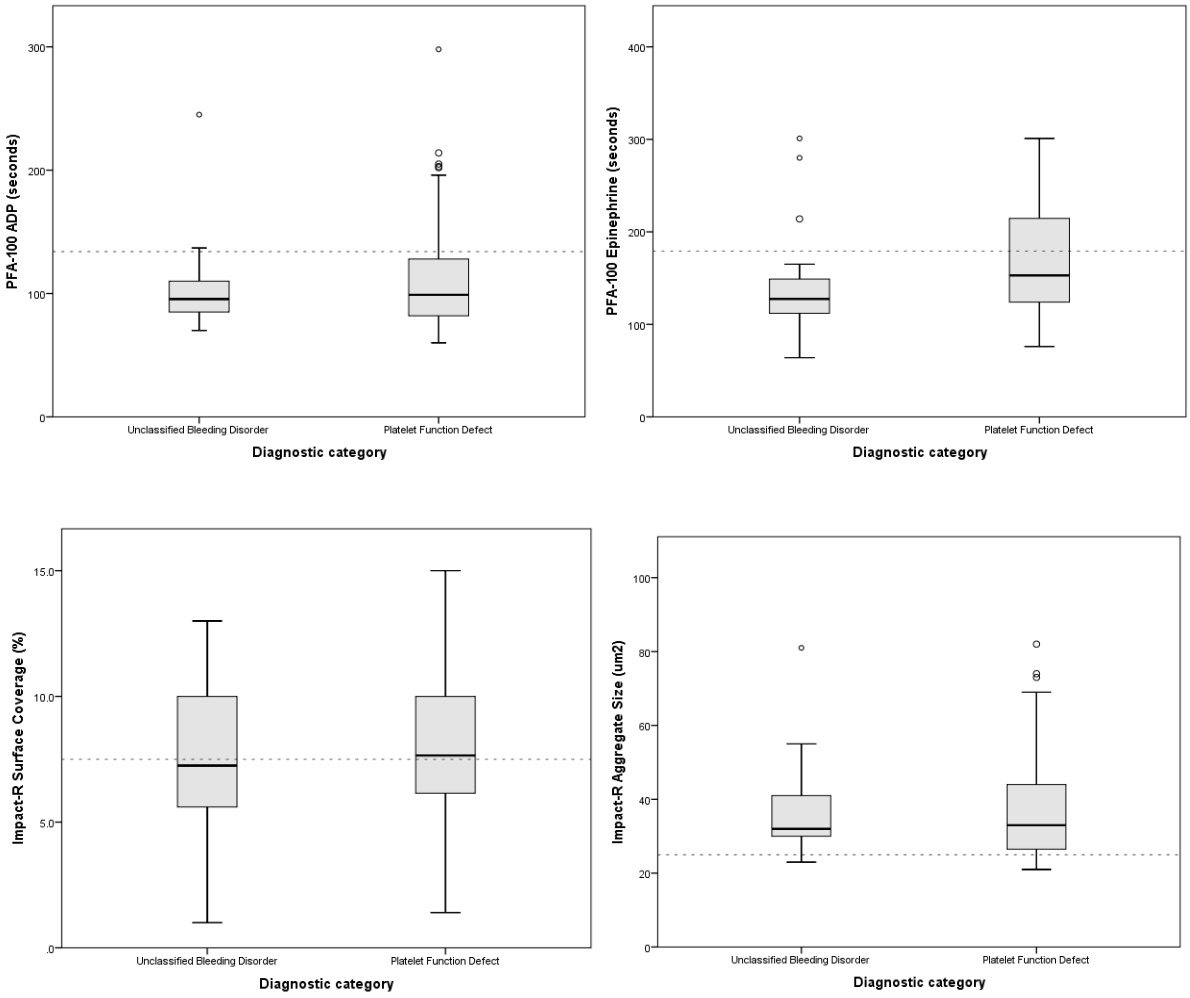


Figure 22 Box and whisker plots showing PFA-100 and Impact-R parameters in cases with platelet function disorders (PFD) and unclassified bleeding disorders (UBD).
Dotted lines denote manufacturers' reference ranges (PFA-100® ADP cartridge <134 seconds, PFA-100® Epinephrine cartridge <179 seconds, Impact-R® SC >7.5%, Impact-R® AS >25µm^2). Shaded area represents interquartile range. Thick horizontal lines represent mean scores. Circles represent outliers.

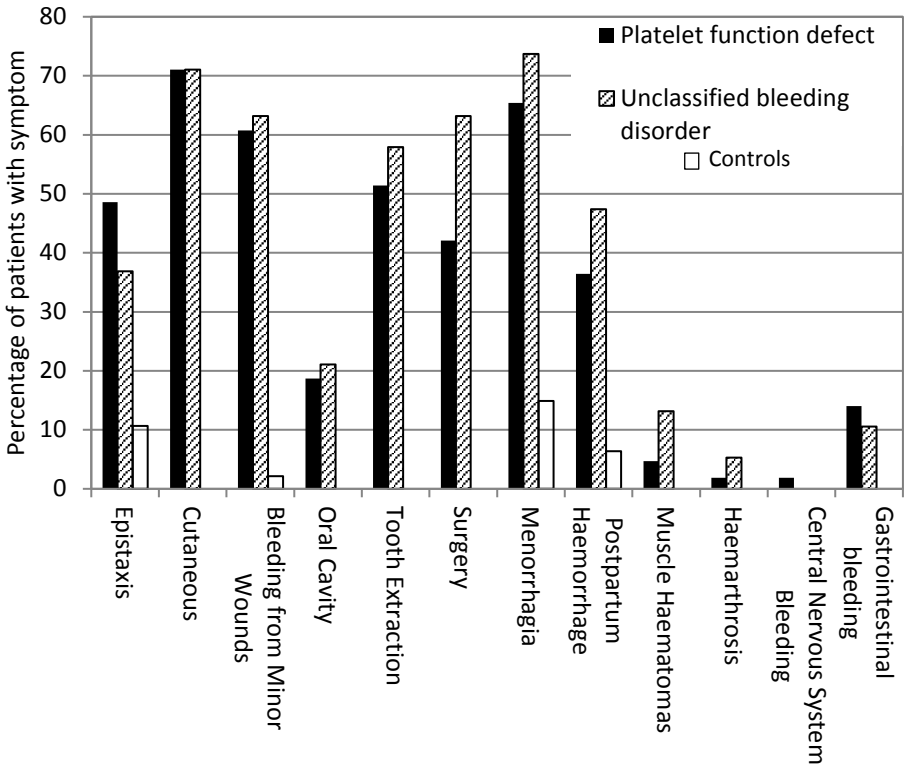


Figure 23 Assessment of variability in symptom categories between cases with platelet function disorders (PFD), unclassified bleeding disorders (UBD) and controls.

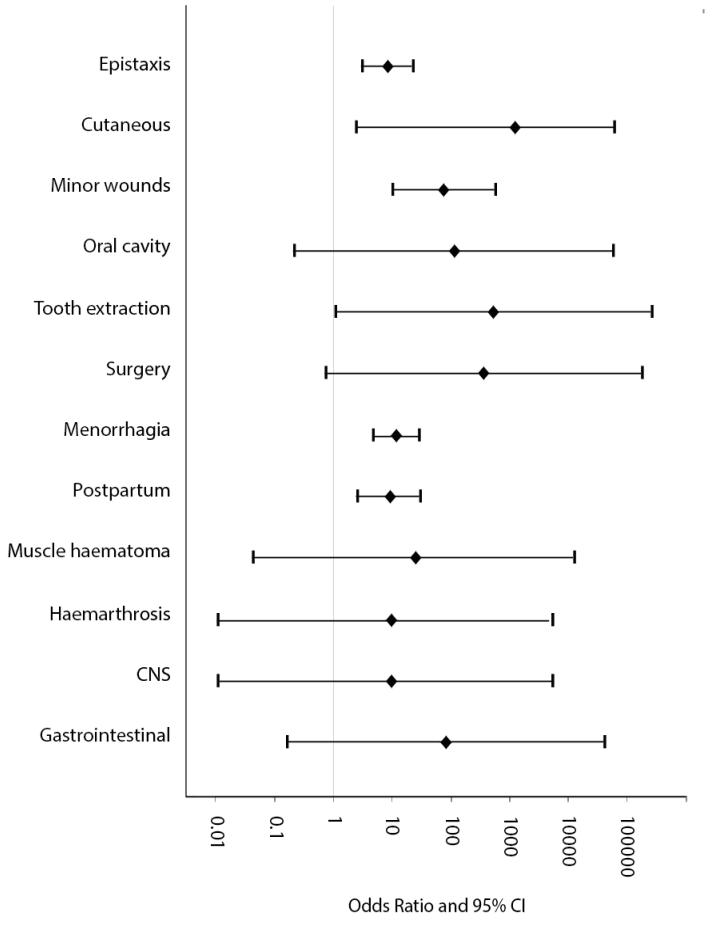


Figure 24 Odds ratios for specific symptom categories in cases with platelet function disorders (PFD) compared with controls. Diamond markers denote absolute odds ratios. Whiskers denote 95% confidence intervals.

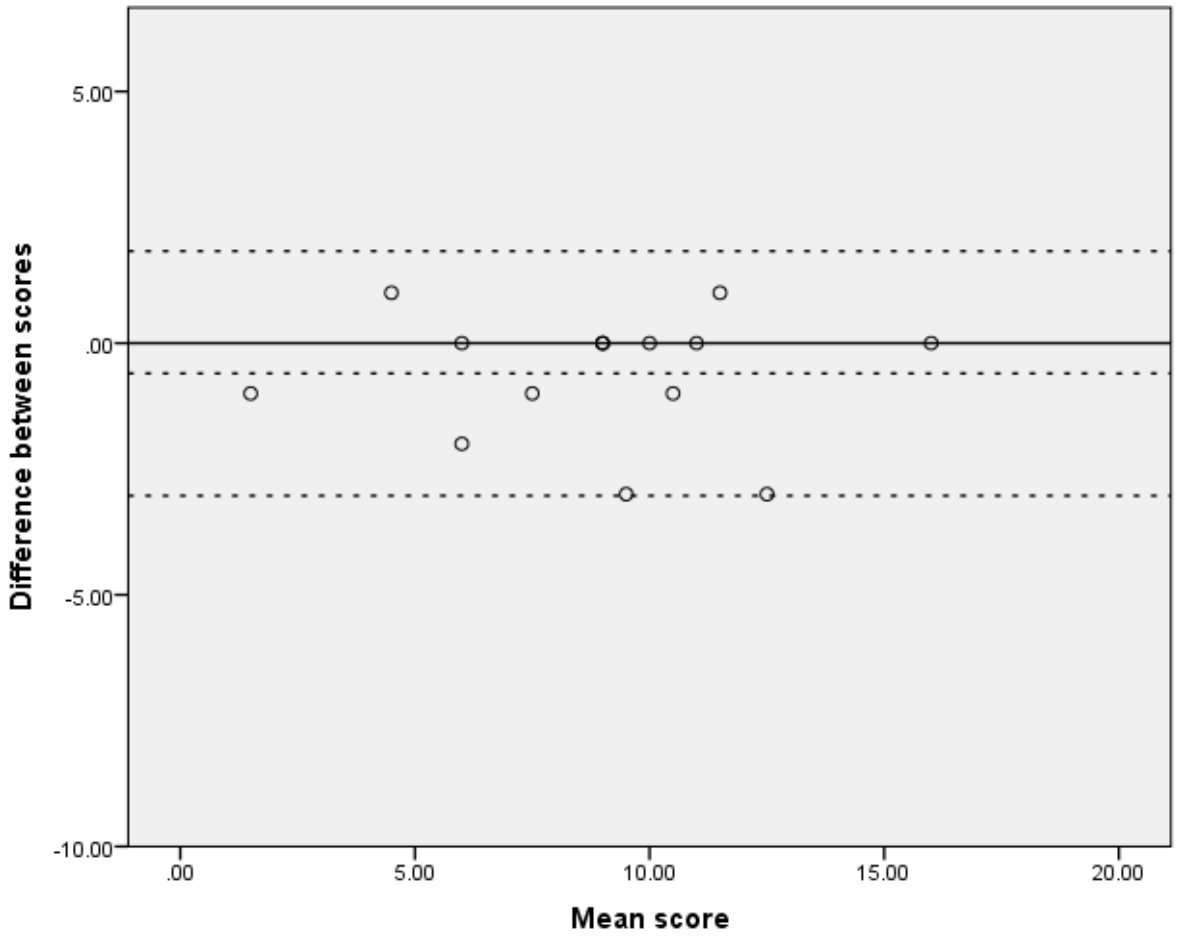


Figure 25 Bland-Altman plot demonstrating moderate inter-observer agreement.
Interrupted lines denote mean difference and 95% limits of agreement.

3.4 Discussion

The findings of this chapter suggest that the NIHR BR-RD BPD study has performed well in terms of recruitment for sequencing and dissemination of data. To facilitate research WGS data has been deposited in the European Genome-phenome Archive (EGA, https://www.ebi.ac.uk/ega/). The implementation of a strategy to detect and feedback medically actionable results in a clinically accredited format to recruiting clinicians has resulted in 177 new genetic diagnoses. In addition to the publication of novel discoveries (SRC [5], DIAPH1 [6] and TPM4 [7]), several manuscripts have been published [8-10] or are in preparation describing the genotype-phenotype analyses in large cohorts of cases with variants in known BPD genes, especially where the number of published cases is relatively low. Further genes are under evaluation by co-segregation, extended cellular phenotyping and cellular modelling, including *ABCC4* and *KDSR* which are discussed in Chapters 4 and 5.

The local analysis of the vast number of sequence variants per case required filtering of variants to be assessed. A minor allele frequency of <1/1000, given the rarity of the diagnoses we are assessing, is probably generous, and most penetrant variants of interest are likely to have a MAF at least tenfold lower or absent from reference cohorts. However penetrance of milder bleeding disorders may be variable. Presentation usually follows significant haemostatic challenges which may be absent, for example in young males, leading to inappropriate inclusion in databases of ostensibly 'healthy' individuals. Therefore a more generous limit was selected. The use of rarity in databases of 'normal' variation to infer likelihoods of pathogenicity is also flawed when cases are non-Caucasian due to the lack of adequate representation of these groups in research studies. At RFH 36/103 index cases were non-Caucasian.

Relatively few variants were found in the RFH cases in known genes. This reflects the strong emphasis on clinical genetic diagnosis taught and practiced at RFH prior to the NIHR BR-RD BPD study, at a time when the mainstay of diagnosis was Sanger sequencing of a gene of interest. As a result cases recruited were truly unexplained. The variants that were detected at RFH raise important ethical issues about the interpretation of genetic variants for clinical purposes. Pathogenic *ANKRD26* variants are associated with a 10% risk of acute leukaemia [104]. The screening for common hereditary cancer variants such as *BRCA1* appears to be of great benefit. However the case in rare disorders is less clear; the risk is serious, probably low, and often unknown if few cases have been published or penetrance is variable. Inevitably the answer is pre-emptive, to carefully counsel individuals prior to genetic tests for research purposes to ensure understanding of the test, appropriate expectations regarding results and ascertain whether feedback is desired. Should feedback be chosen, it is unclear how screening should be undertaken, to minimise the psychological impact of cancer risk on the individual yet maximise efficacy in detection.

The results of the assessment of HPO as a tool for gene discovery demonstrate that phenotypic similarity is quantifiable using HPO terms, and can therefore be used to prioritize novel candidate genes causing BPD. This approach was validated in pedigrees, unrelated cases with provisional clinical diagnoses of syndromic BPD and cases with *ACTN1* variants and macrothrombocytopenia. The experiment also identified areas of improvement for the algorithm through focused analysis of pedigrees assigned low phenotypic similarity values, to account for differences in the depth and gender-specificity of coding prior to application to the whole cohort. As proof-of-principle, the HPO clustering algorithm was used in both the *SRC* [5] and *DIAPH1* [6] discoveries.

However, despite the standardized format of HPO, it proved insufficient in isolation due to the lack of standardization between recruiting centres in terms of the historical clinical work-up of BPD cases. This was exemplified by the *TBXA2R* variants in the cohort; the same mutation was evident in several cases with the common phenotype of bleeding. However only a subset of cases was annotated with a specific, compatible U46619 agonist defect. Based on HPO coding alone, statistical algorithms cannot differentiate between a normal test or a test that had not been performed, therefore additional, more detailed laboratory phenotypes were sought from recruiting centres.

Even using the standardized format of the HPO, the quality and quantity of clinical coding between centres was variable. Much time was spent by the working group chasing, checking and correcting phenotypes and pedigree information. This common drawback of the multicentre approach that was imperative for successful recruitment was likely mitigated, but not resolved by the application of HPO.

Since HPO cannot code severity we evaluated bleeding scores in an interim cohort for this purpose. The comparison of bleeding scores in cases with PFD and UBD must be interpreted with caution because these are both heterogeneous groups of disorders. Taking this caveat into consideration, the data suggest that PFD and UBD are phenotypically similar disorders, principally characterised by mucocutaneous bleeding. That cases with UBD have higher scores than patients with PFD should also be viewed with skepticism as in the absence of laboratory abnormalities, a higher threshold of clinical symptoms may be required to make this entirely clinical diagnosis. This supports the findings in a recently published paper [84] that a high bleeding score is not predictive of a demonstrable PFD, because the referral population may have included patients that would have been classified in this study as UBD. The diagnostic criteria for UBD are not clearly defined, and there may be value in using the bleeding score in a future consensus definition.

Application of the MCMDM-1 VWD score in this study allowed comparison of scores obtained at different time points by different clinicians at a single institution (Royal Free London NHS Trust). No statistically or clinically significant difference was found between historical and retrospective scores

collected, however the number of cases with multiple scores was small. This suggests that scores contributed by different individuals to the study may be comparable.

The condensed MCMDM-1 VWD score was evaluated in this study because at the study outset it was the best validated score. Subsequently the ISTH BAT has been published, which has some differences in how individual symptoms are scored, including a narrower score range from 0-3. In a head-to-head comparison [160] neither score format was clearly superior in the assessment of Type 1 VWD. The authors felt that the MCMDM-1 VWD score was slightly more sensitive to lab parameters that they felt correlate with bleeding tendency in VWD type 1 (VWF:Ag and FVIII:C). The addition of negative scores to demonstrate absence of bleeding in response to haemostatic challenges has the potential to add useful detail to the score. The relatively high odds ratios for tooth extraction and post-surgical bleeding suggest that these symptoms are particularly discriminant. This indicates that the absence of bleeding following these challenges may make a bleeding disorder less likely. There is therefore some advantage in having a negative score attached to the absence of bleeding in some symptom categories. However, the statistical correction made to allow calculation of odds ratios where there were zero controls with symptoms is likely to underestimate the discriminant power of the rarest symptoms, CNS bleeding, haemarthrosis and muscle haematomas.

There are clearly limitations to the general application of bleeding scores. Personality, education, lifestyle, socioeconomic group, family history of bleeding and the experience and interest of the attending physician in bleeding symptoms may affect the perception of severity, and therefore the score. They are, by design, a retrospective review of historical events, even if used at the point of first referral. The time elapsed since bleeding occurred and its severity may also affect reporting. Our study may suffer bias towards recruitment of more severely affected patients. These patients may have a strong link with the haemophilia centre making them more likely to accept an invitation to a face-to-face interview. Further bias is likely introduced by the use of ostensibly healthy controls; the study design could have been improved by the use of cases referred for bleeding that ultimately received neither a diagnosis of UBD nor PFD as controls, especially if performed prospectively. On the other hand, the external validity of these findings is improved by the use of a multicentre approach and high patient numbers. I used strict and explicit inclusion and exclusion criteria and laboratory diagnostic criteria to maximise the uniformity in the study population, which comprised a representative cross-section of specific platelet defects. In summary, limited conclusions can be made from this data and therefore the bleeding score was considered only as an adjunct to other phenotypes following candidate gene nomination by other means.

Chapter 4: KDSR

4 KDSR

4.1 Discovery of *KDSR* as a candidate gene

In 2015 a single case (IPD123) was identified in the NIHR BR-RD cohort with a rare, severe phenotype of neonatal-onset thrombocytopenia and juvenile myelofibrosis. Exome sequencing in a single case revealed novel compound heterozygous variants in *KDSR,* a gene that was not, at that time, known to influence haematopoiesis. A search for similar cases in our cohort yielded no other cases. The work described in this chapter was commenced to explore the hypothesis that the *KDSR* variants were indeed responsible for the phenotype, in the knowledge that with a single case a greater burden of proof is required to establish causality.

In 2017 *KDSR* mutations were implicated in severe skin pathology and shortly thereafter associated with severe thrombocytopenia in a small number of cases [3, 4]. No mechanistic exploration of the role of KDSR in haematopoiesis has been published to date.

The primary objective in this chapter is to evaluate the hypothesis that the observed *KDSR* variants in the proband cause thrombocytopenia. The secondary objective is to develop and test new hypotheses related to the mechanism of action.

Objectives:

- To explore and expand the clinical and laboratory phenotype of the pedigree
- To evaluate the variant effects in the context of the known structure, function and associated pathobiology of *KDSR.* This will include:
    - determination of the metabolic profile of plasma samples from the proband to determine whether there is loss of function
    - development of cellular models to investigate the hypothesis
    - evaluation of whether directed differentiation of patient-derived stem cells and reprogramming of patient-derived induced pluripotent stem cells (iPSC) recapitulates the phenotype
    - use of lentiviral expression of the normal open reading frame of *KDSR* in patient-derived iPSC to determine whether the in vitro phenotypes observed are attenuated
    - interpretation of the results of zebrafish *kdsr* knockdown performed in collaboration on thrombopoiesis

4.2 *KDSR*: structure, function and pathobiology

3-Keto-dihydrosphingosine reductase (KDSR) is an early, essential enzyme in the pathway of de novo sphingolipid synthesis (Figure 1) that catalyses the conversion of 3-ketodihydrosphingosine (KDS) to dihydrosphingosine (DHS) on the cytosolic leaflet of the endoplasmic reticulum (ER) [161]. The canonical transcript encodes a 332 amino acid, 36,187 Da protein. It is highly conserved [162, 163] and widely

expressed [103, 161], consistent with the integral roles of the sphingolipid family in forming lipid rafts that facilitate membrane trafficking and in the regulation of fundamental cellular functions that include apoptosis, differentiation and proliferation [164]. The importance of sphingolipid synthesis for normal cellular functioning is illustrated by the complex multisystem phenotypes of mice that are null for key enzymes or receptors in the pathway, including defective platelet formation, activation and thrombus formation [165, 166]. Sphingosine-1-phosphate (S1P) is synthesized downstream of KDSR and its reduction in sphingosine-kinase 1 knockout mice has been shown to inhibit platelet activation and aggregation without evidence of thrombocytopenia. Conversely, in mouse studies of sphingosine-kinase 2 and S1P receptor knockout, severe thrombocytopenia was observed and attributed to impaired proplatelet extension into bone marrow sinusoids and shedding [34, 167]. Dysregulation of the sphingolipid pathway has also been shown to influence platelet number and function in humans. Gaucher's disease is an autosomal recessive lysosomal storage disorder characterised by the accumulation of intracellular glycosphingolipids and is associated with thrombocytopenia and abnormal platelet function, although the pathobiology is as yet unexplained [168]. The de novo synthetic pathway has been shown to be active in megakaryocytic cell lines but plays a minimal role in mature platelets, which instead generate essential sphingolipids by incorporation from plasma or recycling of plasma membrane sphingomyelins, largely independently of KDSR [169].

Consistent with these important roles, loss of function variants in *KDSR* have recently been identified as causal of the severe skin disorder erythrokeratodermia variabilis et progressiva 4 (EKVP4, MIM617526), a condition characterised by neonatal onset of well-demarcated, thick, scaly skin on the face and genitals and milder erythematous palmo-plantar scaling [4]. Systemic retinoid therapy led to near-complete resolution in two of two treated probands. This observation established a role for *KDSR* in the homeostasis of keratinisation. It is unclear whether there were haematological abnormalities in this cohort. A subsequent study described EKVP4 cases with compound heterozygous variants in *KDSR,* severe thrombocytopenia and platelet dysfunction in infancy [3]. Metabolic profiling demonstrated reduced downstream sphingolipids S1P and ceramide in plasma and reduced ceramide levels in skin. Bone marrow morphology in one case was normal and in a second case demonstrated increased megakaryopoiesis leading to a diagnosis of immune-mediated thrombocytopenia and subsequent splenectomy with minimal improvement in platelet count. No further experimental exploration of the mechanism of thrombocytopenia was undertaken.

KDSR may have haemato-oncogenic properties; it is translocated with BCL2 in variant follicular lymphoma [170] and is differentially expressed in subtypes of primary effusion lymphoma [171] and diffuse large B-cell lymphoma [172], though the mechanism of such an effect remains unclear. A homozygous missense mutation in *KDSR* has also been associated with severe bovine neurodegeneration [173].

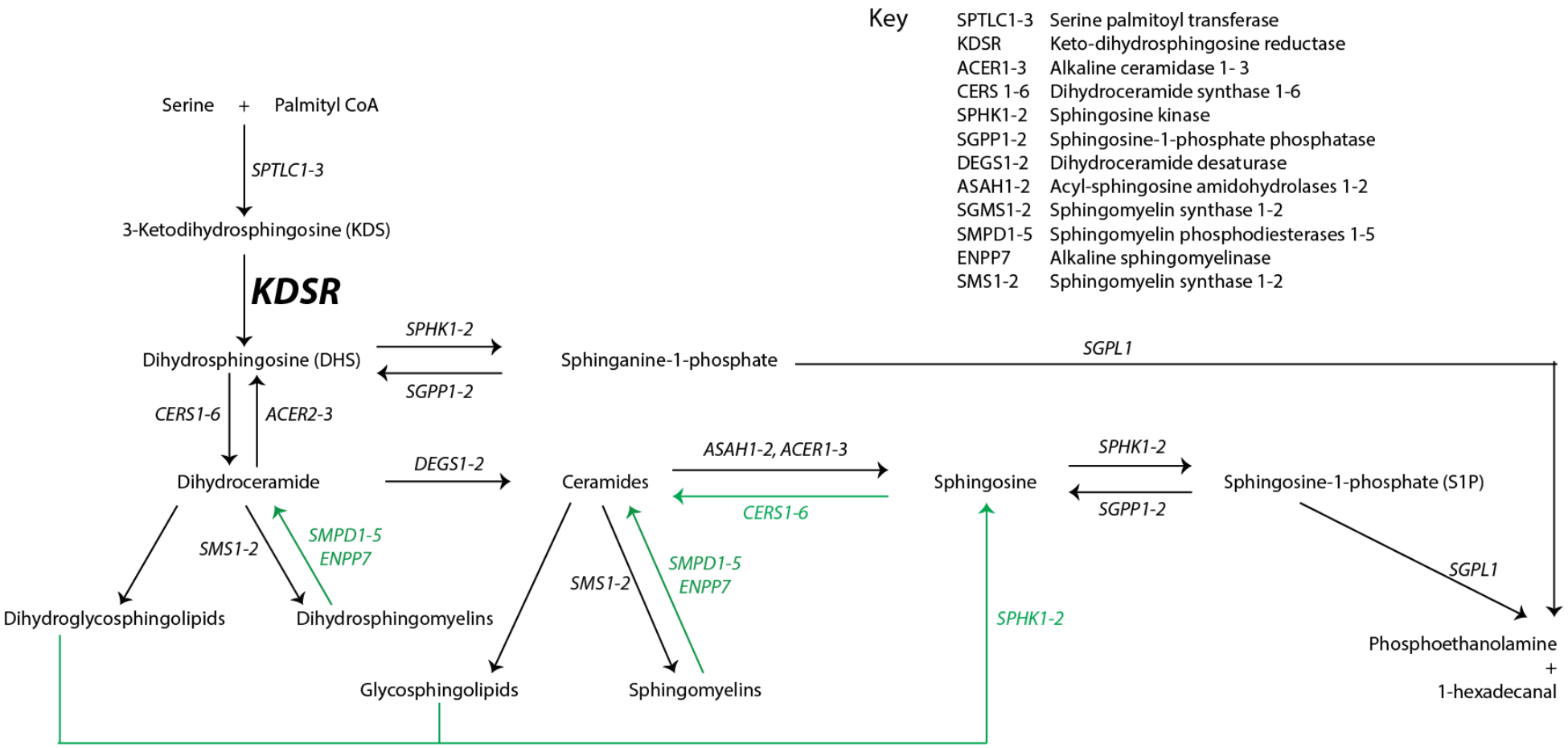


Figure 26 Sphingolipid pathway

Simplified sphingolipid pathway highlighting the role of *KDSR* in de novo synthesis (black arrows). The green arrows indicate the recycling pathway. This pathway was not previously thought to contribute significantly to DHS levels. Adapted from Rego *et al.* [174]

4.3 Clinical characteristics

IPD123 is an eight-year-old male and was born to healthy genetically unrelated parents of European descent (Figure 27a) without complications by Caesarean section. At the age of 4 months he presented to collaborators in Leuven, Belgium with respiratory syncytial virus infection and was found incidentally to have a platelet count of 65 x 10^9/L and a mild normocytic, normochromic anaemia (haemoglobin 106 g/L, Figure 27b, c) with normal iron studies (serum iron 65 µg/dl, normal range (NR) 65-175 µg/dl, transferrin 2.61 g/dl, NR 2-3.6 g/dl, ferritin 63 µg/L, NR 30-400 µg/L). The reticulocyte count was shifted towards immaturity (count 45 x 10^9/L, range 20-100 x 10^9/L, immature fraction 28%, range 5-21%) indicating increased red cell turnover. On several occasions he presented with platelet counts <10x10^9/L accompanied by intermittent rectal bleeding following constipation, gingival bleeding when teething, marked skin bleeding during an episode of chicken pox, excessive ecchymosis following minor injury and recurrent epistaxis. He did not display features suggestive of EKVP4. However the only dermatological phenotype was a historical account of a slow-to-heal perianal wound following rectal manometry, which had healed at the time of recruitment. Over the last three years the proband's haemoglobin has normalised and his platelet count has improved, though not normalised (Figure 27c).

4.4 Pedigree analysis

The brother of IPD123 is unaffected. A sister was born in September 2017 and presented at birth with thrombocytopenia Figure 27a, c), dry skin under her eyes and ears and mild ichthyosis in her left axilla. Her skin symptoms improved spontaneously over the first month. The thrombocytopenia did not improve and no bleeding symptoms have been observed. The parents have no history of thrombocytopenia.

4.5 Further clinical laboratory studies

These were performed by the staff in the clinical laboratory of Professor Kathleen Freson, KU Leuven on IPD123. The affected sister was too young to provide further samples.

Light transmission aggregometry (LTA) was undertaken using platelet-rich plasma at a reduced concentration of 90x10^3 platelets/µl. The propositus had normal aggregation responses for ristocetin (105%), arachidonic acid (95%), U46619 (111%), collagen at 2 µg/ml (70%) and ADP at 5 µM (97%). The response to stimulation with 1µg/ml collagen was slightly reduced compared with two apparently healthy, unrelated controls. PRP was concentrated equally in the case and controls to 90x10^3plt/µl (Figure 27d). The LTA findings were similar to those reported [3]. Analysis by EM demonstrated platelets of normal size but with abnormal α- and δ- granules, some agranular platelets and abnormal internal membrane complexes (Figure 28).

Bone marrow examinations at presentation, at the age of five and eight years revealed increased numbers of dysplastic MK and progressive severe myelofibrosis (Figure 29a,b). Known causes of

myelodysplasia or primary myelofibrosis were excluded by analysis of bone marrow by Trusight Myeloid NGS gene panel testing. Fluorescence in-situ hybridisation studies excluded BCR-ABL fusion and karyotyping excluded structural abnormalities including of chromosomes 5 and 7 (Figure 29c).

4.6 *KDSR* variants may cause the observed phenotype

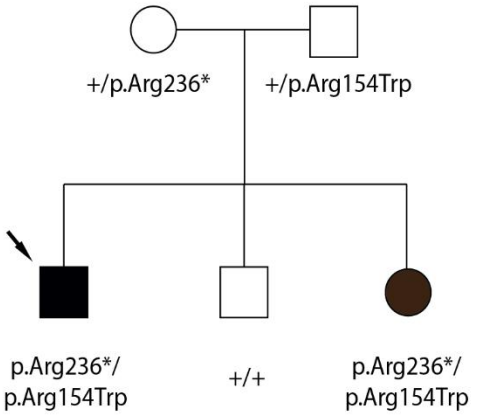
IPD123 and his affected sister carry nonsense variant 18:61006104 G>A (p.Arg236*) and missense variant 18:61018270 G>A (p.Arg154Trp) in KDSR. The variants have minor allele frequencies (MAF) in Europeans of $4.82x10^{-5}$ and $2.32x10^{-4}$ and were of maternal and paternal origin respectively (Figure 27a) [95, 102]. The results of co-segregation study were concordant with the assumed autosomal recessive nature of the condition (Figure 30).

There are no individuals in the GnomAD database with homozygous loss of function variants or compound heterozygous missense and loss of function variants in *KDSR*, and only nine individuals are heterozygous for one of six documented loss of function variants (accessed 3.1.18, [95]). This is consistent with the rarity of the phenotype.

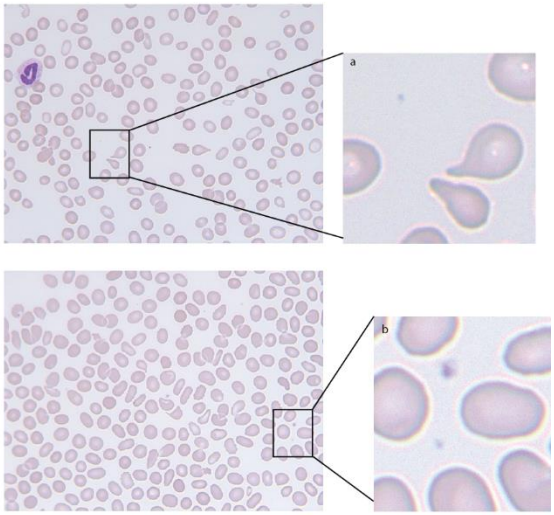
Figure 31 shows all pathogenic variants published to date alongside the novel variants, and the broad phenotypes that those individuals carry. There is no clear correlation between the location of the five missense mutations and the presence of skin or platelet phenotypes. However four out of the five known missense variants, including p.Arg154Trp in IPD123, lie in the region coding for the catalytic domain, with the other one lying in the second transmembrane anchor.

The variant p.Arg154Trp introduces a large, uncharged tryptophan instead of a smaller, positively charged arginine. These differences may interfere with the tertiary structure of the catalytic domain to reduce its activity. There is also conservation of p.Arg154 back to the zebrafish (Figure 32). These findings are reflected in a high CADD score of 35 [102]. Both variants are found in the most abundant *KDSR* transcripts in platelets and MK (Figure 31) [103].

a



b



c

	Lymphocytes	Erythrocytes	Hemoglobin	Platelets
Control values Adults	*4.5-13.5*	*4-5.2*	*11.5-15.5*	*150-450*
Father	5.9	4.88	15.5	258
Mother	5.22	4.25	12.8	415
Control values Children	*5.5-15.5*	*3.9-5.3*	*11.5-13.5*	*150-550*
Healthy brother, 4 year	5.04	3.94	11.5	521
Proband, 4 Months	3.0	3.8	10.6	65
Proband, 4 year	7.84	3.41	9.1	9
Proband, 5 year	14.53	3.69	10.7	41
Proband, 6 year	7.61	3.77	10.6	15
Proband, 8 year	7.06	4.67	13.6	77
Proband, 8.5 year	9.49	4.24	12.5	152
Control values Neonates	*5.5-21*	*3.9-6.3*	*13.5-21.5*	*150-550*
Affected Sister, 1 week	11.64	5.73	19.1	73
Affected Sister, 5 Months	9.34	3.33	8.6	148

d

	Proband*	2 healthy controls
Amplitude for Horm Collagen 1 µg/ml	25%	63 and 66%
Amplitude for ADP 2.5 µM	52%	48 and 28%

Figure 27 Pedigree tree, clinical characteristics, blood counts and platelet aggregometry

(a) Pedigree and variants identified in *KDSR*. '+' denotes the major allele. The proband (arrowed) and his sister, but not his healthy brother, carry the nonsense variant 18:61006104 G>A (p.Arg236*) and the missense variant 18:61018270 G>A (p.Arg154Trp). Co-segregation analysis demonstrated that the mother carries the former and the father the latter variant.

(b) Peripheral blood smears show dysmorphic erythrocytes with poikilocytosis and tear drop cells and an obvious low number of platelets of normal volume.

(c) Serial blood results for proband and pedigree members

(d) Light transmission platelet aggregometry studies for proband and healthy controls. The proband had normal aggregation responses for ristocetin, arachidonic acid, U46619, collagen at 2µg/ml and ADP at 5µM.

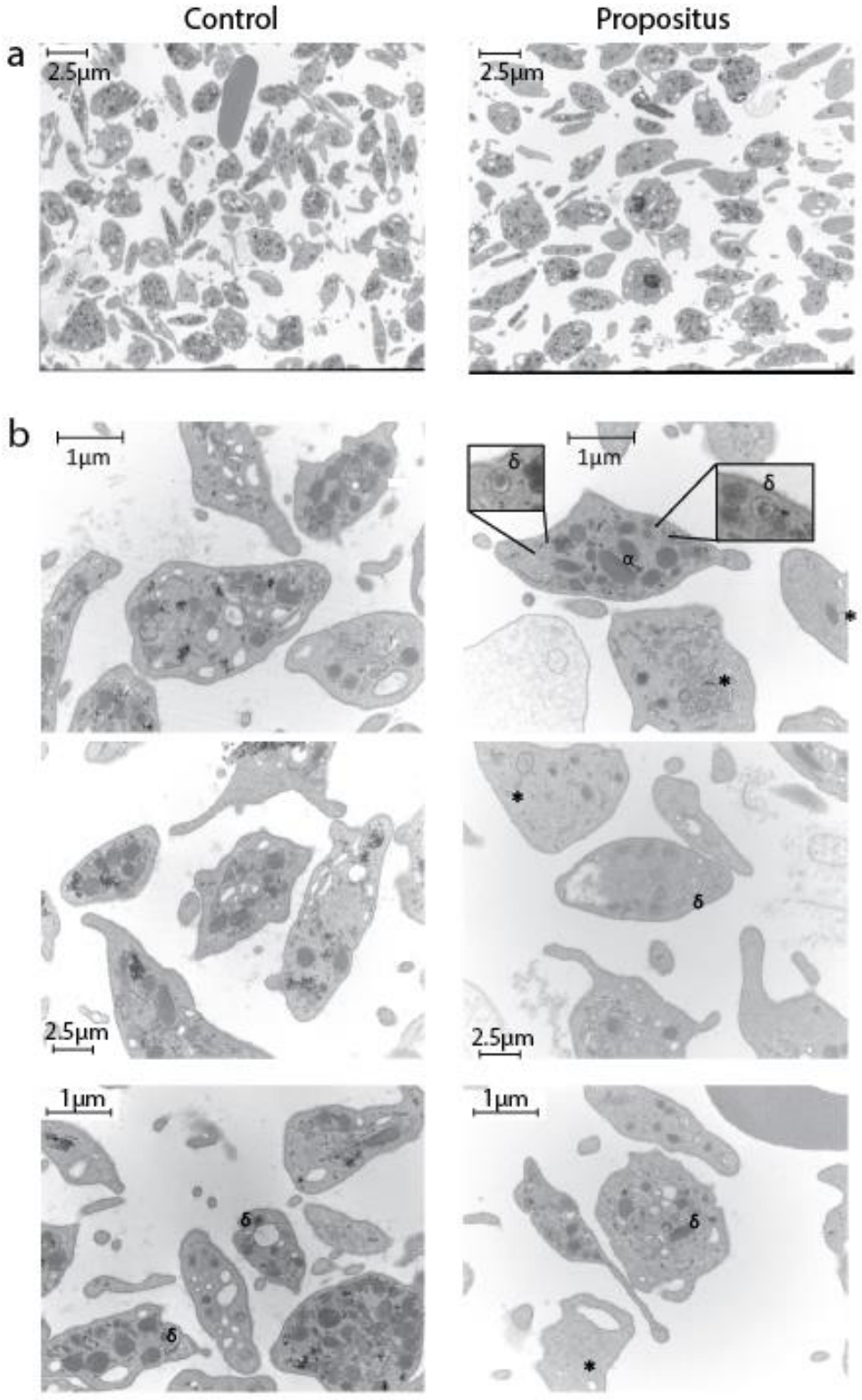
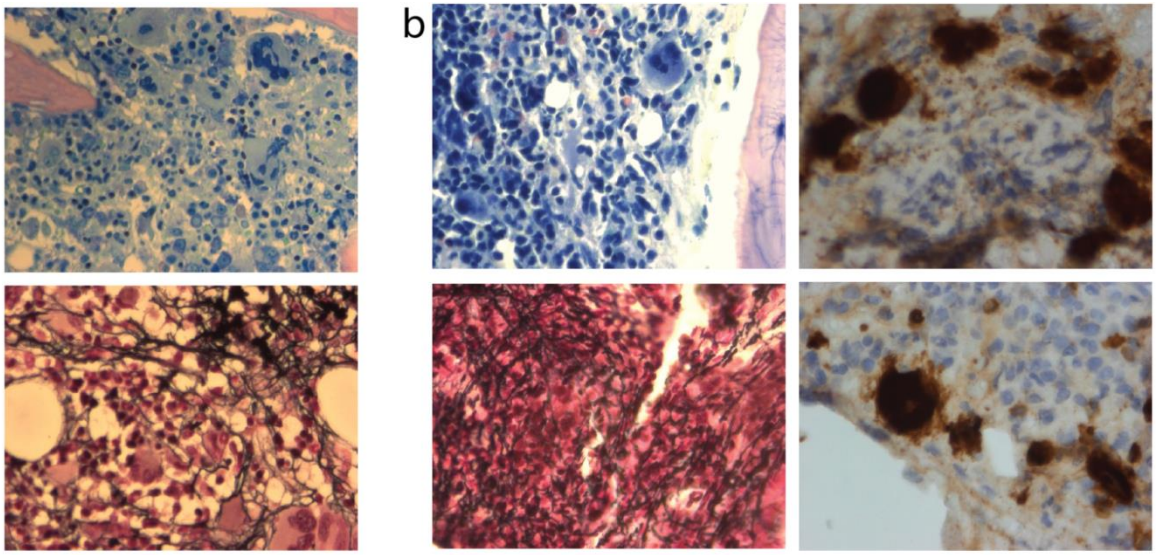


Figure 28 EM from the proband and an unrelated apparently healthy control.

(a) Proband platelets have similar size and shape to control platelets.

(b) Proband platelets demonstrate enlarged α- granules and smaller δ- granules with a diffuse content. Platelets marked by * have reduced α- and δ- granule number and prominent intra-platelet membrane complexes.

a b



c

	Peripheral blood smears	Bone marrow biopsy	Bone marrow genetics
II:2 (Proband) 4 years	White blood cells normal, mild normocytic normochromic anaemia with mild aniso-poikilocytosis. Severe thrombocytopenia.	Erythrophagocytosis. Dysplastic megakaryocytes that are located in clusters. Normal lymphoid lineage with granulocytes. Myelofibrosis grade 2 (reticulin).	Negative: BCR-ABL, FISH (monosomy 7), JAK2, CALR and MPL, chromosomal breakage
II:2 (Proband) 5 years	Mild monocytosis and neutrophilia , mild normocytic normochromic anaemia with anisopoikilocytosis and obvious thrombocytopenia	Left shifted myeloid lineage. Increased numbers of dysplastic megakaryocytes. Erythroid lineage difficult to study due to insufficient numbers. Myelofibrosis grade 3 (reticulin, collagen). Cellularity 90%, myeloid/erythroid: 4/1. No c-kit or CD34 positive blasts.	
II:2 (Proband) 8 years	Mild poikilocytosis and thrombocytopenia.	Normal cellularity 70-80%. Normocellular but hyperplastic and left-shifted myeloid lineage. Prominent hyperplastic megakaryocyte lineage with strong cluster formation (LAT staining). Megakaryocyte dysplasia with both hyperlobulated and small hypolobulated, hyperchromic cells. Normocellular erythroid lineage without dysplasia. Myelofibrosis grade 3 (reticulin,collagen switch). Normal Myeloid/erythroid: 1/1. No evidence of $CD34^{+}$ blasts. Suggestive of MDS or MPN.	Negative: NGS panel test TruSight Myeloid Sequencing (Illumina) with coverage of 500x. Normal karyotype.

Figure 29 Bone marrow (BM) biopsy of proband. Magnification of images x40

(a) BM biopsy of proband at 4 years of age. Upper: Dysplastic MK with hypolobated nuclei and uncommon sizeable clusters shown with May-Grunwald-Giemsa (MGG) staining. Lower: Fibrosis is shown on stromal reticulin staining.

(b) BM biopsy at 8 years of age. Left upper: Dysplastic MK in clusters remain on MGG stain. Left lower: Marrow fibrosis with increased reticulin staining supporting progression of myelofibrosis. Right upper and lower: large, dysplastic MK are highlighted by dense LAT staining, which preferentially stains atypical MK in myeloid disorders [175].

(c) Serial morphological assessments in proband show progression of myelofibrosis. Bone marrow genetics excluded known genetic causes of myelodysplasia and myeloproliferative neoplasms.

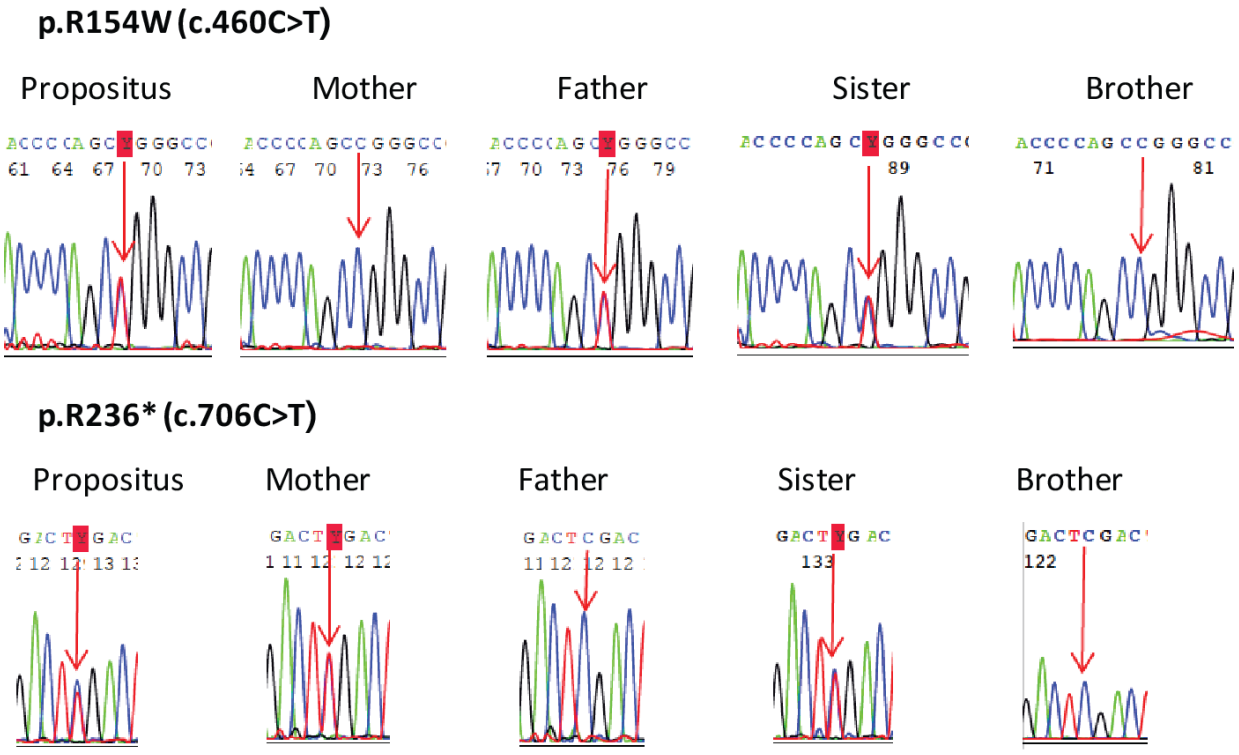


Figure 30 Sanger sequencing of IPD123 pedigree.
Sanger sequencing confirmed two rare variants (18:61006104 G>A and 18:61018270 G>A) in the proband and his affected sister; the healthy brother carried the major allele for both variants. The parents were heterozygous.

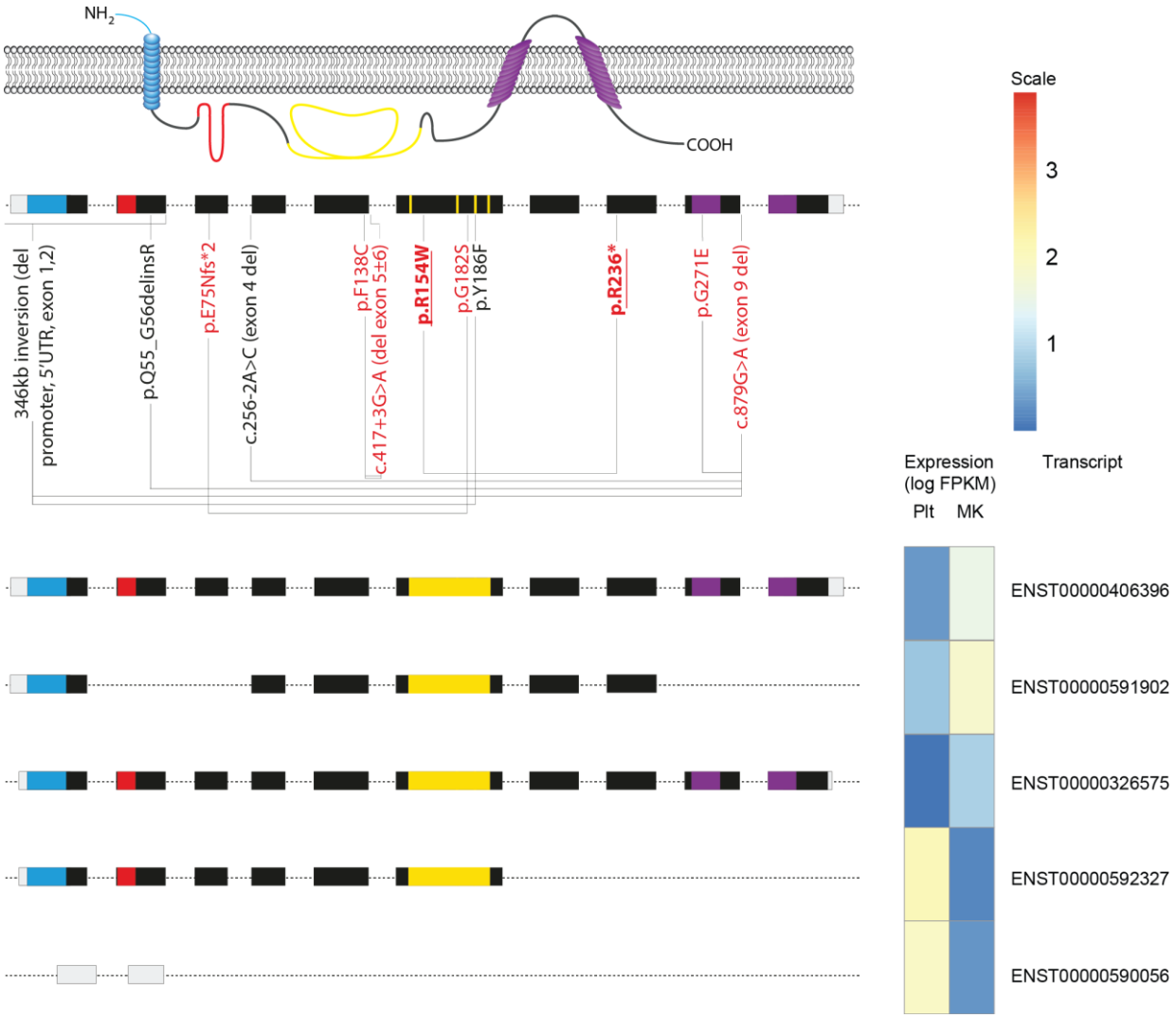


Figure 31 Genotypes, phenotypes and transcript expression of *KDSR*.
Schematic of cDNA and protein showing location of known *KDSR* variants with documented phenotypes in skin only (black), skin and platelets (red) and novel variants in the pedigree reported in this manuscript in bold. Variants are linked by brackets where present in compound heterozygosity in an individual. Key structural elements of KDSR are illustrated: transmembrane anchors (blue, purple), the Rossman folds (red) and a highly conserved domain containing three putative catalytic sites (yellow). The novel p.Arg154Trp variant is within the catalytic domain. Lower panel shows a heat map of *KDSR* transcripts present at >0.5 log FPKM in either megakaryocytes (MK) or platelets (Plt) [103]. The novel p.Arg154Tryp missense variant is present in the most abundant transcripts in MK and the principal coding transcript in platelets. Adapted from Gupta *et al.* [176].

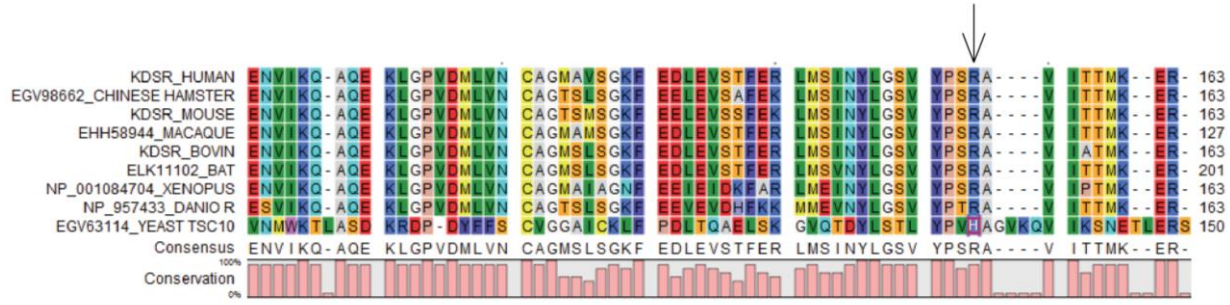


Figure 32 p.Arg154 is highly conserved.
CLC sequence viewer orthologous alignment shows conservation of p.Arg154 (arrow) as far as the zebrafish but not in yeast, and high level conservation in neighbouring amino acids (pink bars show % conservation)

4.7 The proband has elevated KDS, consistent with loss of function

I reasoned that the variants in *KDSR* would cause a loss of enzymatic function leading to a build-up of the substrate KDS (Figure 26). To assess this, global metabolic profiling of plasma from the proband, parents and unaffected sibling was performed in collaboration with Metabolon, Inc. (Durham, North Carolina, US) using a liquid chromatography tandem mass spectrometry (LC-MS/MS) platform as described in section 2.5.6. Results were compared with 496 subjects between the ages of 4 and 55 years of age without thrombocytopenia.

KDS was clearly detectable in the proband's plasma and undetectable in the plasma of the parents, the healthy sibling and controls (Table 14 and Table 15, Figure 33). Unexpectedly, the downstream metabolite dihydrosphingosine (DHS) was also elevated (Z-score 5.4, representing the number of standard deviations from the mean of the cohort, (Figure 34). There was also a trend towards higher scores immediately downstream of DHS.

In view of the unexpected findings, the sphingolipid profile in the proband and parents were assessed on a separate platform, by high resolution LC-MS/MS in collaboration with Dr. Albert Koulman and Dr. Ben Jenkins of the NIHR Biomedical Research Centre Core Metabolomics and Lipidomics Laboratory, University of Cambridge. Once again the chromatographic peak of 3-ketodihydrosphingosine (KDS) was detectable in the proband and absent in plasma from healthy pedigree members. The similarity of peaks in the emission spectrum between the KDS standard and KDS detected in the propositus left no doubt regarding the identity of KDS (Figure 35).

	IPD123	IPD123-mother	IPD123-father	IPD123-brother	Median (controls)	S.D (controls)
KDS	8.5 E5	0	0	0	0	0
DHS	3.6 E5	0	0	0	6.7 E4	3.2 E4
dihydroceramide	6.3 E5	3.3 E5	2.6 E5	2.5 E5	3.1 E5	1.8 E5
sphinganine-1-phosphate	6.7 E5	4.6 E5	3.4 E5	4.0 E5	3.9 E5	1.4 E5
Dihydrosphingomyelin	3.7 E7	2.9 E7	2.5 E7	2.8 E7	2.4 E7	5.5 E6
Ceramide	3.5 E6	3.6 E6	2.9 E6	2.8 E6	2.8 E6	7.9 E6
Sphingosine	1.8 E5	1.0 E5	6.2 E4	9.5 E4	1.5 E5	8.3 E4
sphingosine 1-phosphate	3.4 E6	3.1 E6	2.3 E6	2.7 E6	2.6 E6	6.2 E5
total sphingomyelins	1.3 E9	1.2 E9	1.3 E9	1.2 E9	1.2 E9	1.8 E8
total glycosphingolipids	1.2 E7	8.9 E7	1.2 E7	1.2 E7	1.0 E7	2.6 E6
phosphoethanolamine	1.6 E6	1.1 E6	8.0 E5	1.7 E6	1.6 E6	6.1 E5

Table 14 Raw ion counts from Metabolon analysis of plasma sphingolipids.

The sample was tested once by Metabolon. 3-ketosphinganine (KDS) was detectable in the proband alone. Downstream metabolite dihydrosphingosine (DHS) was unexpectedly high. The relatively abundant sphingomyelins may be a source of DHS by a compensatory mechanism. Median and standard deviation are calculated from 496 unrelated controls without thrombocytopenia. Counts for 17 individual sphingomyelin moieties and 3 individual glycosphingolipids are summarised here and the breakdown is detailed in Appendix 8.2.1.

	IPD123	IPD123-mother	IPD123-father	IPD123-brother
KDS	NA	NA	NA	NA
DHS	5.4	0.3	0.3	0.3
Dihydroceramide	1.6	0.2	-0.4	-0.5
sphinganine-1-phosphate	1.6	0.4	-0.4	0.1
Dihydrosphingomyelin	1.9	0.8	0.2	0.6
Ceramide	0.8	1.0	0.2	0.1
Sphingosine	0.4	-1.2	-2.5	-1.4
sphingosine 1-phosphate	1.2	0.7	-0.4	0.2
total sphingomyelins	-0.6	0.0	-0.2	-0.2
total glycosphingolipids	0.4	-0.3	0.4	1.1
Phosphoethanolamine	0.0	-1.4	-2.6	0.3

Table 15 Z-scores from Metabolon analysis of plasma sphingolipids.

Values represent the number of standard deviations from the median. As KDS was not detected in any other sample, including historical samples, no Z scores can be calculated. Colours reflect the heatmap in Figure 34 Sphingolipid heatmap. Counts for 17 individual sphingomyelin moieties and 3 individual glycosphingolipids are summarised here and the breakdown is detailed in Appendix 8.2.3.

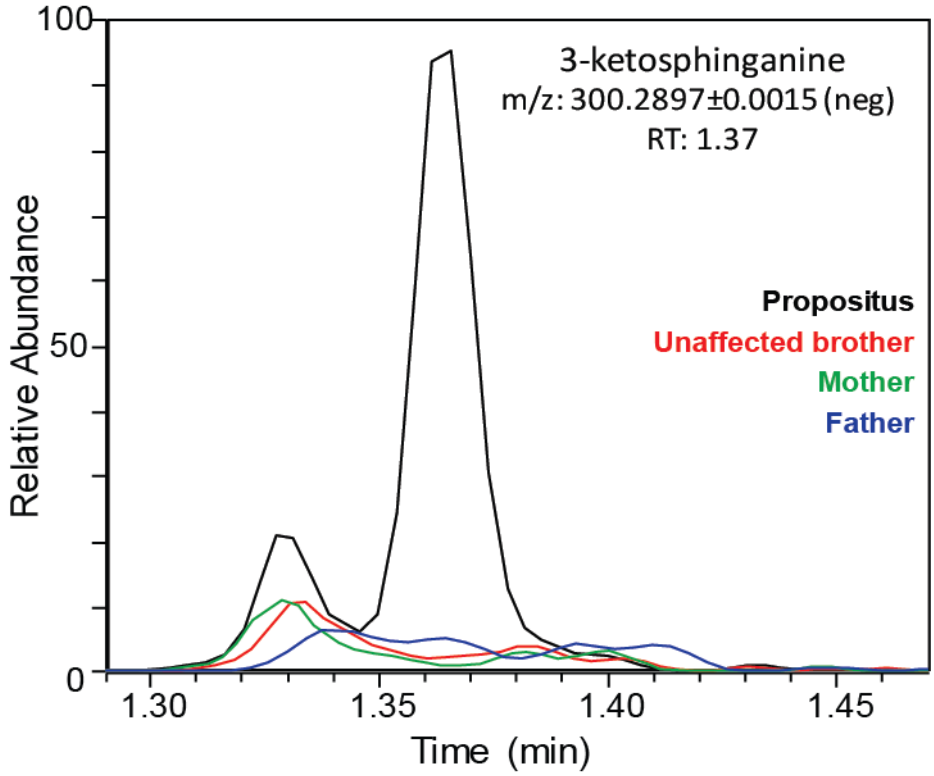


Figure 33 Metabolic profiling of plasma.
Metabolon studies showed that the chromatographic peak of 3-ketodihydrosphingosine (KDS) is present in the proband and absent in plasma from healthy pedigree members; m/z: mass-to-charge ratio; RT: retention time.

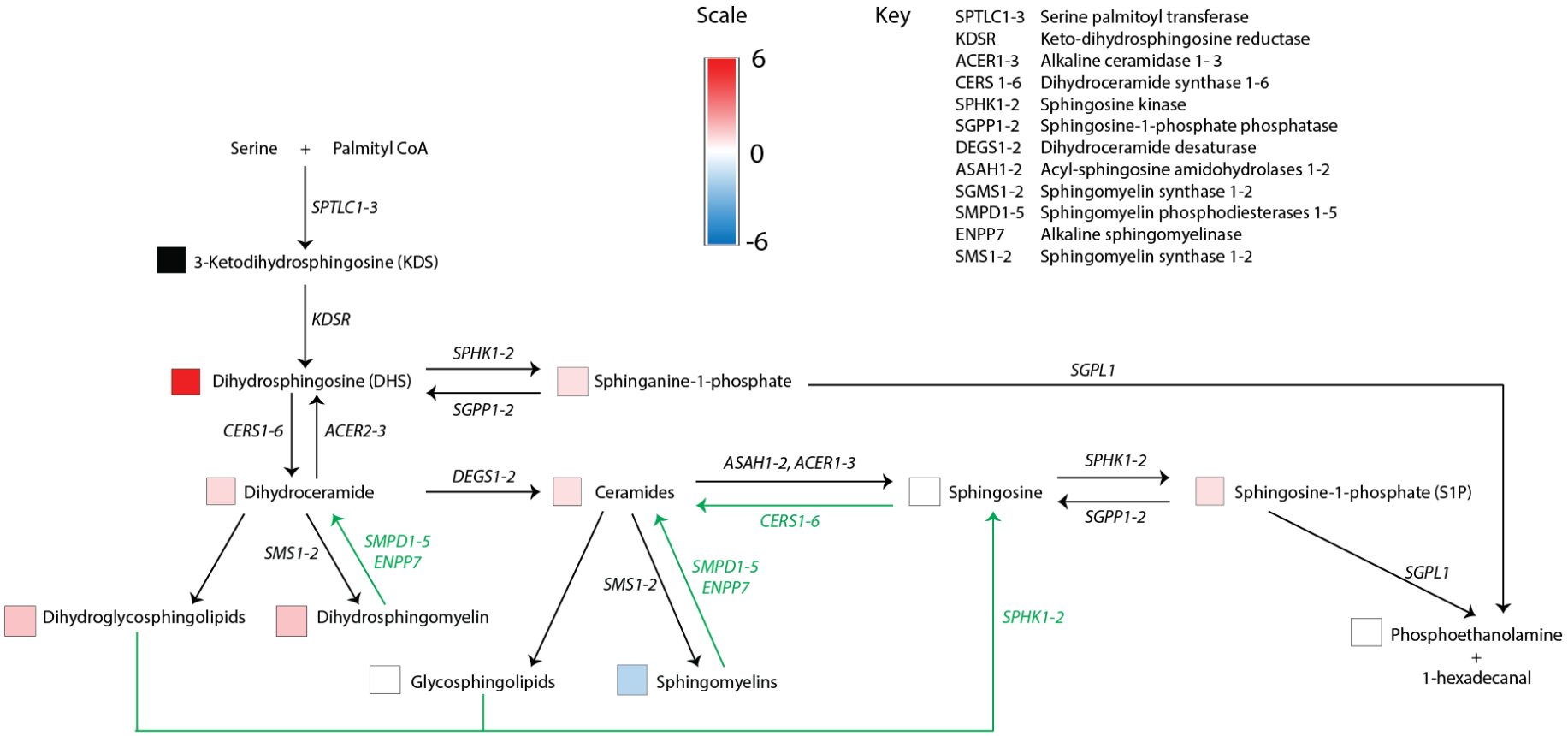


Figure 34 Sphingolipid heatmap.
Pathway as detailed in Figure 26 with an embedded heatmap, where coloured squares next to a compound represent Z-scores from Table 15. Controls were 496 subjects between the age of 4 and 55 years without thrombocytopenia. The black square adjacent to KDS indicates that it was detected in the index case but not in any controls, therefore Z-scores could not be calculated for KDS.

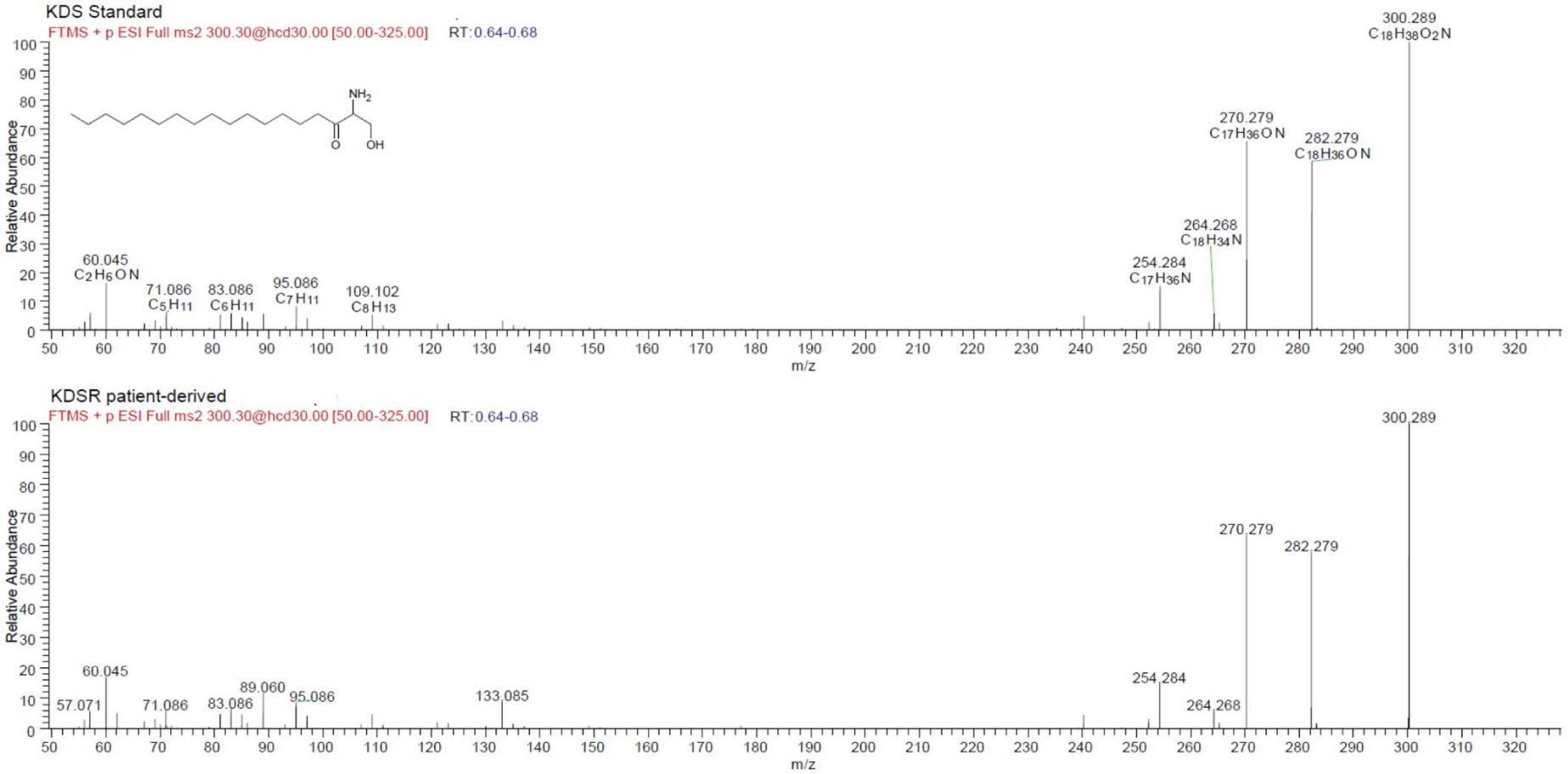


Figure 35 Focused sphingolipid profiling showed conformity of peaks between KDS standard and KDS signal in proband plasma.
FTMS, Fourier transformation mass spectrometry; +pESI, positive mode electrospray ionisation; HCD, higher energy collisional dissociation. Set at the specific fragmentation of ion 300.3 m/z by HCD at a relative energy of 30%, obtaining data from the range 50 to 325 m/z.

4.8 Cellular model I: CRISPR Cas9 knock-in

4.8.1 Choice of strategy

I evaluated the following hypotheses using haematopoietic cellular modelling:

1. *KDSR* variants cause thrombocytopenia
2. Transcriptional or post-translational mechanisms compensate for loss of *KDSR* function

As a first step CRISPR Cas9 targeted gene editing was used in induced, pluripotent stem cells (iPSC), aiming to introduce a homozygous p.Arg154Trp variant, with a view to forward programming to induced MK (iMK, section 1.10.1.2) to explore the phenotype. The other allele in the proband is a premature stop codon, so this model would test the hypothesis that p.Arg154Trp is hypofunctional, therefore capable of causing the phenotype. The strategy was ultimately unsuccessful, however some interesting results were obtained and are the subject of this section.

4.8.2 Results

4.8.2.1 Confirmation of reagents

I chose the iPSC line S4 for gene targeting following successful CRISPR knockout by other researchers in the Ouwehand laboratory. S4 iPSC were shown by Sanger sequencing to carry the reference sequence at the variant position, and over 454bp 5' and 186bp 3' of the variant, which includes all targeted sequence (Appendix). sgRNAs were shown by Sanger sequencing to have been cloned into the pSpCas9n (BB)-2A GFP plasmid (Appendix 8.2.5).

4.8.2.2 Nucleofection yield and survival

Three conditions, defined by different sgRNAs were tested, in view of the likelihood of differing efficiency between different sgRNAs. Details of each condition, including sgRNA sequences, are described in section 2.5.3.1. The optimal length of ssODN homology arms was uncertain; initial reports suggested that 90bp was the optimal length [177], whilst subsequently longer ssODN were considered more efficient [178]. For the initial nucleofection I opted to use 170bp ssODNs. This was chosen for practical reasons because longer ODNs than this had a long lead time for production. To recap, the plasmid carrying the Cas9n also carried a GFP marker. Conditions 1, 2 and 3 yielded 2.9%, 2.8% and 3.2% GFP^+ cells respectively. Therefore transfection efficiency with Cas9n was disappointingly low in iPSC; transfection of the same plasmid in HEK293T cells yielded 43% GFP^+ cells. Given that the reported efficiency of HDR-directed gene editing in iPSC was ~1% [177], GFP^+ iPSC were single cell sorted but formed few, sparse colonies that became infected within a few days and died. Electroporation without exogenous DNA and no treatment controls were not infected, suggesting that contamination was acquired during sorting.

Nucleofection was repeated, this time using 200bp ssODNs following subsequent reports supporting efficiency with this template length [126]. The proportion of cells expressing Cas9n-GFP^+ was similar at

2.5%, 3.1% and 3.6% for conditions 1, 2 and 3 respectively. This time penicillin-streptomycin was added to the culture for 48 hours following single cell sorting. 44 clones were established and of these, 34 survived expansion. Conditions 1, 2 and 3 produced 15, 10 and 8 established clones, respectively.

4.8.2.3 PCR amplification of gDNA showed no knock-in of pArg154Trp or out-of-frame indels

PCR amplicons crossing the targeted region from the expanded clones were Sanger sequenced. 26 high quality traces were checked for the knock-in, which was absent (Figure 36).

Eight traces showed only short regions of alignment to the genome. To establish whether this was attributable to heterozygosity for indels, PCR amplicons from these clones were cloned into the pGEM-T Easy vector to separate alleles for sequencing. Table 16 summarises the results of these findings, and images of the sequence alignments are provided in Appendix 8.2.6. Three clones were heterozygous for an in-frame indel. One clones was homozygous for a 3bp in-frame indel. Two clones yielded colonies with normal amplicons and two clones carried one normal allele but remained inconclusive due to low quality tracings for the other allele.

4.8.2.4 Termination of knock-in strategy

At this stage I decided not to continue with this approach. The knock-in strategy efficiency had been ineffective and the experiment was sub-optimally designed for knock-out production; for gene knock-out the 1st exon would be the ideal target. It was also concerning that the indels seen were all in-frame. The probability of this occurring by chance is $7x10^{-4}$ and suggests that *KDSR* knockout or even haploinsufficiency may be non-viable in this cell line following the specific stresses of nucleofection and cell sorting.

Figure 36 CRISPR Cas9 targeted gene editing of iPSC to introduce a homozygous p.Arg154Trp variant was unsuccessful.
Figure shows Sanger sequencing traces of genomic DNA from twenty-six iPSC clones in which direct alignment to the genome was possible. These clones did not carry the p.Arg154Trp knock-in. Arrow indicates position of p.Arg154.

Condition	iPSC clone	1st allele	2nd allele	Conclusion
1	6	3bp insertion	3bp insertion	Likely homozygous 3bp insertion
1	7	Normal amplicon	33bp deletion	Heterozygous, in-frame deletion
1	18	Normal amplicon	12bp deletion	Heterozygous, in-frame deletion
1	19	Normal amplicon	Normal amplicon	Normal
2	3	Normal amplicon	Normal amplicon	Normal
2	9	Normal amplicon	196bp insertion, 244bp deletion	Heterozygous in-frame deletion
2 1	11 8	Normal amplicon	Could not be aligned to genome	Inconclusive

Table 16 Sanger sequencing of targeted iPSC, following separation of amplicons by cloning into the pGEM-T Easy vector.
Only in-frame indels were identified.

4.9 Cellular model II: Lentiviral rescue of patient-derived iPSC

4.9.1 Choice of strategy

The transduction of iPSC with lentiviral vectors carrying transcription factors induces expression with high efficiency during iMK forward programming (FoP) [123]. In view of this, I pursued a strategy of lentiviral expression of the wild type open reading frame (ORF) of *KDSR* in IPD123-derived iPSC.

4.9.2 IPD123 iPSC fragility in culture

IPD123 patient-derived lines were prepared and validated for pluripotency markers by the Cambridge Biomedical Research Centre from skin fibroblasts as described in section 2.4.1.1, following the protocol of Takahashi and Yamanaka [112, 133]. Three clones were provided, hereafter labelled KCl1, 2, 4 (clone three did not pass the stringent quality control checks provided by the facility).

Sanger sequencing of IPD123 iPSC verified the presence of both putative pathogenic variants in the proband in iPSC (18:61006104 G>A and 18:61018270 G>A, Appendix 8.2.7). IPD123 iPSC showed fragility in cell culture with increased propensity to differentiation and cell death following stress such as passage at high ratio, transduction and addition of polybrene. A puromycin kill curve was performed on the patient-derived iPSC line. The optimal puromycin concentration to cause cell death in 4 days was 0.1µg/ml, tenfold lower than used in phenotypically normal 'stock' iPSC lines including A1ATD-1, S4 and QOLG.

4.9.3 IPD123 *KDSR* variants are associated with increased apoptosis in iPSC

To explore this fragility, and in view of the aforementioned role of sphingolipids as bioactive molecules, apoptosis was induced in the three IPD123 iPSC clones and three stock iPSC lines (A1ATD-1, QOLG, S4) in three independent experiments by addition of 0-500µM hydrogen peroxide to the culture medium. The degree of apoptosis was measured by Apo-BrdU TUNEL TdT assay (as described in section 2.4.1.7). This assay labelled exposed 3'hydroxyl end of genomic fragments produced during apoptosis with BrdUTP by terminal deoxynucleotidyl transferase (TdT). BrdU was then detected using a green-fluorescent Alexa Fluor 488-labelled anti-BrdU antibody. This assay was chosen over annexin V expression because iMK express annexin V at high levels *(personal communication, Dr. C. Ghevaert).*

An ordinary one-way ANOVA test followed by Tukey's multiple comparisons test showed significantly increased apoptosis in the IPD123 line compared with all stock lines at 400 and 500µM H_2O_2 supplementation ($p < 0.05$, Figure 37, Table 17). However stock lines are chosen in part due to their ability to withstand stress induced by manipulation, limiting the conclusions that can be made from this experiment.

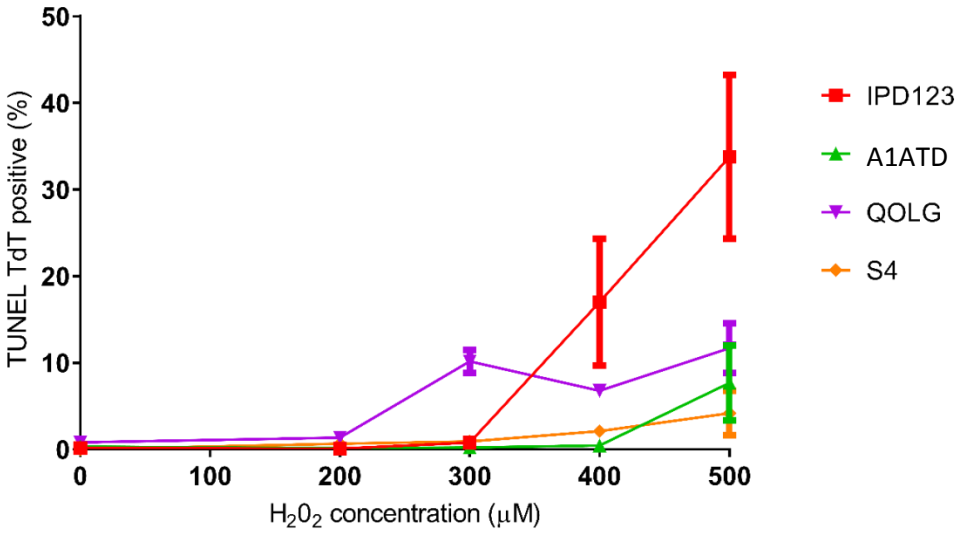


Figure 37 iPSC carrying IPD123 *KDSR* variants undergo higher rates of apoptosis than stock iPSC lines
Apoptosis was induced by incubating IPD123 iPSC and three other stock iPSC lines (A1ATD-1, QOLG and S4) with hydrogen peroxide (H_2O_2). Cells were processed using a TUNEL TdT assay that labels the ends of genomic fragments produced during apoptosis. The experiment was repeated three times. IPD123 iPSC underwent a higher rate of apoptosis than stock iPSC lines.

[H_2O_2]	Tukey's multiple comparisons test	Mean difference between lines	95% CI of difference	p-value
400μM	IPD123 vs. A1ATD-1	16.54	6.858 to 26.22	0.0026*
	IPD123 vs. QOLG	10.24	0.5543 to 19.92	0.0387*
	IPD123 vs. S4	14.89	5.208 to 24.57	0.0051*
	A1ATD-1 vs. QOLG	-6.303	-15.99 to 3.379	0.2364 (ns)
	A1ATD-1 vs. S4	-1.65	-11.33 to 8.032	0.9451 (ns)
	QOLG vs. S4	4.653	-5.029 to 14.34	0.4601 (ns)
500μM	IPD123 vs. A1ATD-1	26.11	11.67 to 40.56	0.0018*
	IPD123 vs. QOLG	22.11	7.664 to 36.55	0.0052*
	IPD123 vs. S4	29.59	15.15 to 44.04	0.0008*
	A1ATD-1 vs. QOLG	-4.007	-18.45 to 10.44	0.8111 (ns)
	A1ATD-1 vs. S4	3.48	-10.96 to 17.92	0.8651 (ns)
	QOLG vs. S4	7.487	-6.956 to 21.93	0.4014 (ns)

Table 17 Ordinary one-way ANOVA of Figure 37.
IPD123 showed significantly higher rates of apoptosis by TUNEL TdT following exposure to H_2O_2 than stock lines at 400μM and 500μM concentrations. * p<0.05; ns, non-significant.

4.9.4 pLenti-EF1a-KDSR-mGFP expression studies

iPSC were transduced with pLenti-EF1a-KDSR-mGFP as described in Section 2.5.5, which carries the wild-type ORF of *KDSR* transcript ENST00000591902 fused to mGFP. This transcript has the highest expression in MK (Figure 31). Transduction of IPD123 iPSC clones with all vectors was complicated by cell death, either following transduction or following addition of selective antibiotics, despite successful transduction of control line A1ATD-1. IPD123 clone 4 survived transduction, with 63% of cells GFP^+ at day ten after transduction and is hereafter referred to as $KGFP^+$.

4.9.4.1 The KDSR-mGFP$^+$ fusion protein trafficks to the ER in HEK293T cells

An important aspect of a functional protein is correct cellular trafficking. It has previously been shown that KDSR localises in the ER in HeLa cells [161]. Therefore pLenti-EF1a-KDSR-mGFP was co-transfected with ER marker plasmid mCh-Sec61β, as detailed in section 2.4.3.3. mGFP clearly co-localised with mCh-Sec61β, supporting the idea that the KDSR-$mGFP^+$ fusion protein is at least partly functional (Figure 38) as well as confirming previous published findings.

4.9.4.2 TUNEL-TdT assay is not informative using the KGFP$^+$ model

The TUNEL TdT assay was performed in $KGFP^+$ using concentrations between 0-500μM H_2O_2 to induce apoptosis. At baseline cells were clearly gated as GFP^+ (38%) and GFP^- (52%), with the latter providing a transduction control. However at higher concentrations of H_2O_2 cells converged at a fluorescence intensity between the gates (Figure 39). Possible explanations are that GFP^- cells began to auto-fluoresce during cell death and/or that GFP^+ cells lost fluorescence during cell death. Therefore no conclusions could be taken from this experiment.

4.9.4.3 Expression of pLenti-EF1a-KDSR-mGFP was lost, including at iMK

With serial passages the $mGFP^+$ signal was progressively silenced to 30%, which may suggest that expression of the normal ORF confers no advantage to the cells despite correct localisation in the ER. Furthermore, following FoP to iMK the GFP^+ signal was lost completely, as shown in
Figure 40. Therefore no further work was undertaken with this model.

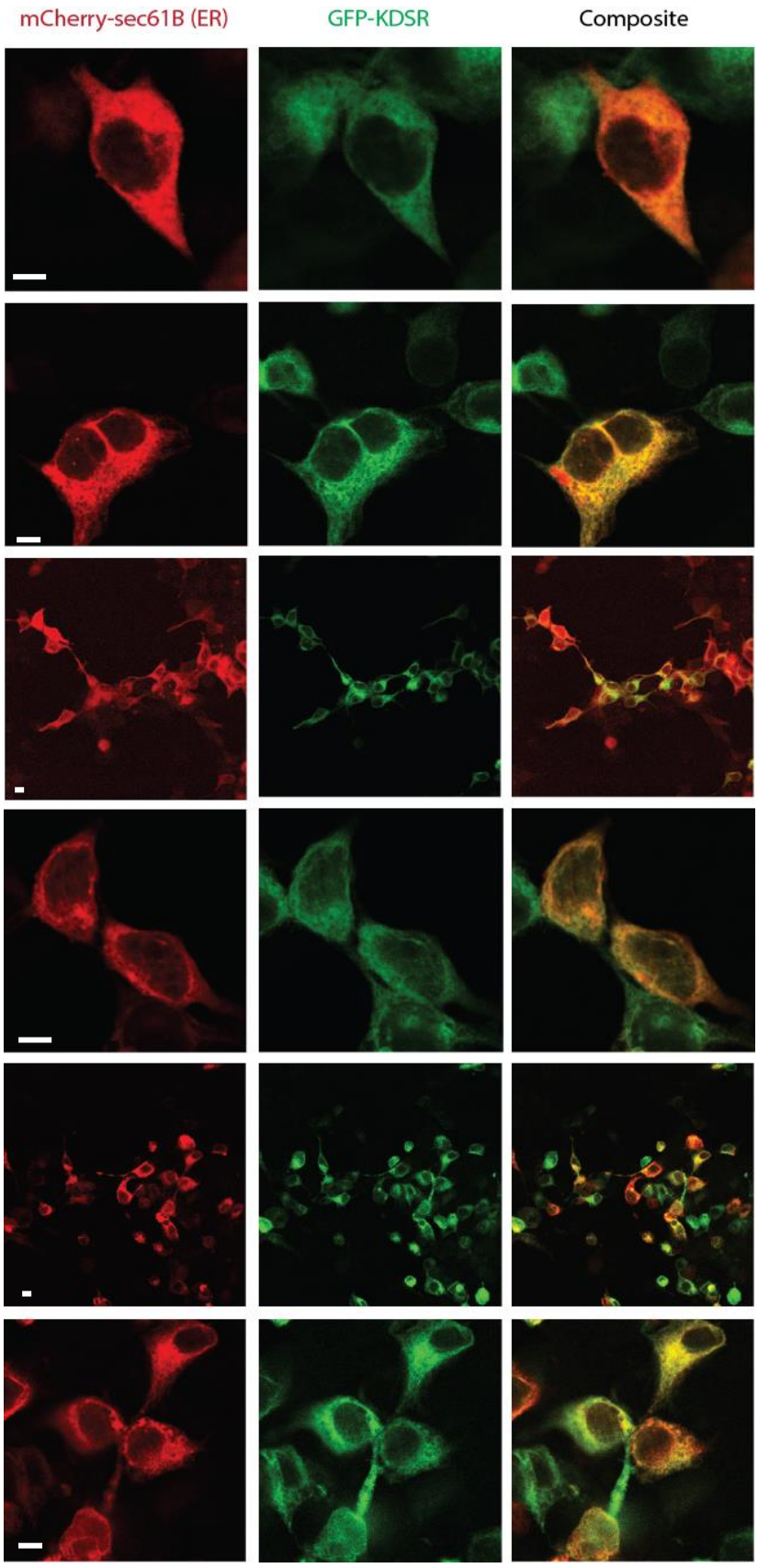
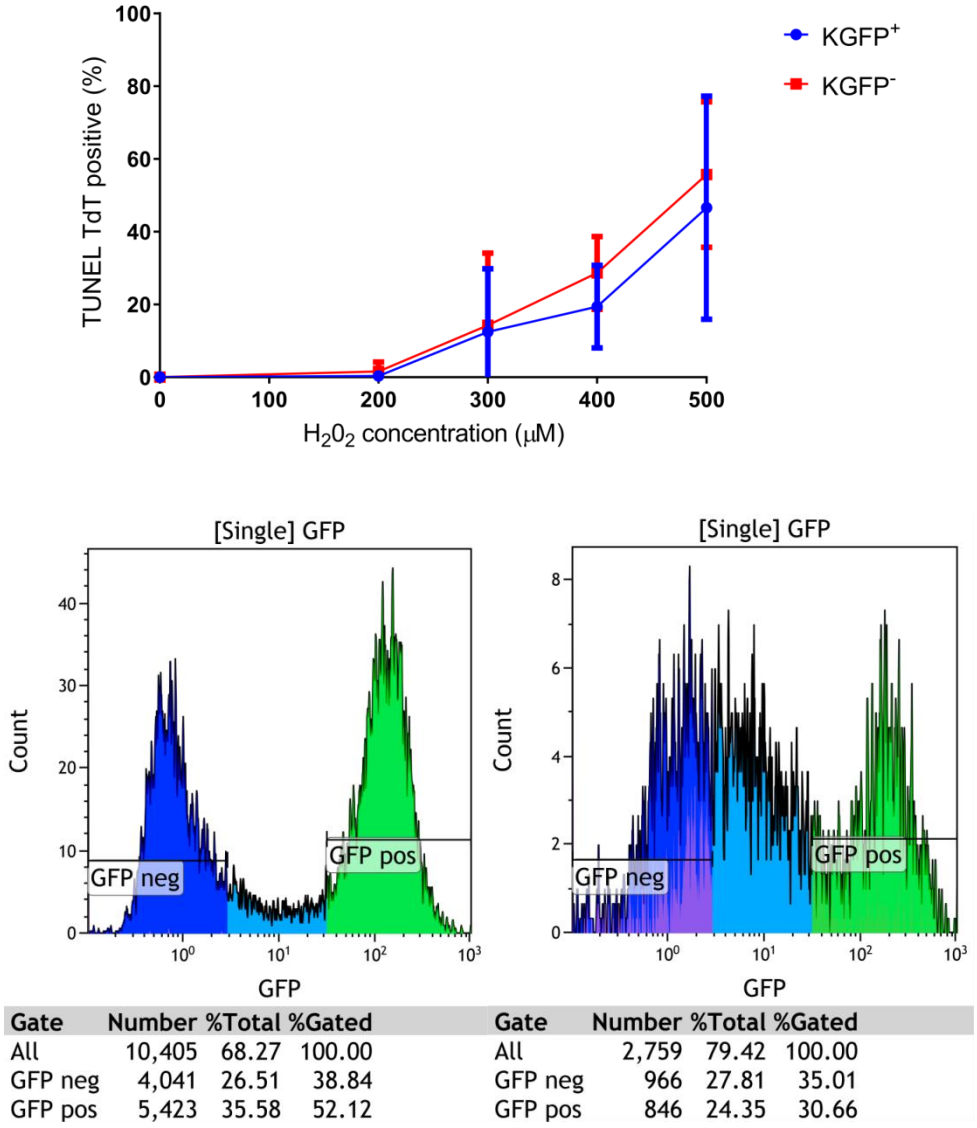


Figure 38 Transfection of HEK293T cells with pLenti-EF1a-KDSR-mGFP.
HEK293T were plated for 4 hours prior to transfection with pLenti-EF1a-KDSR-mGFP. Images were taken 48 hours following transfection. Left: red mCherry signal corresponds to mChSec61β, which labels endoplasmic reticulum. Middle: green GFP signal corresponds to pLenti-EF1a-KDSR-mGFP, which contains the open reading frame of *KDSR* with a GFP tag). Right: overlay shows co-localisation, indicated in yellow, suggesting that GFP-tagged KDSR trafficks to the endoplasmic reticulum. White scale bar indicates 10µm. Multiple images from the same transfection are shown.



Gate	Number	%Total	%Gated
All	10,405	68.27	100.00
GFP neg	4,041	26.51	38.84
GFP pos	5,423	35.58	52.12

Gate	Number	%Total	%Gated
All	2,759	79.42	100.00
GFP neg	966	27.81	35.01
GFP pos	846	24.35	30.66

Figure 39 TUNEL TdT assay of KGFP$^+$ iPSC was inconclusive.
Apoptosis was induced by incubating KGFP$^+$ iPSC with hydrogen peroxide (H_2O_2). Cells were processed using a TUNEL TdT assay that labels the ends of genomic fragments produced during apoptosis. The experiment was repeated three times. Top panel: There was no clear difference in genomic fragmentation between GFP$^+$ and GFP$^-$ populations using the gating strategy set at baseline. Lower left panel: clear gates at baseline. Lower right panel: cell gates are no longer clear at 400μM H_2O_2.

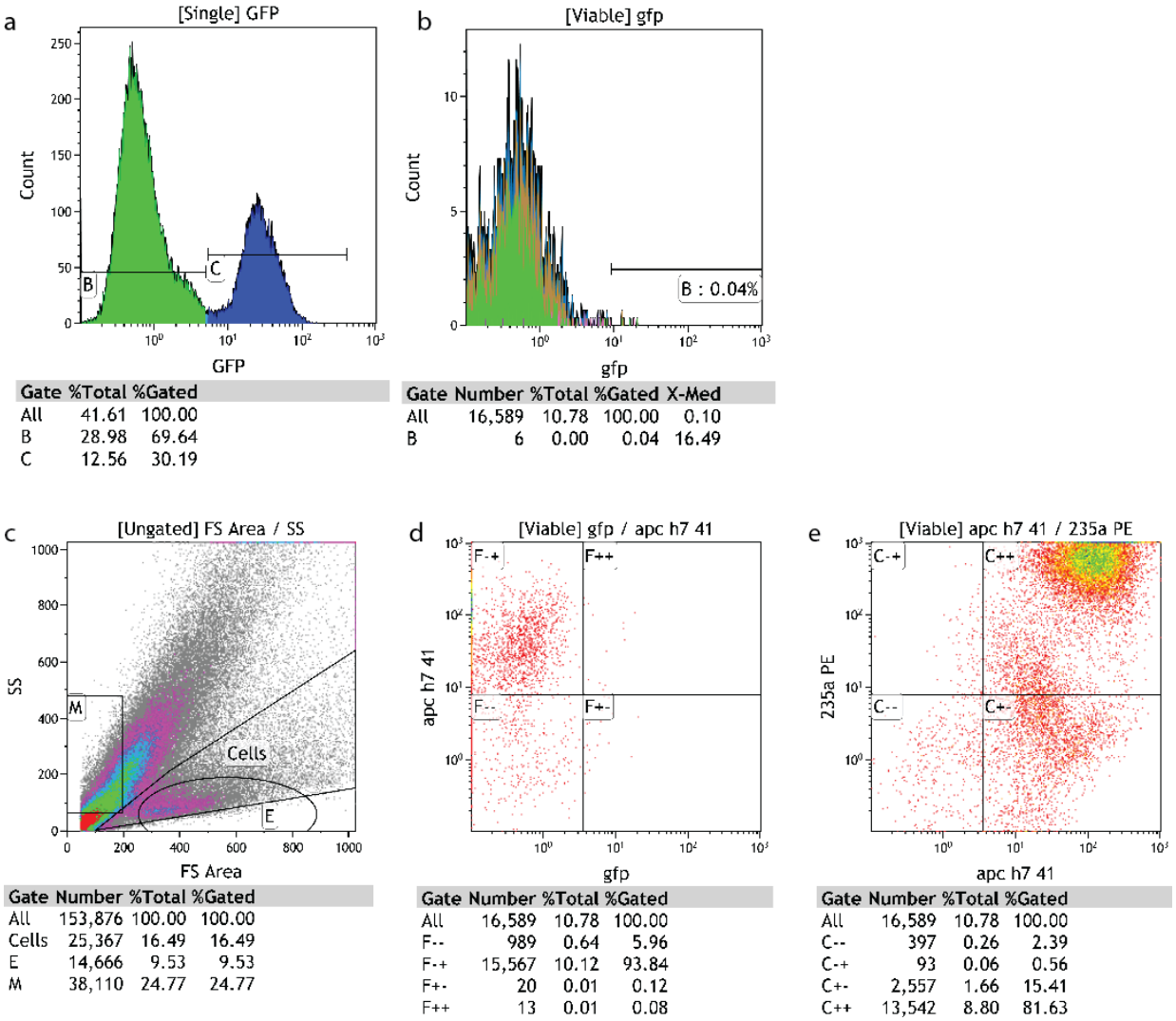


Figure 40 iMK derived from iPSC expressing the GFP-labelled KDSR ORF (KGFP$^+$) did not express the GFP$^+$ signal, suggesting silencing of the transgene.
KGFP$^+$ iPSC were forward programmed to iMK in three independent experiments.
(a) The proportion of mGFP$^+$ cells measured at day 0 FoP (i.e iPSC) was 30%, suggestive of a degree of gene silencing prior to reprogramming as only GFP$^+$ iPSC were seeded following transduction.
(b) The GFP$^+$ signal was lost by day 10.
(c-e) At day 10 the majority of events in the cell gate had successfully formed myeloid-erythroid progenitors, shown by high double positive CD41$^+$ CD235$^+$ expression, indicative of successful forward programming.

4.9.5 pLenti-EF1a-KDSR-myc-DDK-IRES-Puro expression studies

IPD123 iPSC and QOLG stock iPSC were transduced with pLenti-EF1a-KDSR-myc-DDK-IRES-Puro containing the wild-type ORF of *KDSR* transcript ENST00000591902 (to form cell line K^{resc}) and empty vector plasmid PS10085 (to form cell line K^{ev}) as described in Section 2.5.5. For clarity, the experimental set-up is illustrated in Figure 41. pLenti-EF1a-KDSR-myc-DDK-IRES-Puro carries the ORF of KDSR, fused to a myc-DDK (myc-FLAG) tag followed by an internal ribosome entry site (IRES) and a puromycin resistance element. Once again transduction was complicated by recurrent cell death and only KCl4 survived transduction and puromycin selection.

4.9.5.1 *KDSR* wild-type expression was confirmed by RTqPCR and immunofluorescence

K^{resc} showed 187-fold higher expression of *KDSR* than KCL4 in iPSC by RTqPCR using the KQ3 primer set, and two-fold higher than K^{ev} (Figure 42a). K^{resc} were compared with K^{ev} for all subsequent statistical analyses. The 97-fold difference in expression between K^{ev} and KCl4 by RTqpCR is difficult to explain. A possible explanation is that the stress of transduction even with an empty vector caused an upregulation of *KDSR*. There was no difference in expression by RTqPCR of KDSR between targeted and stock (QOLG) iPSC, therefore the QOLG experiment was terminated (Figure 42b).

Immunostaining with anti-myc antibody as described in section 2.4.1.9 showed that the majority of K^{resc} cells were positive (Figure 42c). There was no myc-staining in K^{ev} (Figure 42d), because the myc-DDK sequence in the empty vector is not preceded by a start codon. K^{resc} and K^{ev} iPSC were morphologically indistinguishable from IPD123 clones.

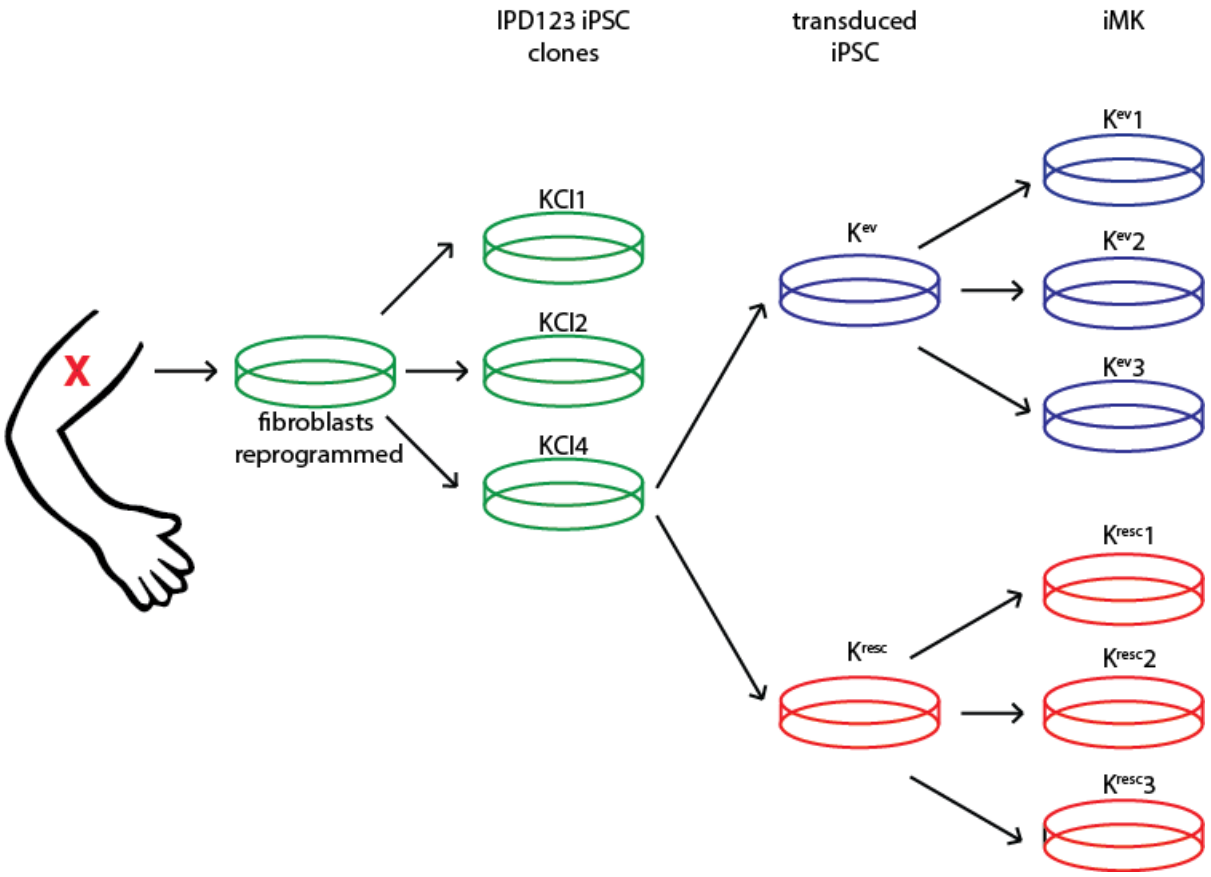


Figure 41 Experimental protocol for KDSR cellular model
Fibroblasts from the proband were differentiated to iPSC and three clones, KCl1, 2 and 4 were picked at an early stage and validated independently. KCl4 was transduced to create K^{ev} and K^{resc} iPSC, carrying an empty vector (PS100085) and the normal open reading frame of *KDSR (*pLenti-EF1a-KDSR-myc-DDK-IRES-Puro), respectively. These lines were forward programmed to iMK in three independent experiments, denoted by the suffix 1-3.

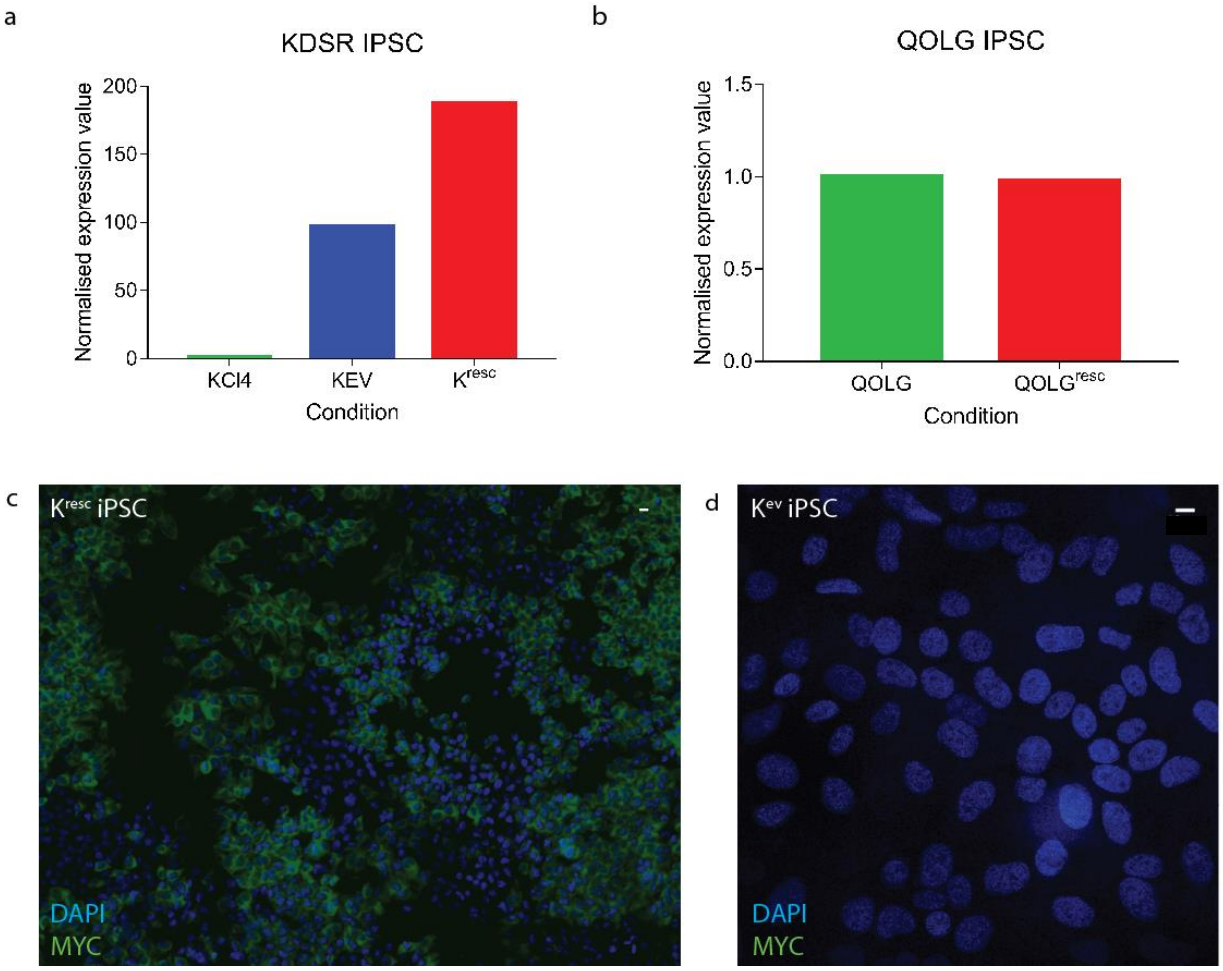


Figure 42 Confirmation of K^{resc} iPSC by RTqPCR and immunofluorescence.
An iPSC line was derived from case IPD123. Three clones were generated from a single skin biopsy, KCl1, KCl2 and KCl4. KCl4 was successfully transduced with a vector containing the normal open reading frame of *KDSR (*pLenti-EF1a-KDSR-myc-DDK-IRES-Puro) and an empty vector (PS100085) to generate K^{resc} and K^{ev}, respectively. In-house stock iPSC line QOLG was also transduced with the same vectors to attempt to generate $QOLG^{resc}$.

(a) RTqPCR normalised expression values using KQ3 primers on cDNA from KCl4, K^{resc} and K^{ev}. 187x overexpression is shown in K^{resc} iPSC. K^{ev} expression is also increased 97x relative to KCl4. Values were normalised relative to expression of housekeeping gene *GUSB*. This was performed once. Further validation by RNA-seq (section 4.9.5.10) was contradictory, suggesting similar levels of KDSR expression in the three lines at the gene-level.

(b) RTqPCR normalised expression values using KQ3 primers on cDNA from lines derived from in-house stock iPSC line QOLG. $QOLG^{resc}$ showed no increase in expression over baseline.

(c) In view of (a) K^{resc} iPSC were stained with anti-myc antibody (green, to label the myc-DDK tag) and DAPI (blue, to label nuclei). 68% are positive for myc, demonstrating that the overexpression is reasonably uniform, yet some cells are not green, suggesting puromycin-resistance in the absence of vector uptake. White bar represents 10µm.

(d) K^{ev} iPSC were myc-negative, showing that the upregulation of *KDSR* in this condition was of the native, mutant form. White bar represents 10µm.

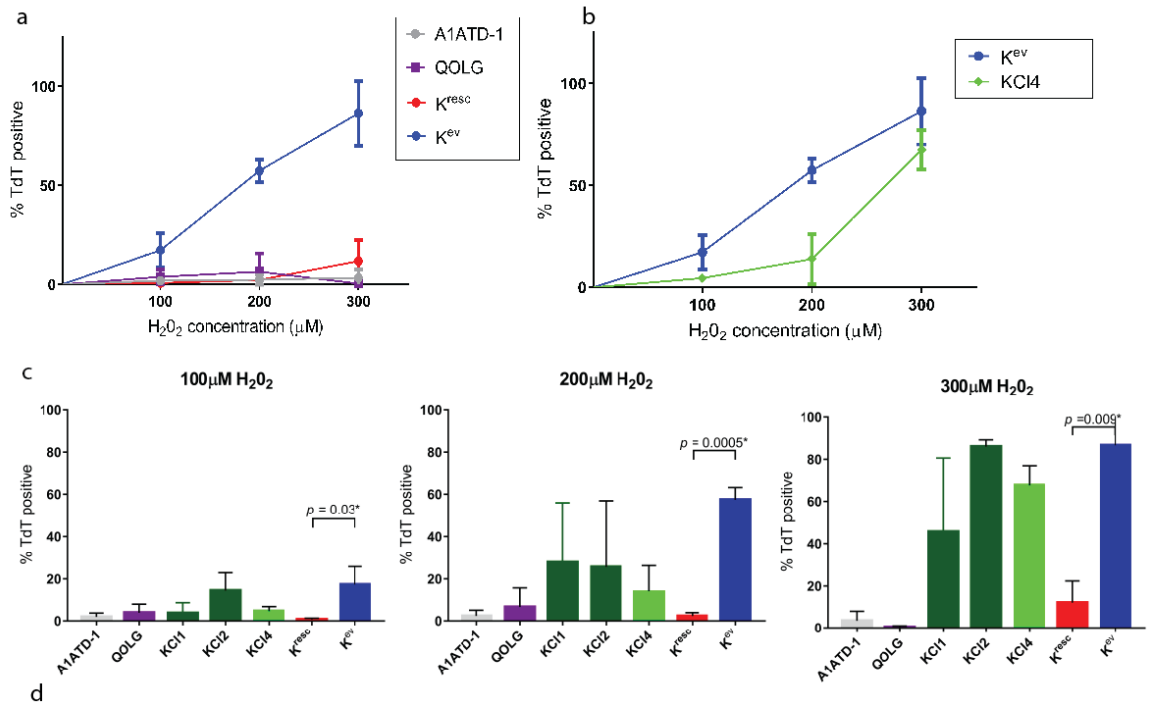
4.9.5.2 *KDSR* wild-type expression rescues the apoptotic phenotype

The TUNEL TdT assay was performed in K^{resc}, K^{ev}, KCl4 and stock lines QOLG and A1ATD-1. The experiment was repeated four times, except QOLG and KCl4 for which two and three experiments were analysed, respectively. A paired, two-tailed Student's t-test was performed at each concentration of H_2O_2 comparing K^{resc} with K^{ev}.

Figure 43a and b confirm that like KCl4, K^{ev} has an increased rate of apoptosis which is improved by lentiviral rescue to similar levels to stock iPSC. At all concentrations H_20_2 K^{resc} showed a significantly lower apoptosis rate than K^{ev} (p <0.05 in all cases, Figure 43c, d), supporting the association of *KDSR* variants with a phenotype of increased apoptosis.

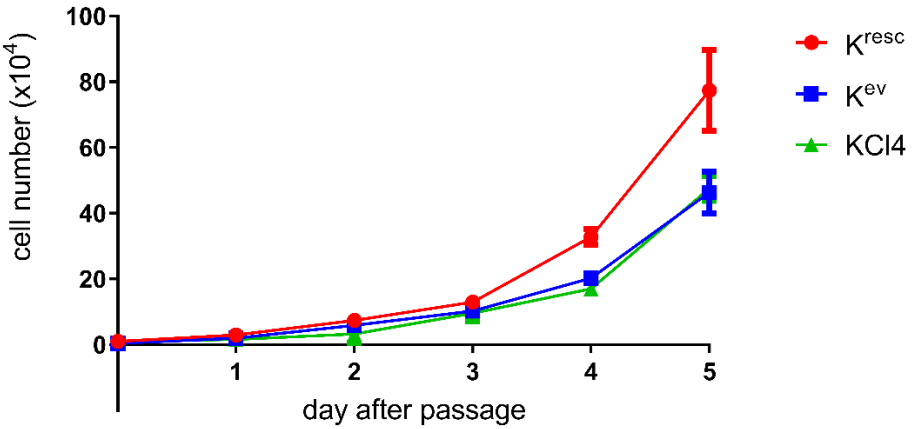
4.9.5.3 KDSR wild-type expression rescue increases the rate of iPSC proliferation

It was observed that K^{ev} required higher seeding density and less frequent passage than K^{resc}. To quantify this cell lines a growth curve was constructed for K^{ev}, K^{resc} and KCl4 as described in section 2.4.1.8. The experiment was performed three times. A paired, two-tailed Student's t-test showed that K^{resc} produced more cells than K^{ev} at 4 days (p =0.017) and 5 days (p =0.016), in further support of the proposed phenotype of increased apoptosis (Figure 44).



	Kresc			Kev			
	mean %	s.d	n	mean %	s.d	n	p
100μM	0.72	0.55	4	17.2	8.5	4	0.03*
200μM	2.3	1.53	4	57.4	5.7	4	0.0005*
300μM	11.8	10.5	4	86.4	16.3	4	0.009*

Figure 43 TUNEL TdT assay was repeated in K^{resc}, K^{ev}, IPD123 clone 4 (KCl4) and stock lines QOLG and A1ATD-1. Apoptosis was induced by incubating iPSC with hydrogen peroxide (H_2O_2). Cells were processed using a TUNEL TdT assay that labels the ends of genomic fragments produced during apoptosis.
(a, b) Like KCl4, K^{ev} has an increased rate of apoptosis which is improved by lentiviral rescue in K^{resc} to similar levels to stock iPSC. (c, d) At all concentrations of H_20_2 K^{resc} showed a lower level of apoptosis than K^{ev}.



	Kresc			Kev			
	mean %	s.d	N	mean %	s.d	n	P
Day 4	32.8	2.4	3	20.2	0.66	3	0.017*
Day 5	77.4	12.3	3	46.3	6.3	3	0.016*

Figure 44 Growth curve of iPSC.
K^{resc} showed an improved growth rate compared to K^{ev} and its parent line KCl4 that was significant at days 4 and 5. The effects were measured by paired, two-tailed Student's t-tests; $p<0.05$ was considered significant.

4.9.5.4 *KDSR* wild-type expression is maintained in iMK

K^{resc} and K^{ev} were each forward programmed three times to generate iMK. The three replicate FoP experiments are denoted by the suffix 1-3. RTqPCR confirmed wild-type rescue expression of *KDSR* at in K^{resc} iMK (Figure 45). The integrated ORF does not carry the 5'UTR sequence, and this was exploited to ensure that the difference in expression was truly due to the integrated vector, rather than clonal differences in native KDSR expression. The level of expression in K^{resc} iMK was five-fold higher than in K^{ev} iMK by RTqPCR. Figure 46 shows the morphology from each condition at day 20. The majority of cells resemble immature MK with a single nucleus. 0.1% of cells in each condition appear more mature, with a low nuclear-cytoplasmic ratio and polyploid nucleus. Representative images of the flow cytometry markers used to confirm this are shown in Figure 47 and Figure 48.

4.9.5.5 *KDSR* wild-type expression rescue is associated with improved iMK yield, viability and maturity

K^{resc} and K^{ev} were forward programmed to iMK in three independent experiments. Cells were collected at day 10 and day 20 and stained as described in section 2.4.2.1 with anti-CD41, anti-CD42 and DAPI antibodies prior to flow cytometry. At Day 20 cells were also incubated with count beads to quantify viable cells. Unstained cells were used to gate cells and spectral overlap was compensated. All results were acquired using the same protocol. All conditions started with a total of $2x10^5$ iPSC plated on day zero.

Paired, two-tailed Student's t-tests were performed to assess differences between cell lines. There were no significant differences in $CD41^+$ at either time point, or $CD42^+$ at day 10 (Figure 49a-c). However K^{resc} generated a higher proportion of $CD42^+$ cells at day 20 (p =0.018), a higher proportion of viable cells at day 10 (p = 0.004) and day 20 (p =0.021) and higher yield at day 20 (p = 0.001) (Figure 49d-g). These results suggest that *KDSR* variants present in the proband affect MK maturity and viability. However these are not independent measures, as viability will influence maturity and yield. It is also worthwhile to note that average iMK yield was very low compared to the $1x10^5$ iMK per starting iPSC reported in the literature using a similar protocol (6.7 iMK and 0.8 iMK per starting iPSC for K^{resc} and K^{ev}, respectively) [123]. The reasons for this are unclear but may be attributable to differences in the protocol or to biological differences in the viability of this line during reprogramming.

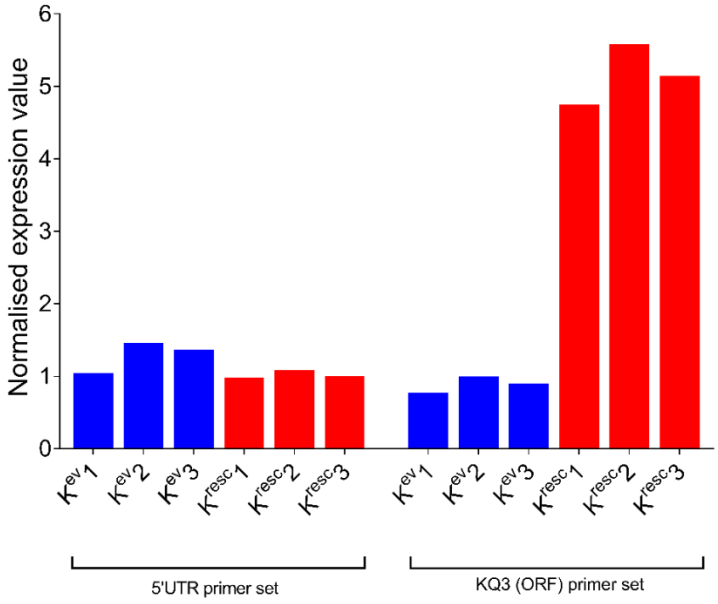


Figure 45 Normalised expression values by RTqPCR in iMK
K^{ev} and K^{resc} iPSC were forward programmed in three independent experiments as described in Figure 41 to generate K^{ev} 1-3 and K^{resc}1-3. Each replicate was plated in triplicate for RTqPCR and mean expression is plotted. Left: the 5'UTR primer picks up mRNA derived from the native genome of IPD123. There was similar expression across experiments. Right: the KQ3 primer set detects mRNA derived from native or vector-derived *KDSR*. Expression was approximately 5-fold higher in K^{resc} replicates.

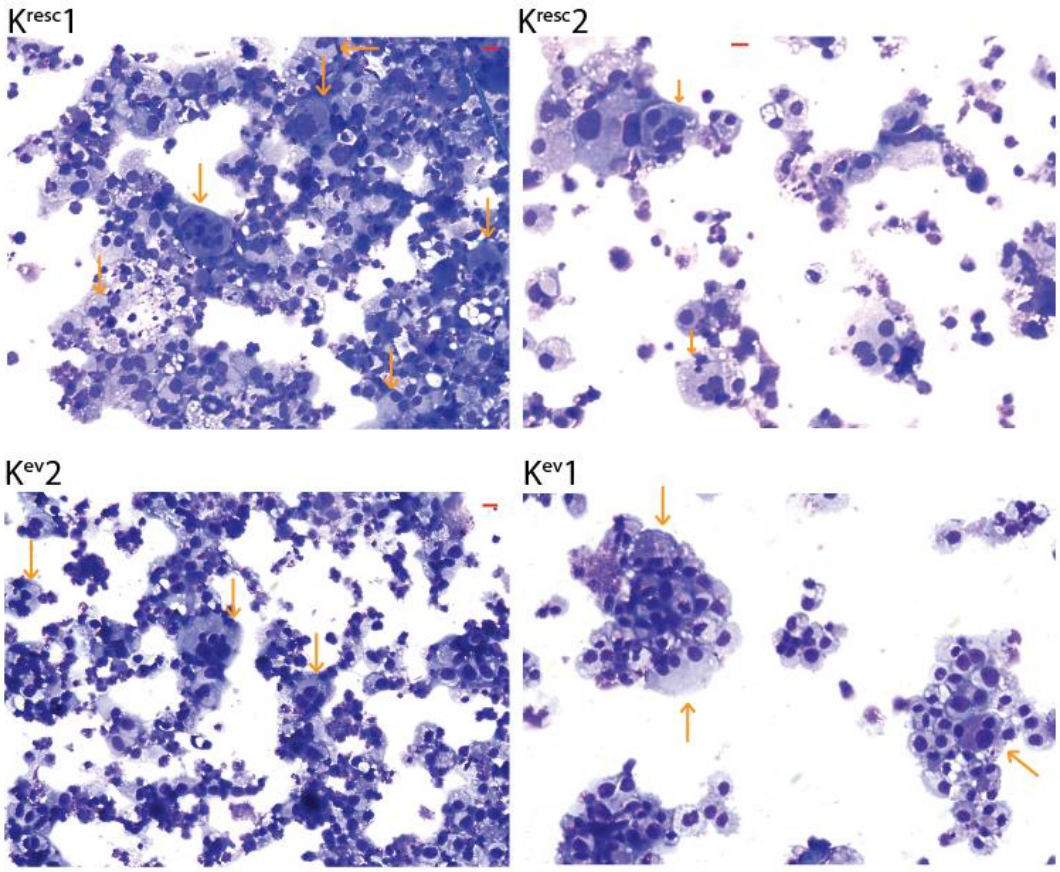


Figure 46 Morphology of iMK was similar in both K^{ev} and K^{resc}.
Images show cytospins of K^{ev} and K^{resc} iMK, stained by the Romanowsky stain. Representative images are shown. Arrows denote more mature forms that resemble *in vivo* MK with a low nuclear-to-cytoplasmic ratio and high ploidy.

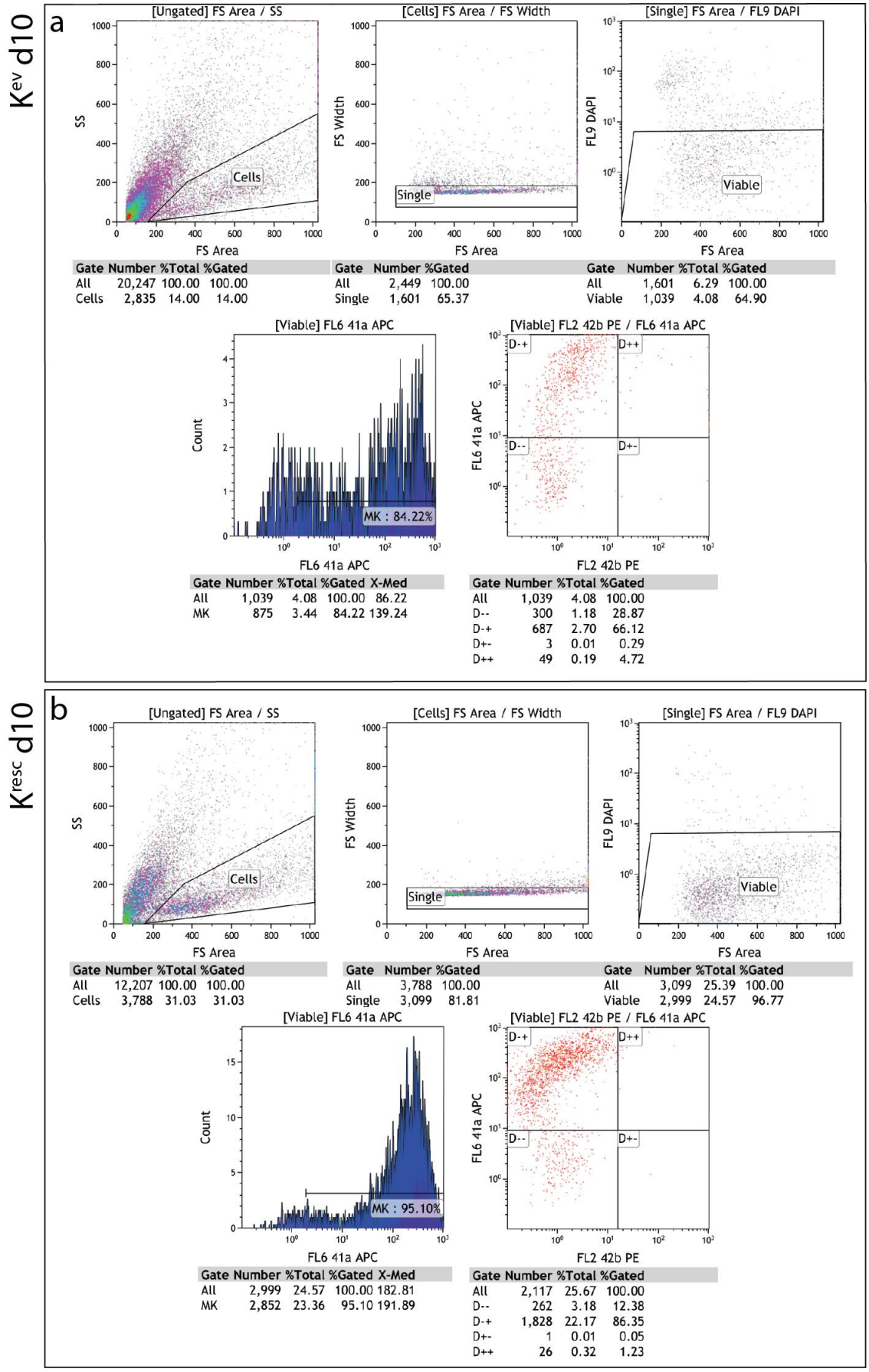


Figure 47 Examples of representative flow cytometry of $K^{ev}1$ and $K^{resc}1$ iMK at day 10.
During forward proramming at day 10, single, viable cells were assessed by flow cytometry for markers of iMK. In both conditions, as in other cell lines, the majority of cells were $CD41^{+}$ $CD42^{-}$, in keeping with differentiation to myeloid-erythroid progenitors. Viability was assessed by lack of DAPI uptake.

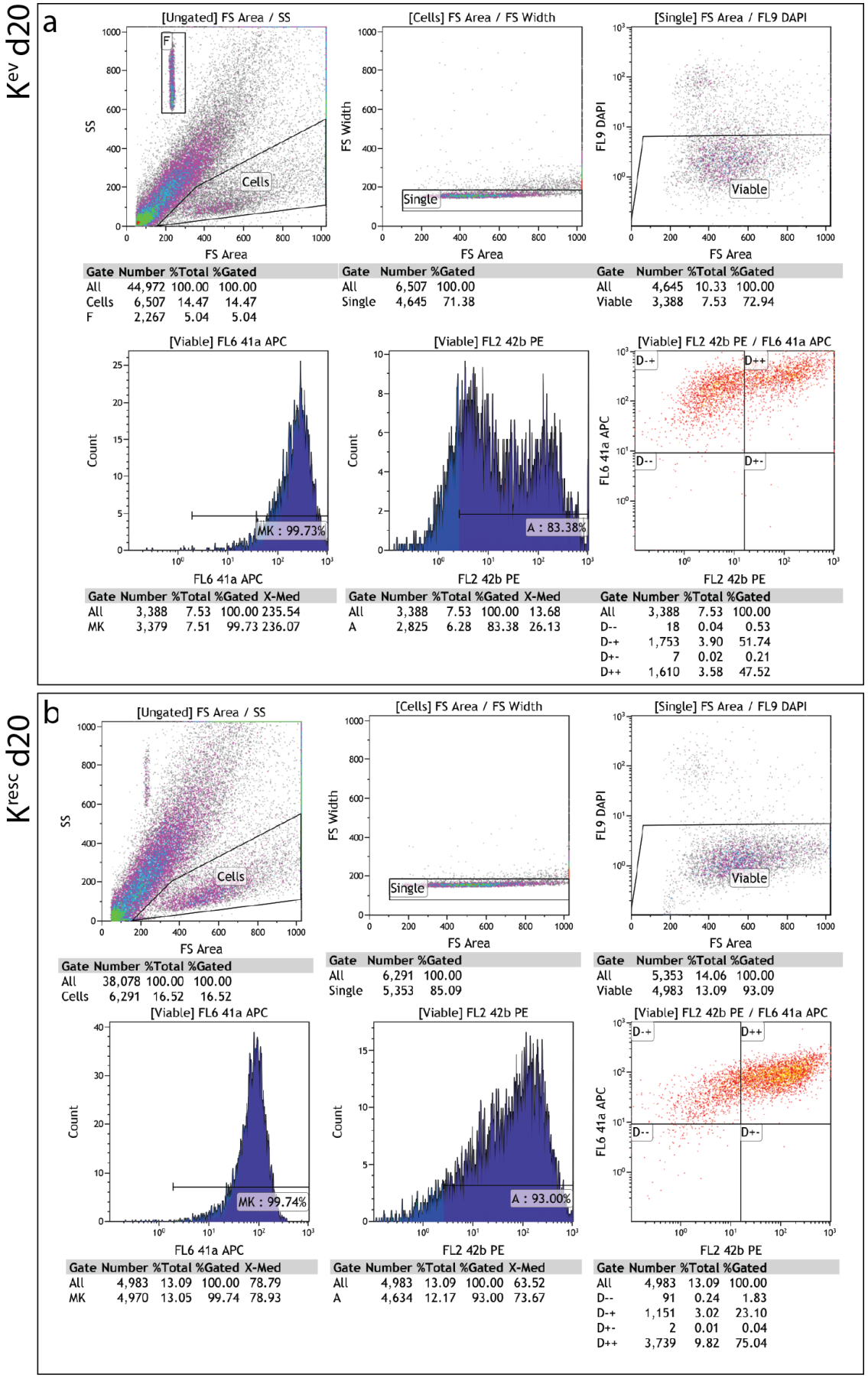
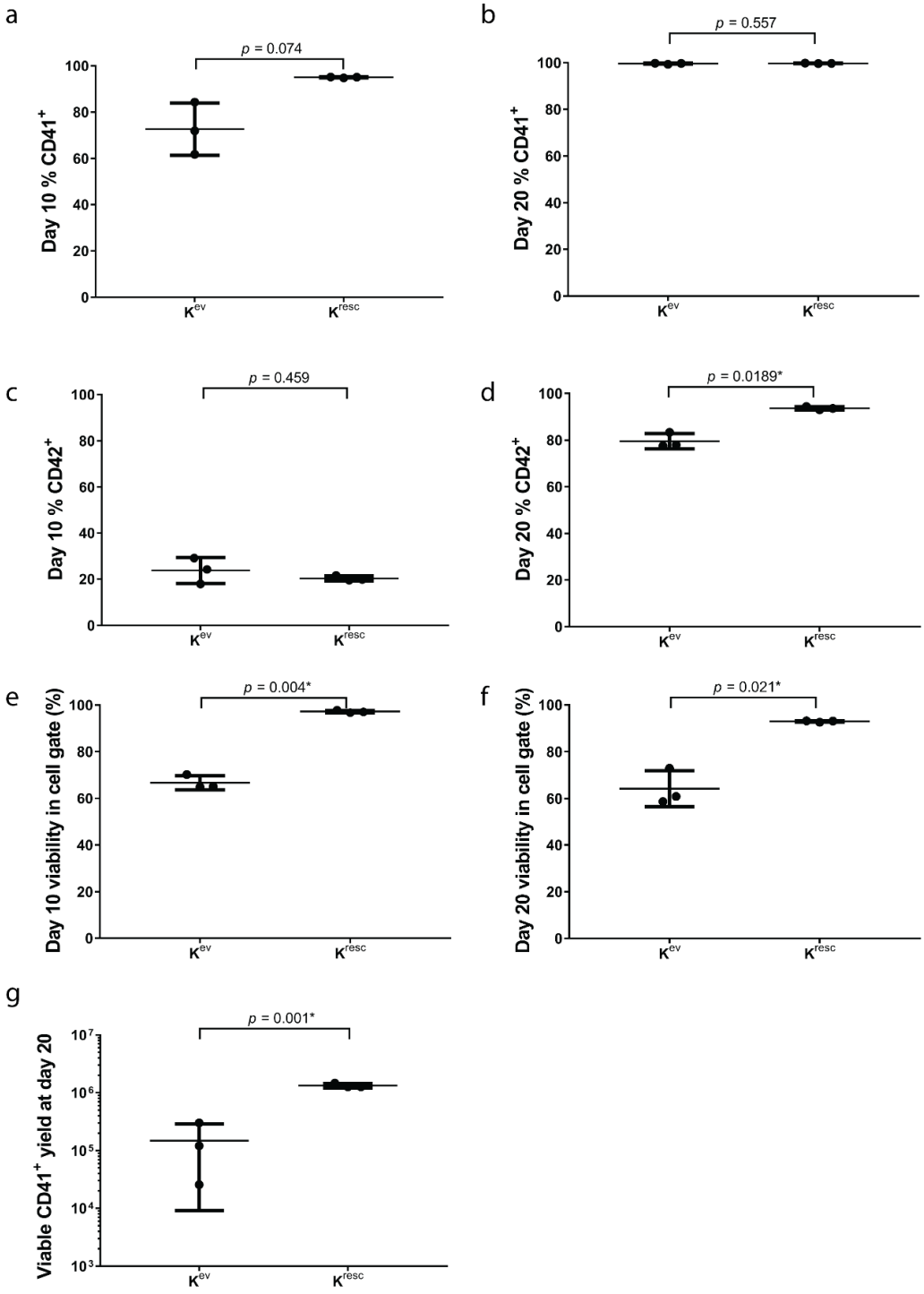


Figure 48 Examples of representative flow cytometry of $K^{ev}1$ and $K^{resc}1$ iMK at day 20.
During forward proramming at day 20, single, viable cells were assessed by flow cytometry for markers of iMK. In both conditions, as in other cell lines, the majority of cells were double positive $CD41^{+}$ /$CD42^{+}$ at this stage, in keeping with differentiation to iMK. Viability was assessed by lack of DAPI uptake.



		Kresc		Kev		
	Day	mean %	s.d	mean %	s.d	p
CD41+	10	94.93	0.2	72.6	11.2	0.074 (ns)
	20	99.7	0.1	99.6	0.2	0.0557 (ns)
CD42+	10	20.3	1.1	23.8	57	0.459 (ns)
	20	93.7	0.7	79.5	3.3	0.019*
Viability	10	97.2	0.6	66.7	3.0	0.004*
	20	93.0	0.3	64.2	7.6	0.021*
Yield	20	1.33E6	1.21E5	1.50E5	1.40E5	0.001*

Figure 49 (previous page) K^{resc} and K^{ev} iMK specific markers, viability and yield K^{resc} and K^{ev} were FoP to iMK in three experiments. Cells were collected at day 10 and day 20 and stained with antibodies to MK markers CD41 and CD42 and viability marker DAPI prior to flow cytometry. At Day 20 cells were also incubated with count beads to count viable cells. The effects were measured by paired, two-tailed Student's t-tests, $p<0.05$ was considered significant. The K^{resc} condition generated a higher proportion of $CD42^+$ cells at day 20 (d), of viable cells at day 10 and day 20 (e, f) and a higher yield at day 20 (g) than the K^{ev} condition.

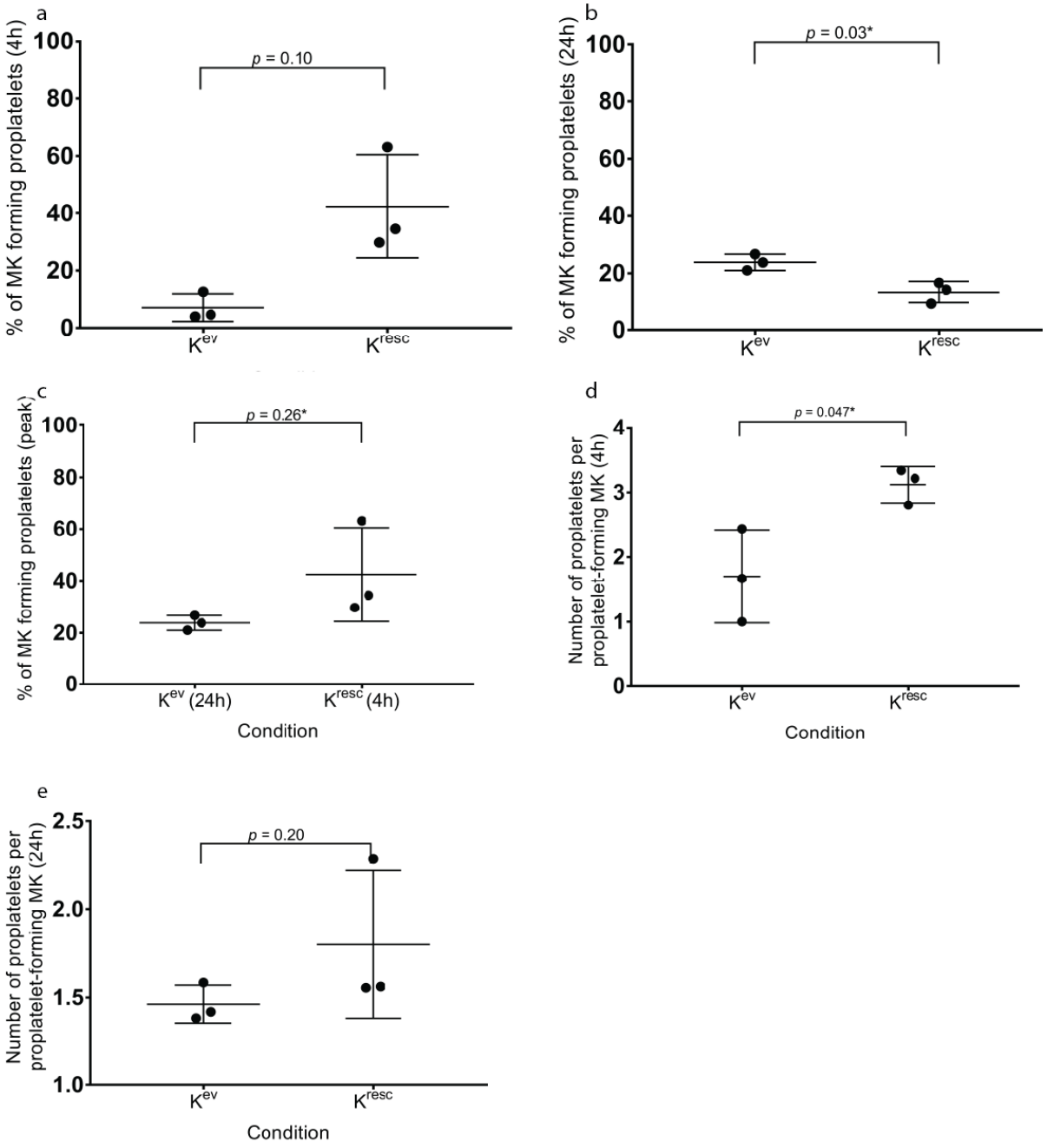
4.9.5.6 *KDSR* wild-type expression rescue is associated with increased proplatelet formation at an earlier timepoint

To further assess the contribution of *KDSR* variants to the phenotype in this model a proplatelet assay was performed on each experimental replicate, as described in section 2.4.2.2. Five to ten representative fields were examined to ensure at least 100 iMK were observed in each condition. I performed paired, two-tailed Student's t-tests of each measure and considered a *p*-value <0.05 to be significant.

The mean number of proplatelet-forming MK (PPFMK) was assessed and findings are summarised in Figure 50a-c. The % of PPFMK at 4 hours was higher for K^{resc} than K^{ev} but the difference was not statistically significant ($p = 0.10$). At 24 hours this pattern was reversed, with K^{resc} producing fewer PPFMK than K^{ev} ($p = 0.03$). There was no difference in the peak % of PPFMK ($p = 0.26$).

To explore this further the number of proplatelets formed per PPFMK was measured (Figure 50d-e). K^{resc} PPFMK produced significantly more proplatelets at 4 hours ($p = 0.047$), with no difference at 24 hours ($p = 0.2$). These data suggest that K^{resc} PPFMK produce more proplatelets than K^{ev} and at an earlier time-point. This suggests that *KDSR* variants in this proband cause thrombocytopenia through a mechanism of impaired proplatelet formation.

The morphology of the iMK shown in Figure 51 and Figure 52 support the quantitative findings. K^{resc} iMK are large and produce multiple proplatelets at 4 hours. By 24 hours there is very little cytoplasm remaining and evidence of significant fragmentation, while K^{ev} proplatelet formation is peaking.



		Kresc		Kev		
	Time	mean %	s.d	mean %	s.d	p
%PPFMK	4h	42.4	18.	7.0	4.8	0.10
	24h	13.4	3.8	23.8	2.8	0.03*
Peak PPFMK		42.4	18. 0	23.8	2.8	0.26 (ns)
Proplatelet no. per PPFMK	4h	3.1	0.3	1.7	0.7	0.047*
	24h	1.8	0.4	1.5	1.4 6	0.2 (ns)

Figure 50 (previous page) Proplatelet assay of K^{ev} vs. K^{resc} iMK.
High purity iMK from each replicate experiment were plated on fibrinogen and fixed at 4 and 24 hours. Cells were stained with cytoskeletal marker α-tubulin. At least 100 iMK were observed in each condition. Paired, two-tailed Student's t-tests of each measure were performed and a p value <0.05 was considered significant.

(a) There was a non-significant trend towards higher mean % number of proplatelet-forming MK (PPFMK) at 4 hours for K^{resc} than K^{ev} (42.4% vs. 7.1%, p = 0.10).

(b) At 24 hours this pattern was reversed, with K^{resc} producing fewer PPFMK than K^{ev} (p = 0.03).

(c) There was no significant difference between peak %PPFMK (p = 0.26).

(d) The number of proplatelets formed per PPFMK for K^{resc} was higher at 4 hours (p = 0.047).

(e) There was no difference in the number of proplatelets formed per PPFMK at 24 hours (p = 0.2)

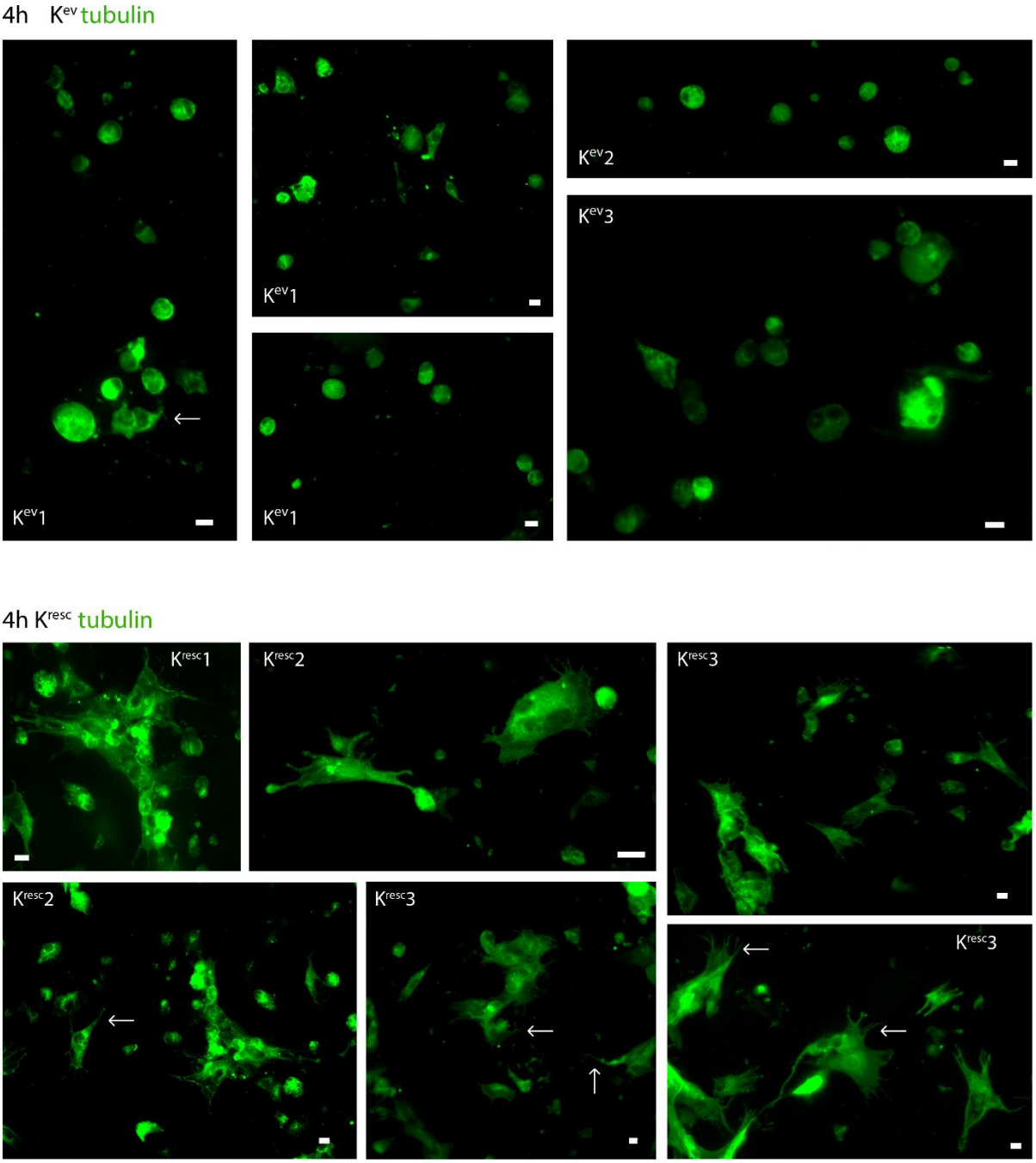


Figure 51 Proplatelet assay morphology at 4 hours.
iMK at day 20 of forward programming from each of the three experiments were seeded on fibrinogen-coated plates. Cells were inspected regularly and fized at 4 hours. Cells were stained with DAPI (blue) and anti-α-tubulin antibody (green). At least 100 iMK were observed in each condition. Representative images are shown and proplatelets are indicated by white arrows. White scale bars indicate 10µm. Upper panel shows K^{ev} iMK which appear smaller and are producing very few proplatelets. Lower panel shows K^{resc} iMK which appear larger and produce numerous proplatelets.

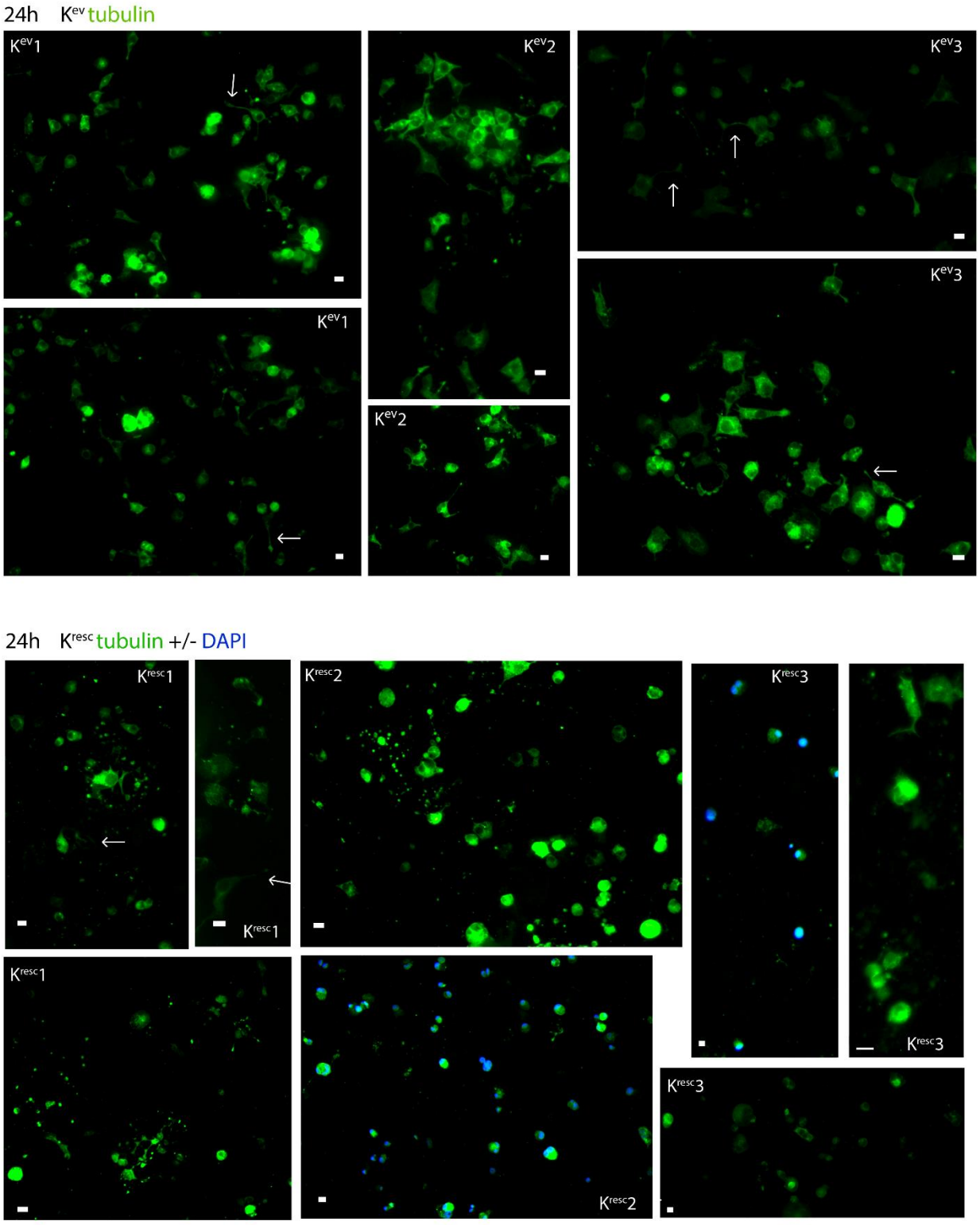


Figure 52 Proplatelet assay morphology at 24 hours
iMK at day 20 of forward programming from each of the three experiments were seeded on fibrinogen-coated plates. Cells were inspected regularly and fixed at 24 hours. Cells were stained with DAPI (blue) and anti-α-tubulin antibody (green). At least 100 iMK were observed in each condition. Representative images are shown and proplatelets are indicated by white arrows. White scale bars indicate 10µm. Upper panel shows K^{ev} iMK which appear smaller and are producing proplatelets. Lower panel shows K^{resc} iMK. The appearance is strikingly different from that at 4 hours with many iMK showing very little cytoplasm and apparent fragmentation into free proplatelets and platelet-like particles.

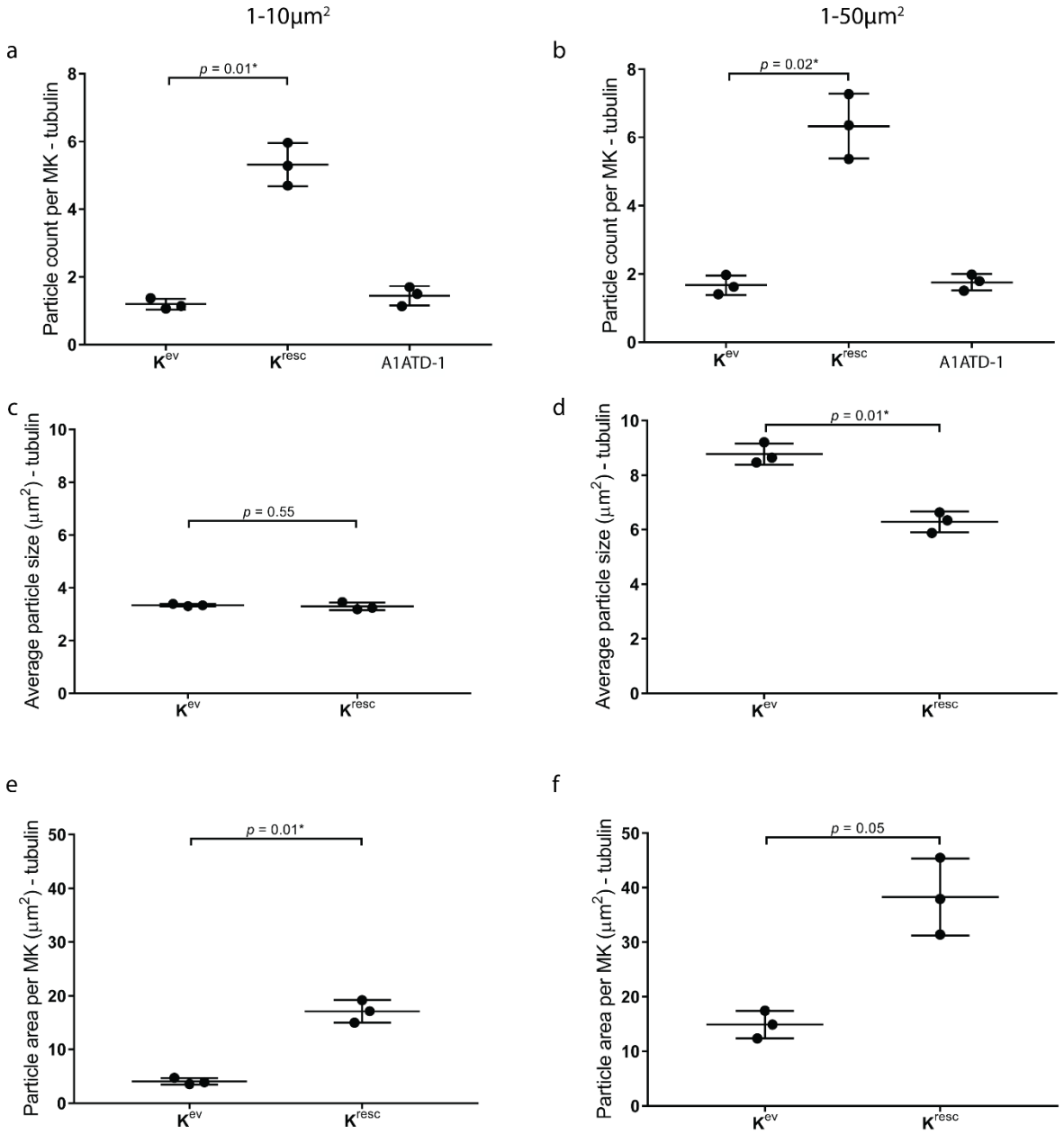
4.9.5.7 *KDSR* wild-type expression rescue is associated with increased platelet-like particle formation

To quantify platelet-like particle (PLP) formation at 24 hours, images recorded for the proplatelet assay were used for automated particle counting using the Analyze Particle plugin of the FIJI software package [134]. Several representative fields were examined for each of the three experimental replicates. Platelets *in vivo* are 9-25μm^2 and iMK nuclei measured by DAPI staining from both conditions were >200μm^2. Therefore for this analysis arbitrary size thresholds were set to count PLP that had an area between 1-10μm^2 or between 1-50μm^2. However, it should be noted that coalescent PLP cannot be differentiated from individual large PLP by this assay. Whilst unlikely that bare nuclei or small iMK were counted as particles, some large, coalescent particles may have been missed. Furthermore as this assay was performed post hoc using images from the proplatelet assay, it does not provide independent observations from that assay. There were insufficient high quality slides or residual iMK for a dedicated PLP assay.

With these limitations in mind, Figure 53 shows that K^{resc} iMK produced more PLP than K^{ev} (p <0.05 for both size thresholds), though on average K^{resc} particles were smaller (p =0.01 at 1-50μm^2 threshold, no difference at 1-10μm^2.) To adjust for this limitation, the particle area, normalised for the number of iMK in the fields examined, was calculated and was higher for K^{resc} than K^{ev}, confirming that by this assay, K^{resc} produce more PLP than K^{ev}.

Of note, K^{ev} did not successfully forward program at the same time as K^{resc}. To control for differences in reagent and plate handling a stock line (A1ATD-1) was seeded on the same plate and the results of this line did not have unexpected levels of fragmentation (Figure 53a, b). This line has been shown to produce infrequent proplatelets as shown in section 5.13.4 and as reported by collaborators in the Ghevaert lab, but was chosen as a positive control because it reprograms reliably with high cell yields.

Only replicates K^{ev}1 and K^{resc}1 provided enough slides for a further, limited examination of CD41$^+$ expression in the particles, to support their identity as PLP. Figure 54 shows the results of analysis of three fields of a single slide for each condition. There is a trend towards higher CD41$^+$ particle count and area in the K^{resc} condition.



		K^resc		K^ev		
	Area (µm²)	mean	s.d	mean %	s.d	P
Particle count per MK	1-10	5.3	0.6	1.2	0.2	0.01*
	1-50	6.3	1.0	1.7	0.3	0.02*
Mean particle size (µm²)	1-10	3.3	0.1	3.3	0.0	0.55 (ns)
	1-50	6.3	0.4	8.8	0.4	0.01*
Particle area per MK (µm²)	1-10	17.1	2.1	4.1	0.6	0.01*
	1-50	38.2	7.0	14.9	2.5	0.05

Figure 53 PLP assay of K^{resc} and K^{ev} iMK
High purity iMK from three FoP experiments were plated on fibrinogen, fixed at 24 hours and stained with DAPI and anti-α-tubulin. Automated PLP counting was performed, using size thresholds of 1-10µm² and 1-50µm². Paired, two-tailed Student's t-tests of each measure were performed and a *p* value <0.05 was considered significant.
(a,b) K^{resc} iMK produced more particles than K^{ev} replicates. A1ATD-1 was plated alongside K^{resc} and showed no excess fragmentation
(c,d) For PLP between 1-50µm², K^{resc} produced smaller particles. There was no difference in particles sized 1-10µm².
(e,f) The particle area, normalised for the number of iMK in the fields examined, was higher for K^{resc} than K^{ev}.

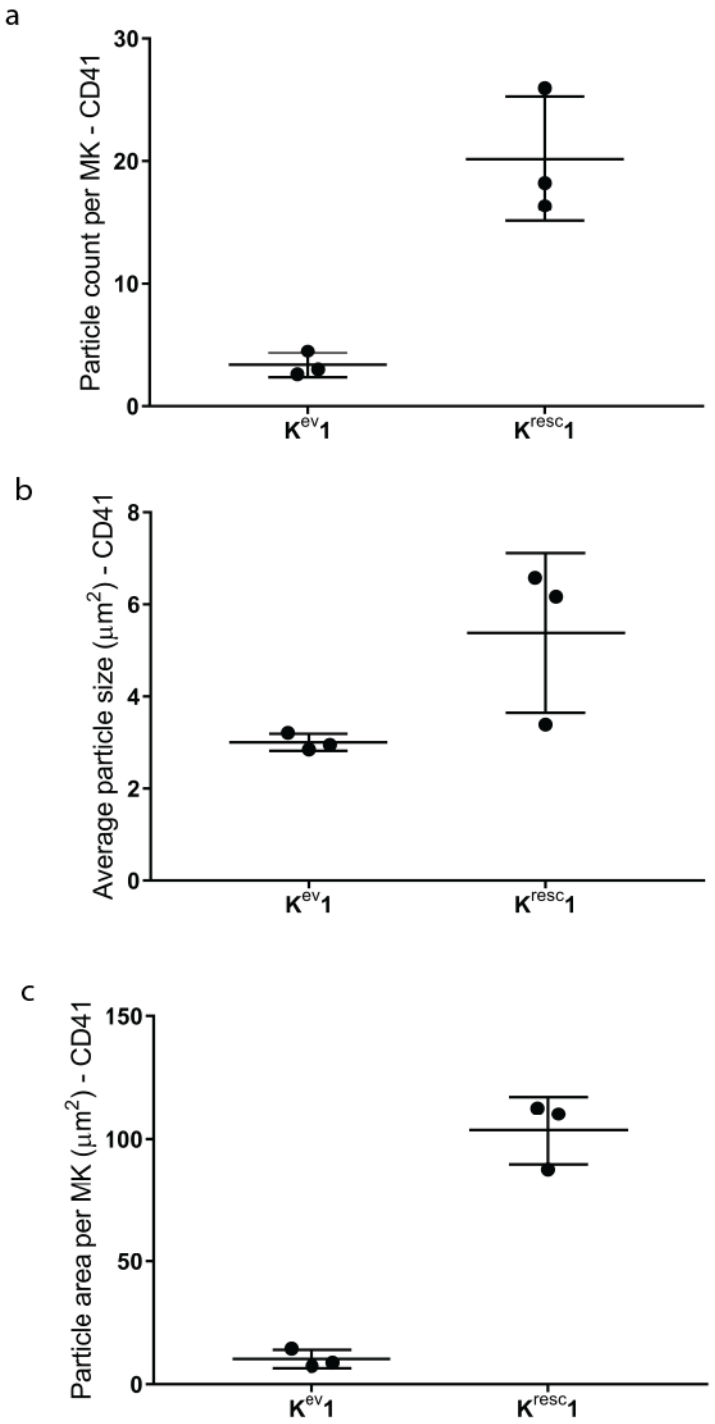


Figure 54 CD41^{+} immunostaining of PLP

Insufficient high quality slides were obtained to allow confirmation of PLP as CD41^{+} in all replicates. Here proplatelet assessment was undertaken on three fields of K^{ev}1 and K^{resc}1 iMK only, examining particles between 1-50µm^2. The results follow a similar trend in terms of a higher particle count and area for K^{resc}.

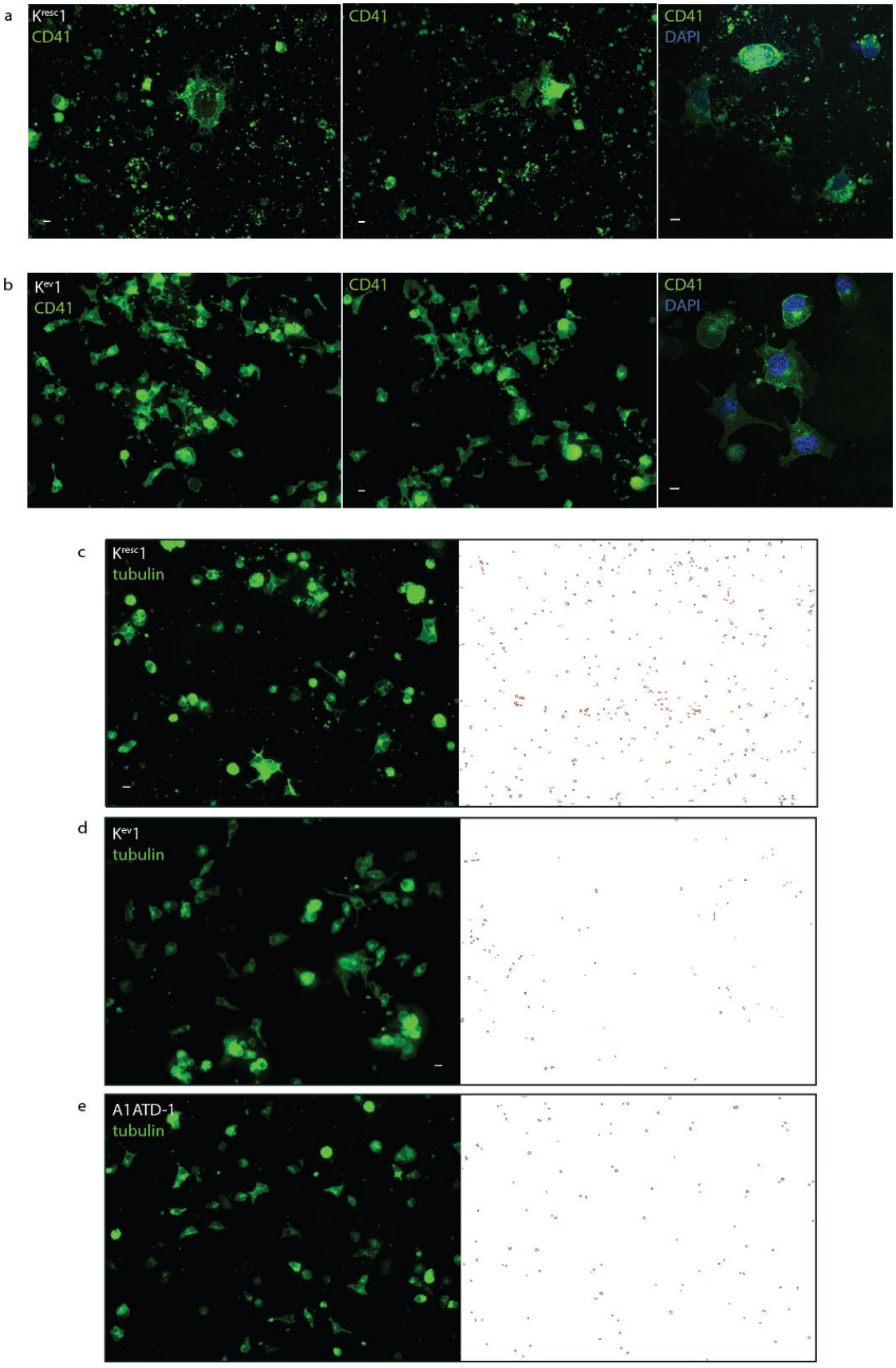


Figure 55 iMK were plated on fibrinogen to promote proplatelet and platelet-like-particle (PLP) formation. Cells were fixed at 24 hours. Immunostaining was performed using antibodies targeting CD41 (green) +/- DAPI (blue) staining or α-tubulin (green), as indicated. Automated PLP counting was performed using the Analyze Particle plugin of the FIJI software package [134]. Several fields were examined for each of the three experimental replicates and representative images are shown. These images count PLP with an area between 1-50μm^2. Higher particle numbers are seen in K^{resc} (a,c) compared to K^{ev} (b,d) and A1ATD-1 plate control (e).

4.9.5.8 *KDSR* wild-type expression rescue is associated with increased iMK size

iMK size was assessed by forward scatter-area (FSC-A) at day 10 and day 20 FoP (Figure 56). A minimum of 2000 viable singlet $CD41^+$ iMK were captured for each replicate, with the exception of K^{ev} at day 10 where a minimum of 500 events was captured due to poor yield. At day 10 there was a non-significant trend towards larger K^{resc} MEP (K^{resc} mean 486, s.d 6.2 vs. K^{ev} mean 517, s.d 7.1, p =0.053). At day 20 this was reversed with significantly larger K^{resc} iMK (mean 593, s.d 4.4) than K^{ev} iMK (mean 521, s.d 18.6, p =0.015). This data suggests that *KDSR* variants in the proband influence iMK size.

Images captured for the proplatelet assay for the three experiments were also analysed to determine size using the Analyze Particle plugin for FIJI, this time counting particles >50µm^2. This threshold was chosen following manual review of different thresholds because it captured all iMK in the field. However it is possible that some large/ coalescent, anucleate particles were picked up by this automated method. There was no difference between sizes at 24 hours, but peak size was higher for K^{resc} (p = 0.01), which is consistent with the results of the proplatelet assay.

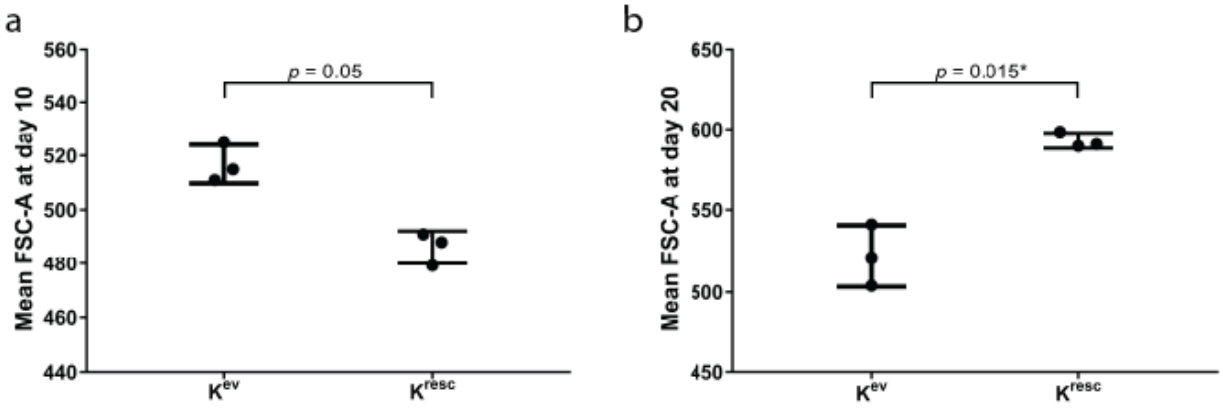


Figure 56 iMK size estimation in suspension by flow cytometry.
iMK were defined as CD41^{+} viable singlets and forward scatter-area (FSC-A) was measured. A minimum of 2000 events were captured for each replicate experiment, with the exception of K^{ev} at day 10 where a minimum of 500 events was captured. At day 10 there was a non-significant trend towards larger K^{resc} MEP (K^{resc} mean 486, s.d 6.2 vs. K^{ev} mean 517, s.d 7.1, p =0.053). At day 20 this was reversed with significantly larger K^{resc} iMK (mean 593, s.d 4.4) than K^{ev} iMK (mean 521, s.d 18.6, p =0.015).

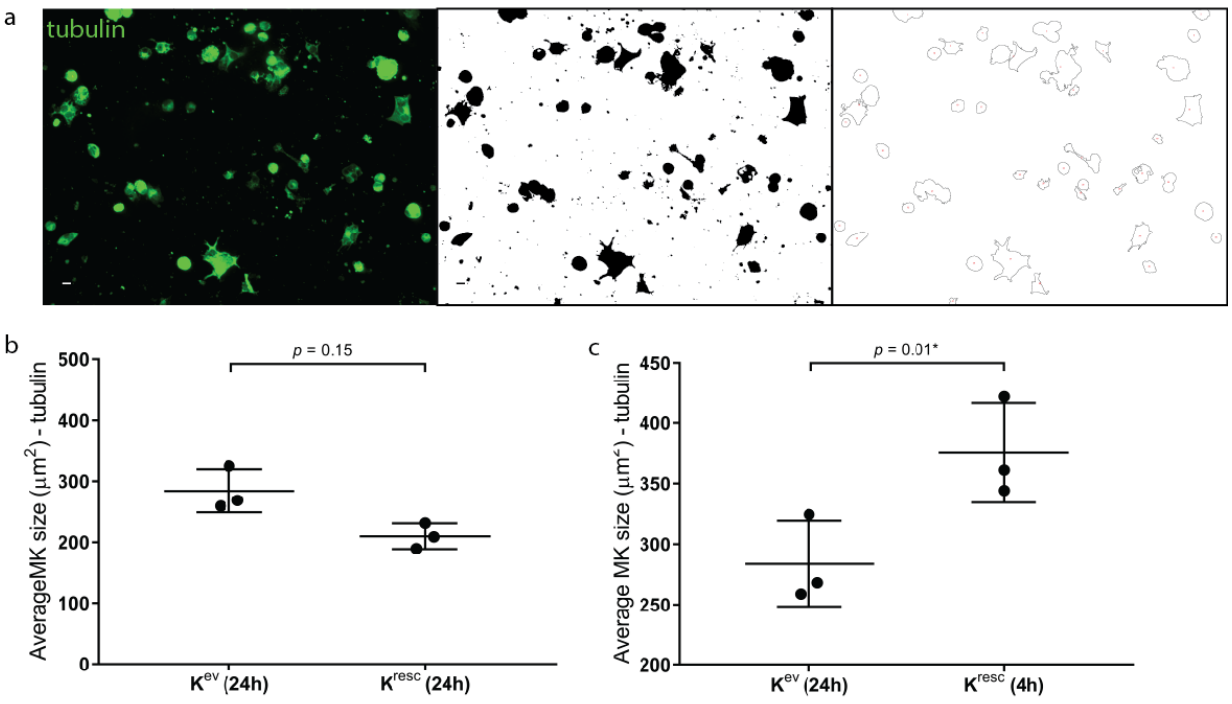


Figure 57 iMK size by immunostaining
High purity iMK from three FoP experiments were plated on fibrinogen, fixed at 24 hours and stained for DAPI and α-tubulin. At least 100 iMK were observed in each condition. The ImageJ Analyze Particle plugin was used for iMK counting, using a size threshold of >50µm^2. Paired, two-tailed Student's t-tests of each measure were performed and a p value <0.05 was considered significant.
(a) Images captured (left) were converted to a black and white image (central panel). A threshold of 50µm^2 picked up the majority of iMK and few anucleate particles (right panel). iMK are multinucleate and were often coalescent so manual counting was essential.
(b) As expected, there was no significant difference in MK size between K^{resc} and K^{ev} replicates at 24 hours.
(c) However, peak size was significantly higher for K^{resc} (at 4 hours, mean 375µm^2, s.d 40µm^2) than K^{ev} (at 24 hours, mean 284.0µm^2, s.d 35.5µm^2) (p = 0.01)

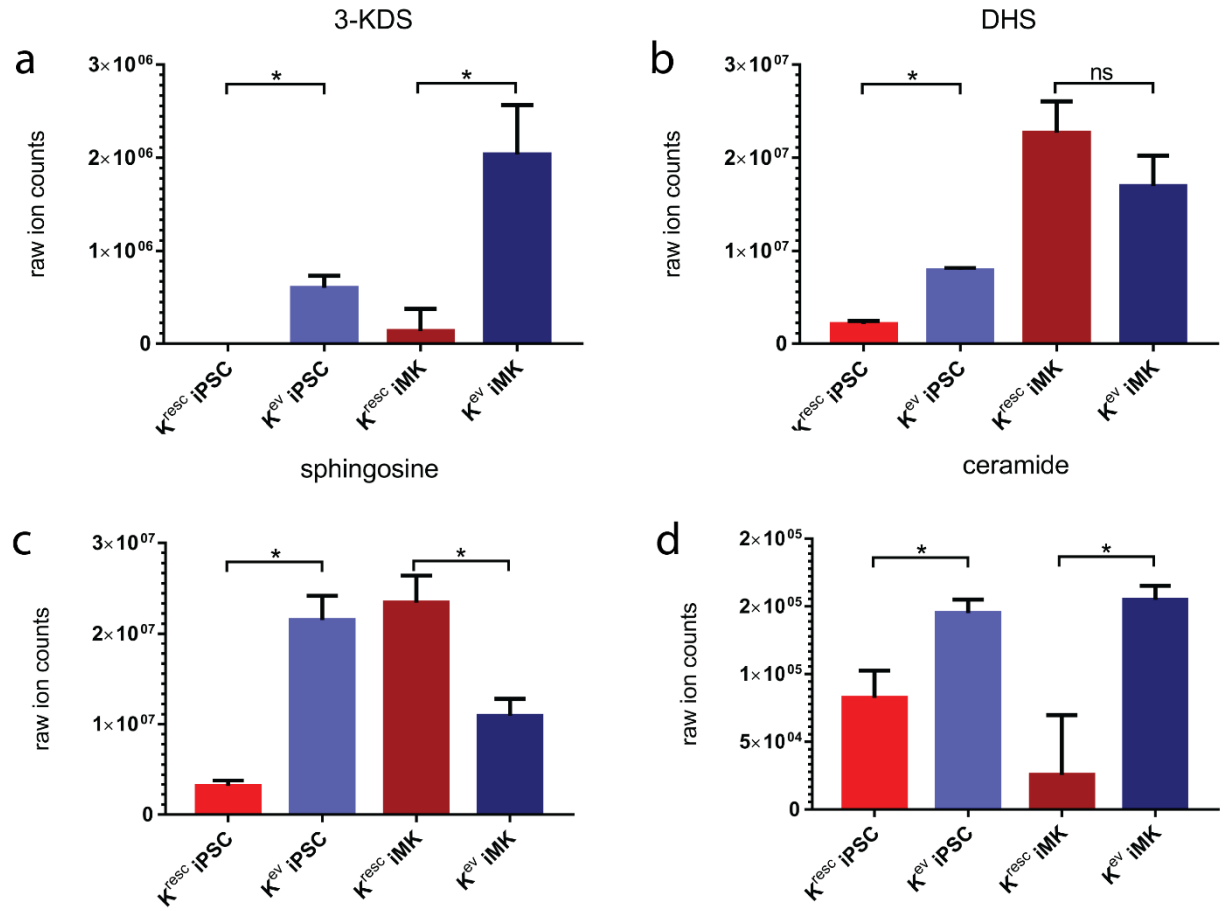
4.9.5.9 The plasma metabolome is recapitulated and corrected in iPSC by *KDSR* wild-type expression rescue

For differences in behaviour between K^{resc} and K^{ev} to be attributable to lentiviral rescue, the metabolic profile observed in plasma (Figure 33) should ideally be recapitulated in K^{ev} and corrected in K^{resc}. Therefore iPSC and iMK were analysed using the Metabolon mass spectrometry platform as described in section 2.5.6. $1x10^6$ iPSC and between $0.2\text{-}1x10^6$ iMK were analysed, and the results were normalised by cell number. Figure 58 and Figure 59 summarise the results, analysed by the paired, two-tailed Student's t-test comparing K^{resc} with K^{ev} in each cell type, for each of sphingolipid shown.

3-KDS was detected in K^{ev} iPSC and iMK. It was undetectable in K^{resc} iPSC and reduced in K^{resc} iMK (p = 0.02) showing that the key *in vivo* phenotype is replicated and corrected or significantly improved by lentiviral rescue in this model.

In iPSC DHS was also lower in K^{resc} than K^{ev} (p = 0.003) suggesting attenuation of this additional phenotype. Downstream metabolites sphingosine, ceramide, sphingomyelins, and complex glycosphingolipids were also reduced in K^{resc} iPSC suggesting global downregulation of sphingolipids in this line. The end product of sphingolipid breakdown, phosphoethanolamine, showed similar raw ion counts in both conditions, suggesting that reduced production, rather than increased breakdown, may be causing this picture.

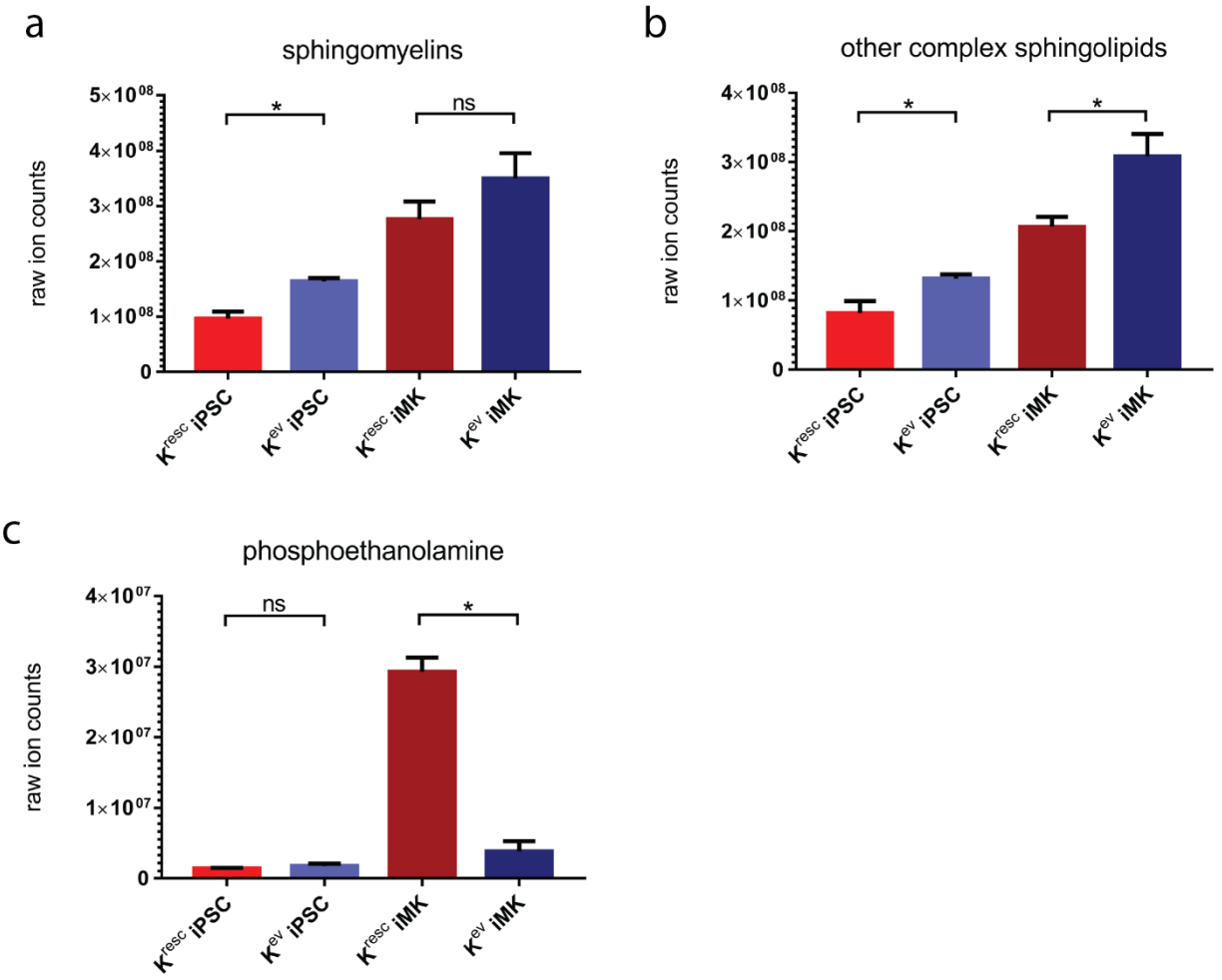
The sphingolipid profile in iMK, other than for KDS, was not so uniform. DHS was higher in K^{resc} than K^{ev} albeit not significantly (p = 0.07), which is at odds with the plasma and iPSC findings. Sphingosine was also raised and ceramide reduced in K^{resc}. This latter profile is interesting because ceramide is known to be raised in apoptosis states, whilst sphingosine-1-phosphate (S1P) is reduced, and sphingosine is the metabolic intermediate between them [179]. It is unfortunate that the sphingosine-1-phosphate and dihydrosphingosine-1-phosphate raw counts did not meet quality control criteria and therefore were not available for interpretation. It is interesting that in K^{resc} iMK breakdown product phosphoethanolamine was dramatically raised, suggesting activation of the breakdown pathway.



e

	Kresc iPSC		Kev iPSC		*p*	Kresc iMK		Kev iMK		*p*
	Mean	s.d	Mean	s.d		mean	s.d	mean	s.d	
3-KDS	0	0	5.98E+05	1.30E+05	**NA**	1.37E+05	2.38E+05	2.04E+06	5.31E+05	**0.02***
DHS	2.11E+06	3.30E+05	7.90E+06	2.51E+05	**0.003***	2.27E+07	3.37E+06	3.39E+06	6.54E+05	**0.07 (ns)**
sphingosine	3.17E+06	6.11E+05	2.15E+07	2.69E+06	**0.006***	2.34E+07	2.98E+06	2.18E+06	3.73E+05	**0.02***
ceramide	8.24E+04	2.02E+04	1.45E+05	1.01E+04	**0.09 (ns)**	2.55E+04	4.42E+04	1.55E+05	1.02E+04	**0.048***

Figure 58 Raw ion counts for K^{ev} and K^{resc} iPSC and iMK generated using the Metabolon mass spectrometry platform.
$1x10^6$ iPSC and between $0.2\text{-}1x10^6$ iMK were analysed, and the results were normalised by cell number. Results shown were analysed by the paired, two-tailed Student's t-test comparing K^{resc} with K^{ev} in each cell type, for each sphingolipid shown.
(a) 3-KDS was detected in K^{ev} iPSC and iMK. It was undetectable in K^{resc} iPSC and reduced in K^{resc} iMK (p = 0.02) showing that the key *in vivo* phenotype is replicated and corrected or significantly improved by lentiviral rescue in this model.
(b) In iPSC DHS was also lower in K^{resc} than K^{ev} (p = 0.003) suggesting attenuation of this additional phenotype. This was not replicated in iMK.
(c,d) Downstream metabolites sphingosine and ceramide were also reduced in K^{resc} iPSC suggesting global downregulation of sphingolipids in this line. In iMK the pattern was different, with higher sphingosine and lower ceramide in the rescued line.



	Kresc iPSC		Kev iPSC		*p*
	Mean	s.d	Mean	s.d	
total sphingomyelins	9.65E+07	1.23E+07	1.64E+08	5.79E+06	**0.02***
other complex sphingolipids	8.18E+07	1.72E+07	1.32E+08	5.90E+06	**0.046***
phosphoethanol-amine	1.49E+06	4.41E+04	1.82E+06	3.48E+05	**0.26 (ns)**

	Kresc iMK		Kev iMK		*p*
	mean	s.d	mean	s.d	
total sphingomyelins	2.77E+08	3.17E+07	7.00E+07	9.21E+06	**0.1 (ns)**
other complex sphingolipids	2.07E+08	1.46E+07	6.17E+07	6.53E+06	**0.041***
phosphoethanol-amine	2.93E+07	2.00E+06	7.76E+05	2.88E+05	**0.005***

Figure 59 K^{ev} and Kresc iPSC and iMK were analysed using the Metabolon mass spectrometry platform. $1x10^6$ iPSC and between $0.2\text{-}1x10^6$ iMK were analysed, and the results were normalised by cell number. Results shown were analysed by the paired, two-tailed Student's t-test comparing K^{resc} with K^{ev} in each cell type, for each sphingolipid shown.
(a,b) Sphingomyelins and complex sphingolipids, of which there are many subtypes, are several fold more abundant than KDS and DHS. Therefore the recycling of these abundant sphingolipids to generate, for example, DHS, would not be expected to alter physiological levels significantly. Therefore it is unclear what conclusions can be taken from the significant differences observed following rescue, at iPSC and iMK. (c) Phosphoethanolamine levels in rescue iMK, but not iPSC, were dramatically increased, suggesting activation of the sphingolipid degradation pathway in rescued iMK but not iPSC.

4.9.5.10 RNA-seq

The primary objectives of RNA-seq were to establish whether there was a transcriptional explanation for the unexpected metabolic profile of elevated DHS levels detected in patient plasma and iPSC, and to investigate whether the pro-apoptosis phenotype observed was associated with dysregulation of known apoptosis and growth genes.

4.9.5.10.1 KCL1-4 vs. PODX, BIMA, QOLG iPSC

RNA was extracted from KCL1, KCL2 and KCL4 iPSC and three stock hiPSC lines (QOLG, PODX and BIMA). RNA-seq libraries were prepared and sequenced as described in section 2.5.2.1. Dr. Denis Seyres and Dr. Luigi Grassi at the University of Cambridge performed read alignment using STAR [180] and read counts were obtained using featureCounts [181] software with respect to Ensembl v.70, of reference genome build GRCh37. Expression of transcripts and genes was quantified using MMSEQ and differential expression was assessed using MMDIFF [69, 70]. This package derives estimates of gene-level expression and associated uncertainty and uses a Bayesian regression-based model to derive a posterior probability of differential expression (where a posterior probability greater than 0.5 was used to declare a transcript or gene differentially expressed. The size and directionality of the effect in terms of normalised expression values is denoted by η, the estimated log fold change. 140 protein-coding genes were significantly differentially expressed in iPSC (Appendix 8.2.8). *KDSR* was not differentially expressed at gene level. This gene list was compared with a list of 1527 genes known to regulate apoptosis, identified using the Gene Ontology online database (http://geneontology.org, [182, 183]). No apoptosis-related genes were differentially expressed. The differentially expressed gene list was also checked for genes involved in sphingolipid synthesis (Figure 26 *Sphingolipid pathway*). No sphingolipid-related genes were identified.

I used the FIDEA package [184] to look for patterns in Gene Ontology (GO) terms that were enriched in these genes. This software computes the statistical significance of enrichment and corrects the resulting *p*-values using the Benjamini and Yekutieli False Discovery Rate method [185]. There was no clear enrichment for a single pathway, and no terms identified that were related to sphingolipid regulation or apoptosis.

4.9.5.10.2 K^{resc} and K^{ev} iMK

K^{resc} and K^{ev} iMK resulting from three forward programming experiments were submitted to RNA-seq and analysis using the same methodology as described for iPSC. As previously stated, the ORF used to express wild-type *KDSR* in the IPD123 line had the same sequence as transcript ENST00000591902, which has the highest expression in MK (Figure 31,[103]). RNA-seq confirmed significant differential expression of only this transcript in the K^{resc} replicates compared with K^{ev} ($p = 0.595$, Figure 60). There was no significant gene-level differential expression of *KDSR* ($p = 0.24$).

Previously it was noted in one experiment that by RTqPCR in iPSC there was significant upregulation of *KDSR* expression between KCL4 and K^{ev} (Figure 42). As shown in Figure 61a, expression by RNA-seq was similar in KCL4 and K^{ev} in triplicate experiments. Assessment of expression of the heterozygous variant 18:61018270 G>A (p.Arg154Trp) present in the proband and affected sibling showed majority wild type expression in K^{resc} (mean count G: 85.2%, A: 14.8%, p = 0.0015 by paired, two-tailed Student's t-test, Figure 61b). Variant expression in K^{ev} was consistent with heterozygosity (mean count G: 35.9%, A: 64.1%, p = 0.25). Counts were considered only from unequivocally mapped reads with a FASTQC quality index in the given position > 30. 2,654 genes were differentially expressed between K^{resc} and K^{ev} iMK. In summary, normalised *KDSR* expression at a gene-level was similar in KCl4, K^{ev} and K^{resc} iMK conditions, yet the majority of reads in the rescue experiment were of the wild-type ORF. This supports the hypothesis that phenotypic variation between the K^{ev} and K^{resc} conditions is attributable to correction of the putative pathogenic variants seen in IPD123.

4.9.5.10.2.1 Wild-type expression rescue is associated with transcriptional dysregulation of genes that regulate the level of DHS and the 'sphingolipid rheostat'

MMDIFF differential expression analysis [70] identified 2797 differentially expressed genes between K^{ev} and K^{resc} conditions. This gene list was compared with a list of genes involved in sphingolipid metabolism (Figure 26). Two sphingolipid enzymes were differentially expressed at the gene level: *ASAH1* (posterior probability of differential expression = 0.61, η = -0.67, downregulated in K^{ev}) and *CERS6* (posterior probability of differential expression = 0.774, η = +0.70, upregulated in K^{ev}). Figure 61c illustrates the potential impact of these transcriptional changes in the K^{ev}. To recap, CERS6 converts DHS to dihydroceramide. In patient-derived plasma and iPSC the DHS level was unexpectedly high, and upregulation of CERS6 in this scenario would be consistent with an appropriate response to high DHS, rather than its cause. However, given that there was no significant difference in iMK DHS levels, and that transcriptional changes were only observed in iMK, no firm conclusions can be made about the role of CERS6 in regulating DHS in the context of *KDSR* variants. Figure 60c also shows that ASAH1 and CERS6 act in opposition to regulate the balance between ceramide and sphingosine. This has been referred to as the 'sphingolipid rheostat.' This concept is based on the pro-apoptotic properties of ceramide, and anti-apoptotic, pro-growth properties of sphingosine-1-phosphate, downstream of sphingosine. Therefore the transcriptional balance in K^{ev} iMK favours a pro-apoptotic, anti-growth phenotype compared with K^{resc}, which is consistent with the observed phenotypes in iPSC. These findings are in keeping with the metabolic findings of higher sphingosine in K^{resc} iMK and higher ceramide in K^{ev} iMK (Figure 58). However once again, conclusions are limited because the apoptotic phenotype was only observed in iPSC. This phenotype was inferred but not confirmed in iMK, as whilst the yield of viable iMK was higher in the rescued line, insufficient cells were obtained from forward programming for the TUNEL TdT assay.

4.9.5.10.2.2 Wild-type expression rescue is not associated with transcriptional dysregulation of other genes that regulate apoptosis

The list of 2797 differentially expressed genes between K^{ev} and K^{resc} iMK was compared with a list of 1527 genes known to regulate apoptosis, identified using the Gene Ontology online database [182, 183, 186]. 116 genes were present in both lists. 110/116 of these genes were confirmed by literature searching to be associated with apoptosis and were manually annotated as pro- or anti-apoptotic. To support the experimental observations, it was expected that the great majority of pro-apoptotic genes would be higher in the K^{ev} condition, and that anti-apoptotic genes would be higher in the K^{resc} condition. However the expected directional effect was only observed in 54/110 of genes (49%, Appendix 8.2.9). This neither supports nor refutes the hypothesis that *KDSR* variants mediate increased apoptosis.

4.9.5.10.2.3 Wild-type expression rescue is associated with transcriptional dysregulation of genes that regulate the cell cycle, cytoskeletal organisation and MK differentiation

A GO term enrichment analysis was performed again using FIDEA [184, 185]. 439 GO terms had a corrected *p*-value < 0.05 (**Table 18**). No apoptosis terms or synonyms were identified in this list. Several high-ranking terms specifically relate to regulation of mitosis and cell cycle checkpoint entry, consistent with regulation of growth (highlighted green in **Table 18**); contributory differentially-expressed genes are listed in Appendix 8.2.10). Several terms related to protein localisation in the ER were also identified (highlighted orange in **Table 18**; contributory differentially-expressed genes are listed in Appendix 8.2.11). This is of interest because I and others have shown that KDSR localises in the ER. A single term related to MK differentiation was identified (GO:0045652, *regulation of megakaryocyte differentiation*, corrected $p = 0.015$). Contributory genes to this term were *MEF2C, HMGB2, KDM1A, PF4, TESC, PRMT1,* none of which are known to be directly regulated by sphingolipids. These gene-lists provide a starting point for further investigation of the mechanism by which *KDSR* variants may influence growth, differentiation and the cytoskeleton.

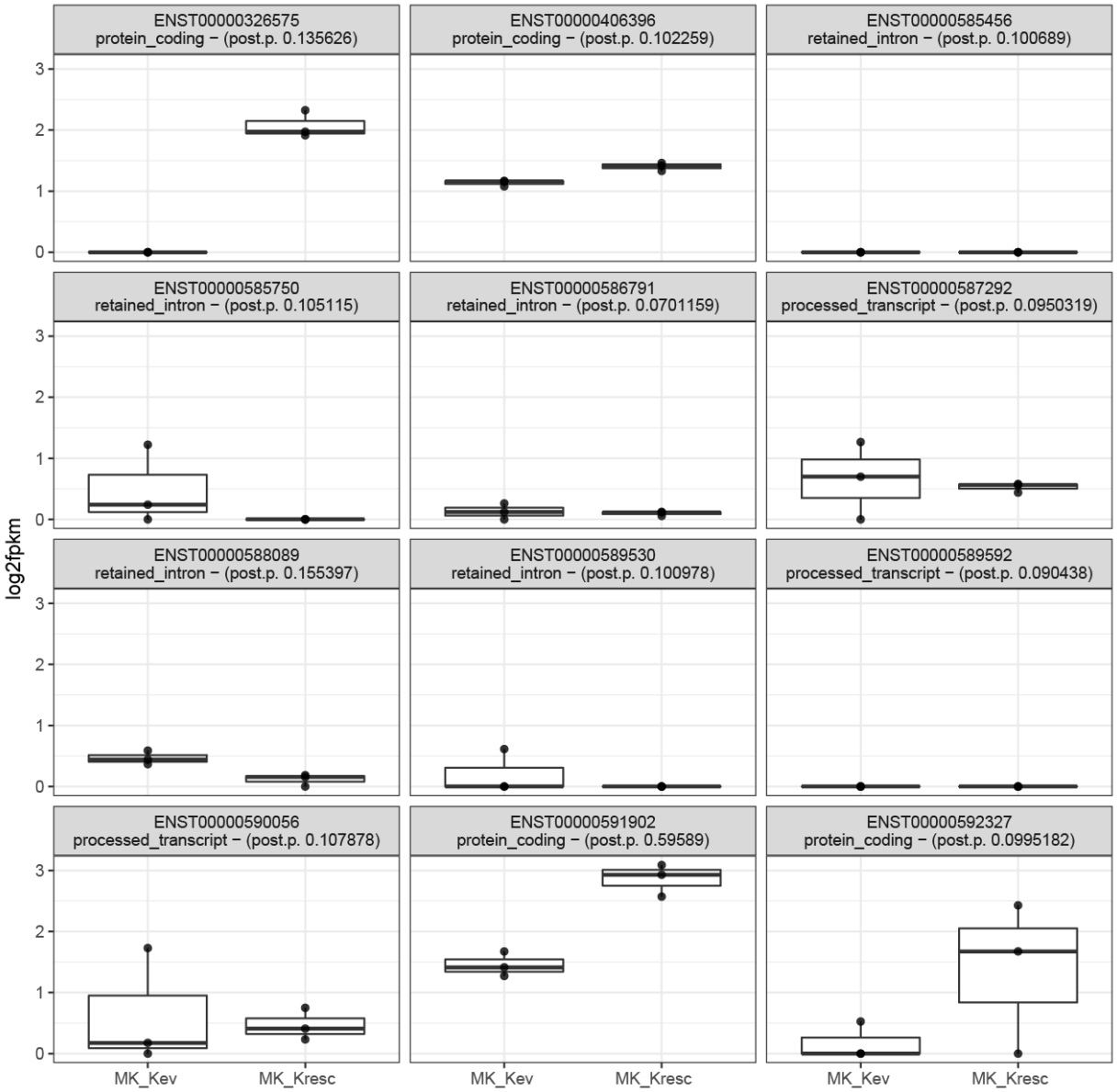


Figure 60 Box plots show differences in normalised RNA expression (log_2fpkm) by RNA-seq between K^{ev} and K^{resc} iMK.

Data are from three independent forward programming experiments. The only significantly differentially expressed transcript was ENST00000591902 (lowest, middle panel), which corresponds to the transcript used as the ORF in the lentiviral expression vector. At a gene-level there was no differential expression of *KDSR* between conditions. 'post p' refers to the posterior probability of differential expression, and a value >0.5 was considered significant.

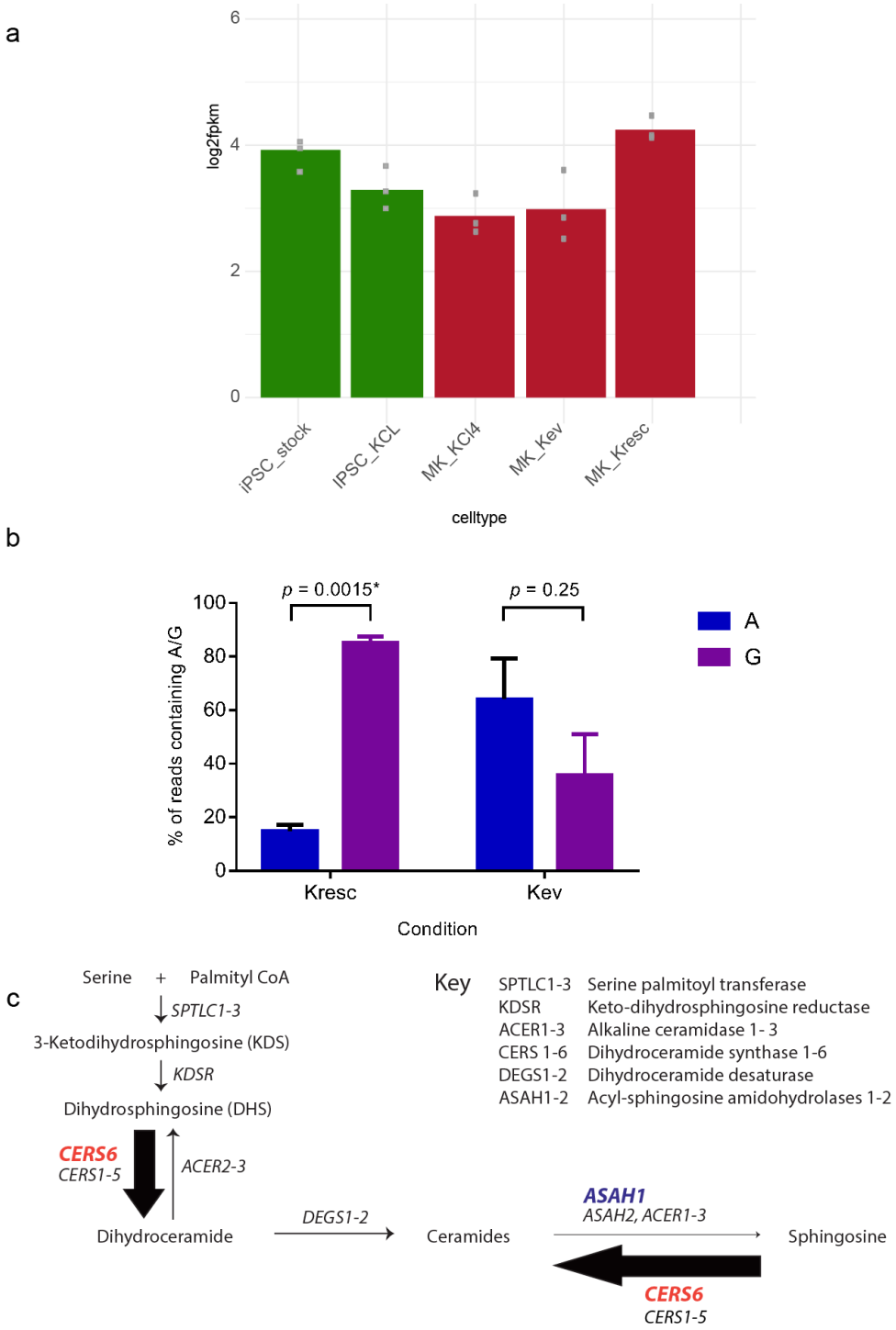


Figure 61 Effect of wild-type expression rescue by RNA-seq.

a. Normalised KDSR expression at a gene-level was similar between conditions at iPSC (green bars) and iMK (red bars).
b. K^{resc} iMK showed majority wild type expression of variant 18:61018270 G>A (p.Arg154Trp) (mean count G: 85.2%, A: 14.8%, p = 0.0015 by paired, two-tailed Student's t-test). K^{ev} iMK expression was consistent with heterozygosity (mean count G: 35.9%, A: 64.1%, p = 0.25).
c. Two sphingolipid enzymes were differentially expressed: *ASAH1* (p = 0.61, ɳ = -0.67, downregulated in K^{ev}) and *CERS6* (p = 0.774, ɳ = +0.70, upregulated in K^{ev}). CERS6 converts DHS to dihydroceramide. ASAH1 and CERS6 act in opposition to regulate the balance between ceramide and sphingosine. This has been referred to as the 'sphingolipid rheostat.'

Table 18 (below) 100 highest-ranking Gene Ontology (GO) terms identified by FIDEA analysis [182],[183] The software computes the statistical significance of GO term enrichment and corrects the resulting *p*-values using the Benjamini and Yekutieli false discovery rate method). The number of genes that are differentially expressed and annotated with the GO term in question is shown. Terms related to cell cycle progression were enriched and are highlighted green. Terms related to protein localisation in the ER were also enriched and are highlighted orange. The lists of genes contributing to the highest-ranked cell cycle term and ER localisation term are listed in Appendices 8.2.10 and 8.2.11.

GO ID	NAME_CATEGORY	ADJUSTED_PVALUE	#GENES
GO:0022403	cell cycle phase	3.82E-66	225
GO:0000278	mitotic cell cycle	2.06E-63	208
GO:0022402	cell cycle process	2.59E-58	236
GO:0007049	cell cycle	1.75E-51	262
GO:0006412	Translation	7.61E-51	132
GO:0006259	DNA metabolic process	3.39E-49	204
GO:0044237	cellular metabolic process	7.97E-43	881
GO:0034621	cellular macromolecular complex subunit organization	1.63E-41	163
GO:0000279	M phase	2.80E-38	140
GO:0000087	M phase of mitotic cell cycle	3.66E-38	113
GO:0000280	nuclear division	2.17E-37	111
GO:0007067	Mitosis	2.17E-37	111
GO:0006413	translational initiation	1.42E-36	70
GO:0048285	organelle fission	1.54E-35	112
GO:0044710	single-organism metabolic process	1.92E-35	736
GO:0008152	metabolic process	2.02E-35	919
GO:0006414	translational elongation	5.16E-35	63
GO:0006415	translational termination	5.96E-35	57
GO:0046483	heterocycle metabolic process	3.90E-34	566
GO:0034641	cellular nitrogen compound metabolic process	5.90E-34	579
GO:0043624	cellular protein complex disassembly	7.87E-34	60
GO:0006807	nitrogen compound metabolic process	1.01E-33	583
GO:0043241	protein complex disassembly	3.05E-33	60
GO:0006725	cellular aromatic compound metabolic process	3.95E-33	563
GO:0019080	viral genome expression	5.51E-33	55
GO:0019083	viral transcription	5.51E-33	55
GO:0000184	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	7.57E-33	64
GO:0071841	cellular component organization or biogenesis at cellular level	1.16E-32	454
GO:0006260	DNA replication	1.78E-32	80
GO:1901360	organic cyclic compound metabolic process	1.93E-32	578
GO:0044238	primary metabolic process	2.29E-32	860
GO:0006614	SRP-dependent cotranslational protein targeting to membrane	3.48E-32	60
GO:0034623	cellular macromolecular complex disassembly	3.88E-32	64
GO:0006613	cotranslational protein targeting to membrane	6.02E-32	60
GO:0045047	protein targeting to ER	6.02E-32	60
GO:0072599	establishment of protein localization to endoplasmic reticulum	6.02E-32	60
GO:0032984	macromolecular complex disassembly	1.13E-31	64
GO:0006139	nucleobase-containing compound metabolic process	1.15E-31	543
GO:0051325	Interphase	1.35E-31	108
GO:0051329	interphase of mitotic cell cycle	2.21E-31	107
GO:0019058	viral infectious cycle	4.01E-29	61

GO:0051301	cell division	5.37E-29	116
GO:0070972	protein localization to endoplasmic reticulum	1.30E-28	60
GO:0043933	macromolecular complex subunit organization	2.16E-28	201
GO:0071842	cellular component organization at cellular level	2.16E-28	430
GO:0000956	nuclear-transcribed mRNA catabolic process	1.89E-27	72
GO:0044260	cellular macromolecule metabolic process	5.37E-27	680
GO:0006401	RNA catabolic process	8.05E-27	79
GO:0071704	organic substance metabolic process	8.05E-27	604
GO:0006612	protein targeting to membrane	9.59E-27	66
GO:0006402	mRNA catabolic process	3.61E-26	73
GO:0072594	establishment of protein localization to organelle	4.43E-26	75
GO:0006996	organelle organization	9.87E-26	298
GO:0006261	DNA-dependent DNA replication	2.26E-25	45
GO:0006281	DNA repair	5.85E-25	104
GO:0071840	cellular component organization or biogenesis	1.00E-24	492
GO:0010564	regulation of cell cycle process	1.23E-24	114
GO:0090304	nucleic acid metabolic process	5.18E-24	470
GO:0016071	mRNA metabolic process	9.09E-24	142
GO:0071845	cellular component disassembly at cellular level	1.07E-23	77
GO:0051276	chromosome organization	1.21E-23	142
GO:0022411	cellular component disassembly	1.89E-23	77
GO:0071156	regulation of cell cycle arrest	2.11E-23	83
GO:0006974	response to DNA damage stimulus	5.05E-23	133
GO:0051320	S phase	7.08E-23	55
GO:0044249	cellular biosynthetic process	1.26E-22	477
GO:0000084	S phase of mitotic cell cycle	2.28E-22	54
GO:0000075	cell cycle checkpoint	3.17E-22	77
GO:0009058	biosynthetic process	4.06E-22	509
GO:0016043	cellular component organization	2.32E-21	468
GO:0043170	macromolecule metabolic process	3.35E-21	703
GO:0044265	cellular macromolecule catabolic process	9.33E-21	138
GO:0000236	mitotic prometaphase	9.44E-21	42
GO:0000082	G1/S transition of mitotic cell cycle	1.13E-20	61
GO:0009057	macromolecule catabolic process	6.11E-20	145
GO:0022616	DNA strand elongation	9.51E-20	27
GO:0022415	viral reproductive process	2.33E-19	116
GO:0006271	DNA strand elongation involved in DNA replication	2.37E-19	26
GO:0044248	cellular catabolic process	5.33E-19	217
GO:0071103	DNA conformation change	5.37E-19	61
GO:0033365	protein localization to organelle	9.34E-19	96
GO:0006310	DNA recombination	1.48E-18	62
GO:0006605	protein targeting	4.47E-18	92
GO:0044267	cellular protein metabolic process	6.79E-18	352
GO:0034622	cellular macromolecular complex assembly	1.43E-17	101
GO:0051726	regulation of cell cycle	3.09E-17	140
GO:0009056	catabolic process	4.10E-17	245
GO:0006323	DNA packaging	6.13E-17	52
GO:0016032	viral reproduction	8.13E-17	136
GO:0048610	cellular process involved in reproduction	1.00E-16	100

GO:0071824	protein-DNA complex subunit organization	2.85E-16	50
GO:0007059	chromosome segregation	3.16E-16	49
GO:0006333	chromatin assembly or disassembly	4.86E-16	48
GO:0071822	protein complex subunit organization	1.22E-15	135
GO:0065004	protein-DNA complex assembly	1.61E-15	46
GO:0022613	ribonucleoprotein complex biogenesis	2.59E-15	67
GO:0009987	cellular process	5.12E-15	1168
GO:0000216	M/G1 transition of mitotic cell cycle	1.62E-14	37
GO:0044281	small molecule metabolic process	2.11E-14	292
GO:0042254	ribosome biogenesis	2.62E-14	50

(Legend for table at top)

4.9.5.10.3 Colocalisation studies in K^{resc} lines

Kihara *et al.* [161] first reported that KDSR is located in the ER in HeLa cells. In section 4.9.4.1 I presented data that confirmed that in HEK293T this is also the case. It is not known whether this is also true in human stem cells or MK. To evaluate this in iPSC, K^{resc} and K^{ev} iPSC lines were stained as described in section 2.4.1.9. Figure 62 and Figure 63 show that FLAG-tagged KDSR in K^{resc} colocalises with ER-specific markers calreticulin and calnexin, respectively.

There was minor carryover of signal from anti-calnexin/ Alexa Fluor 488 into the anti-FLAG/ Alexa Fluor 555 channel which is shown in Figure 63, panel 2. There was no carryover of other signals sued for immunostaining and secondary-only stains showed minimal non-specific staining. FLAG antigen is not expressed in the K^{ev} line, so K^{ev} was used as a negative control for FLAG and myc staining; there is no start codon between the EF1a promoter and the myc-DDK tag in the empty vector PS100085. Unfortunately in iMK anti-FLAG staining was not strong or reproducible and anti-calreticulin and anti-calnexin stains were negative, therefore no conclusions can be made about the location of KDSR in iMK.

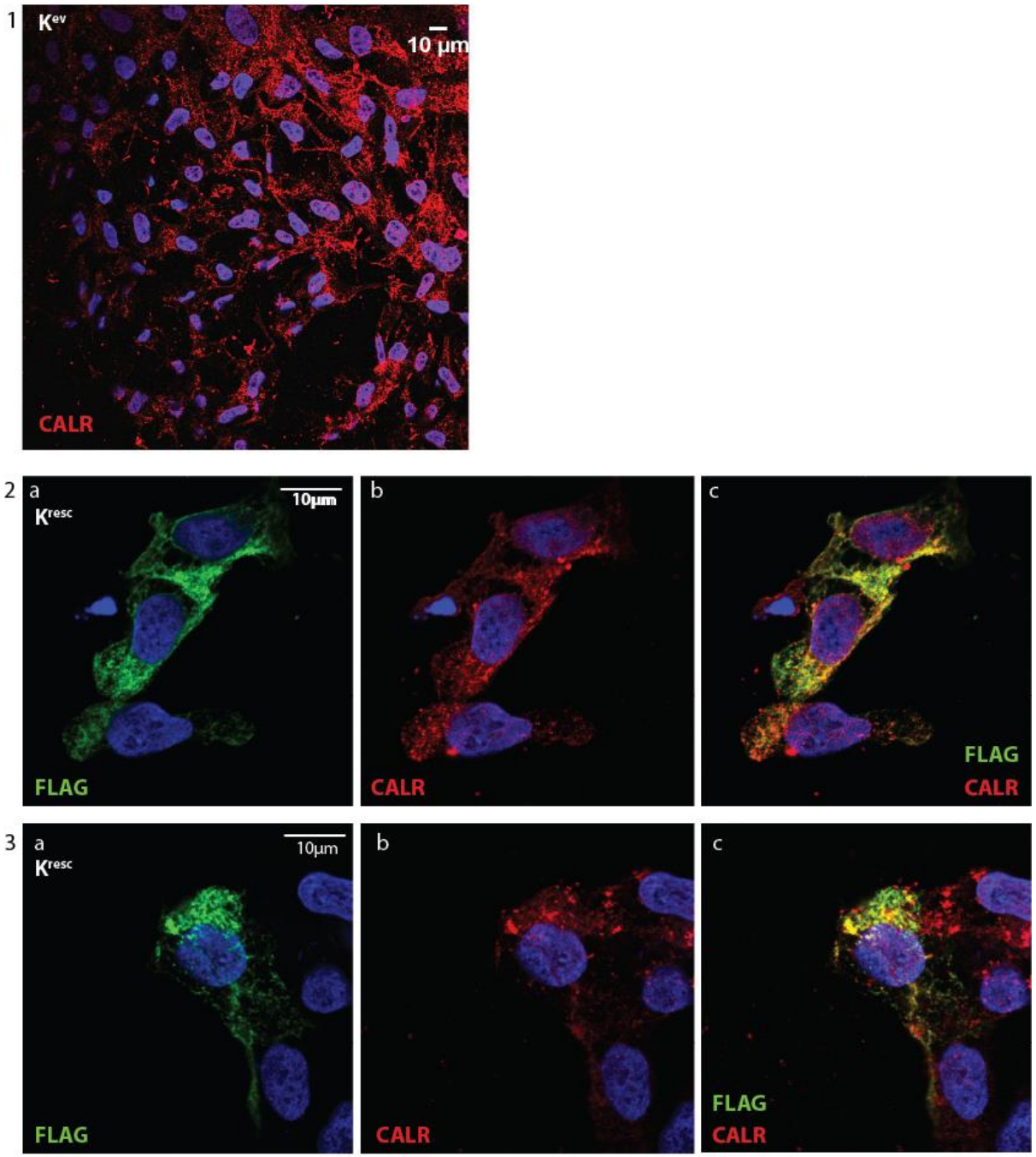


Figure 62 FLAG-tagged KDSR in K^{resc} colocalises with ER-specific marker calreticulin.
iPSC were plated on vitronectin and fixed prior to achieving 80% confluence. Cells were stained with antibodies targeting FLAG antigen (green, labelling the KDSR ORF in wild-type rescued iPSC, on which there is a MYC-DDK label), calreticulin (red, CALR, which is located in the endoplasmic reticulum) and DAPI (blue, staining the cell nucleus). Representative images are shown.

1. K^{ev} iPSC show staining of calreticulin (red) but no FLAG signal, as expected.
2. K^{resc} iPSC staining shows areas of colocalisation between KDSR ORF-FLAG and CALR in all three cells
3. K^{resc} iPSC staining shows areas of colocalisation between KDSR ORF-FLAG and CALR in three cells. The fourth cell shows only CALR staining but no FLAG, further demonstrating the specificity of FLAG for cells transduced with the vector. This cell acquired puromycin-resistance and/ or silenced KDSR-FLAG expression in culture.

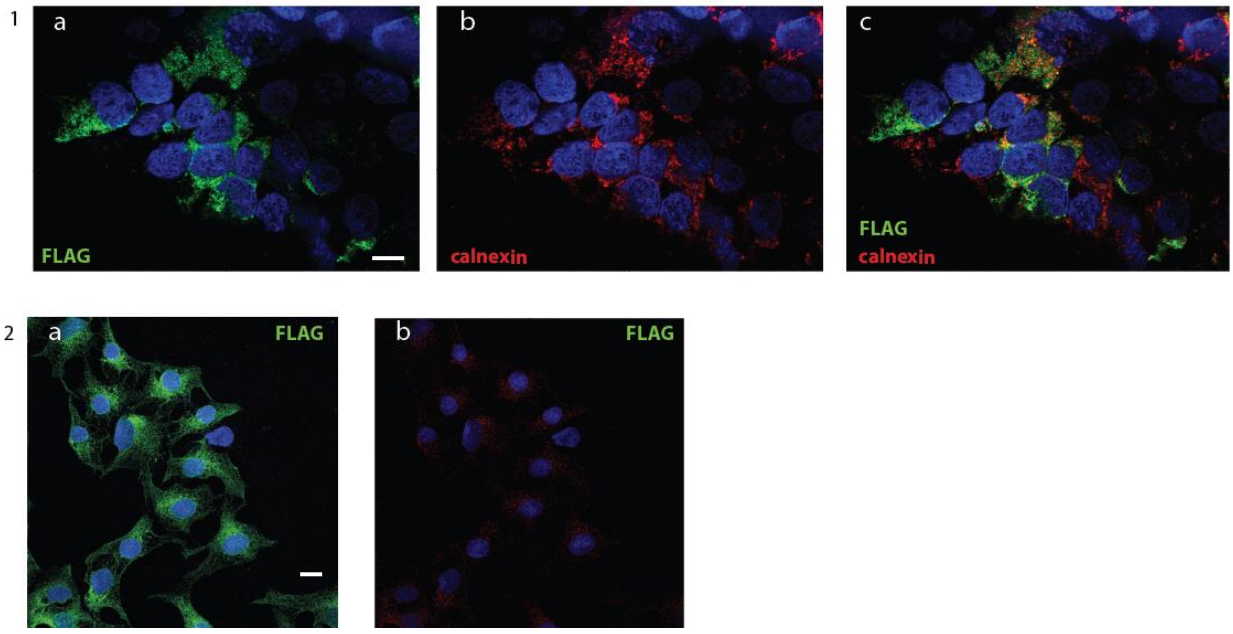


Figure 63 FLAG-tagged KDSR in K^{resc} colocalises with ER-specific marker calnexin
iPSC were plated on vitronectin and fixed prior to achieving 80% confluence. Cells were stained with antibodies targeting FLAG antigen (green, labelling the KDSR ORF in wild-type rescued iPSC, on which there is a MYC-DDK label), calnexin (red, which like calreticulin is located in the endoplasmic reticulum) and DAPI (blue, staining the cell nucleus). Representative images are shown. White bar indicates 10µm.

1. K^{resc} iPSC showing areas of colocalisation in FLAG-positive cells. Once again there are FLAG-negative, calnexin-positive cells in view, providing a negative control.
2. Dedicated single stain anti-FLAG controls:
    a. Single stain shows FLAG-positive cells in 555nm channel in which this fluorophore was imaged.
    b. Single stain observed in 488nm channel in which calnexin was imaged demonstrates minimal carryover.

4.10 Collaborative work using zebrafish and differentiation of CD34$^+$ stem cells to MK

The results in this section were obtained through collaboration with the laboratory of Professor Kathleen Freson of the University of Leuven. Professor Freson, Ms. Chantal Thys and Ms. Mara De Reys performed and interpreted the experiments.

4.10.1 KDSR is a putative regulator of thrombopoiesis in zebrafish (*Danio rerio*)

Morpholino oligonucleotides (MO) are a widely used tool for knockdown in zebrafish and have been used extensively to explore phenotypes associated with candidate genes [187]. MO consist of short primers comprising morpholine bases, containing a 6-ring heterocycle backbone and non-ionic phosphorodiamidate linkages. These are designed to be complementary to the RNA of interest and bind with high affinity to prevent normal transcript processing and translation by steric hindrance [187]**.** To further explore the role of KDSR in megakaryopoiesis, CD41 transgenic zebrafish embryos with GFP-labelled thrombocytes (*Tg(cd41:EGFP)*) were injected with a kdsr ATG MO (KDSR ATG_MO) as described in section 2.5.7. Morphology was normal except for curved tails that often co-occurs with thrombocyte defects ([147] Figure 64a). Flow cytometry for the detection of GFP -labeled thrombocytes was performed at 72 hpf. A dose-dependent reduction in thrombocyte numbers were observed using KDSR ATG_MO compared to controls (Figure 64a, b). Immunoblot analysis confirmed reduced KDSR and GFP levels in pooled samples (n = 5 embryos/condition) from KDSR depleted embryos compared to controls (Figure 64c, d).

4.10.2 Proband CD34$^+$-derived MK have reduced proplatelet formation compared to a control

CD34+ stem cells were isolated from bone marrow aspirates from the proband and a healthy gender-matched donor by magnetic cell sorting. MK directed differentiation was performed once, as described in section 2.5.8. Flow cytometry for CD41$^+$ expression showed a higher number of proband CD41$^+$ cells at day 7 (23.1% vs. 7.6%, Figure 65a), at odds with our findings of reduced yield in K^{ev} iMK. Proband cells differentiated for 11 days produced cells with membrane budding but almost no formation of normal proplatelets ($p<0.0001$, unpaired two-tailed Student's t-test, Figure 65b, c). Staining for actin and tubulin in combination with markers for α- (VWF) and δ- (CD63) granules of MK from the propositus showed reduced granule trafficking to abnormal proplatelet structures (Figure 65d, e).

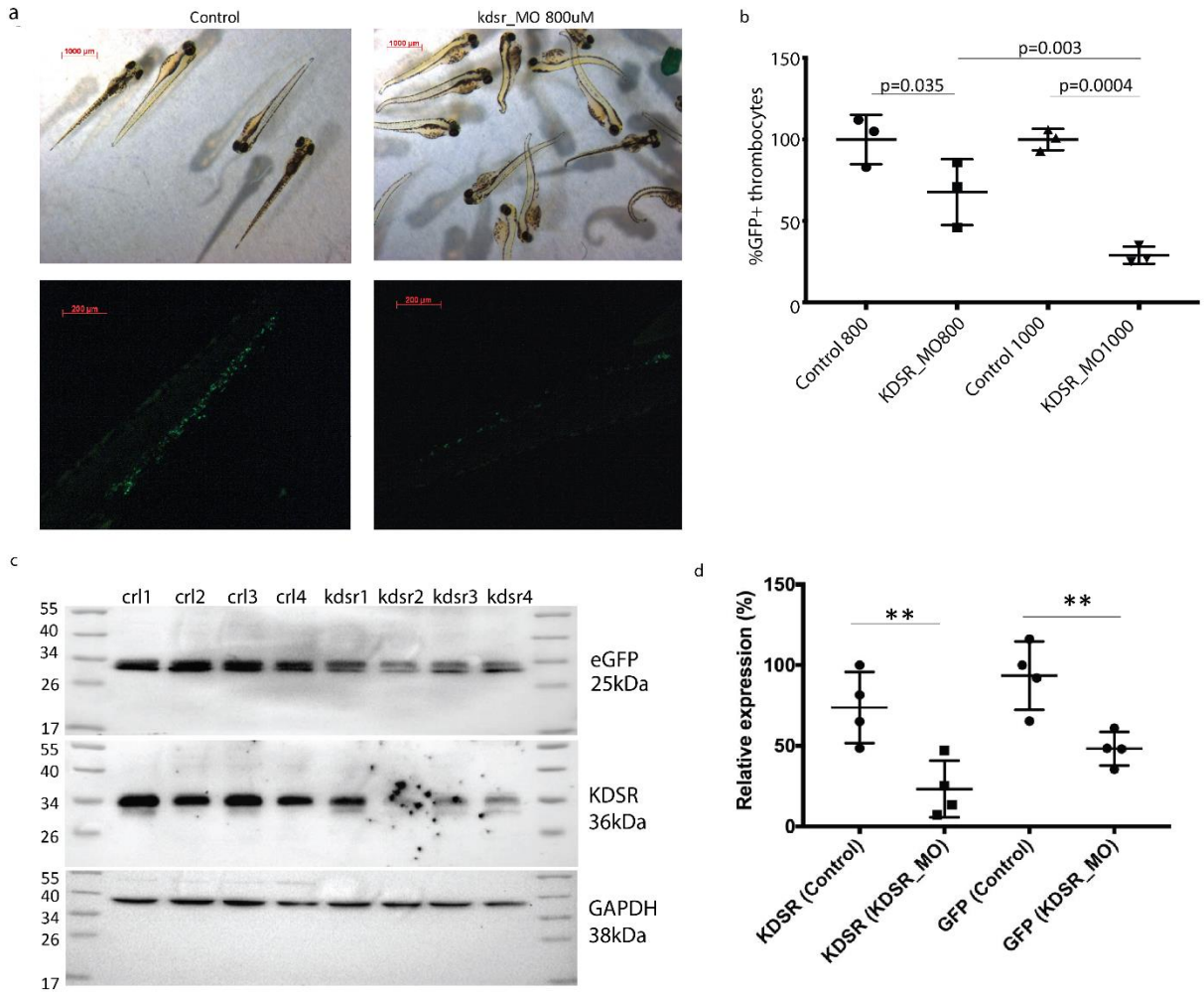


Figure 64 Zebrafish experiments performed by the University of Leuven
(a) Upper panels: phenotype after kdsr knock down in Danio rerio using an ATG antisense MO (800 µM) after 72 hpf. Most embryos develop curled tails but no other dysmorphisms. Lower panels: stereomicroscope images (20x original magnification) in the tail region at 72 hpf to visualize the GFP labeled thrombocytes. Tg(cd41:EGFP) embryos injected with 800 µM kdsr ATG_MO present with reduced number of thrombocytes.
(b) Tg(cd41:EGFP) Danio rerio embryos at 72 hpf were analysed by flow cytometry for presence of GFP-labelled thrombocytes after buffer (control) or kdsr-MO (800 or 1000 µM) injections. Values are means and standard deviation (SD) as quantified for 10 randomly selected embryos for each condition in triplicate. Results were analysed by one-way ANOVA with multiple comparisons test.
(c) Immunoblot for GFP and kdsr in injected and lysed Tg(cd41:EGFP) embryos at 72 hpf. The kdsr ATG_MO was injected at 1000 µM and buffer injections were used as control. GFP and kdsr proteins were reduced after kdsr knockdown. Equal amounts (50 µg) were loaded (5 randomly selected embryos for each of the 4 conditions) and gapdh was used as loading control.
(d) Quantification of immunoblot after normalization for gapdh. Values are means and standard deviation (SD); ** p = 0.004 (for kdsr), ** p = 0.009 (for GFP), one-way ANOVA.

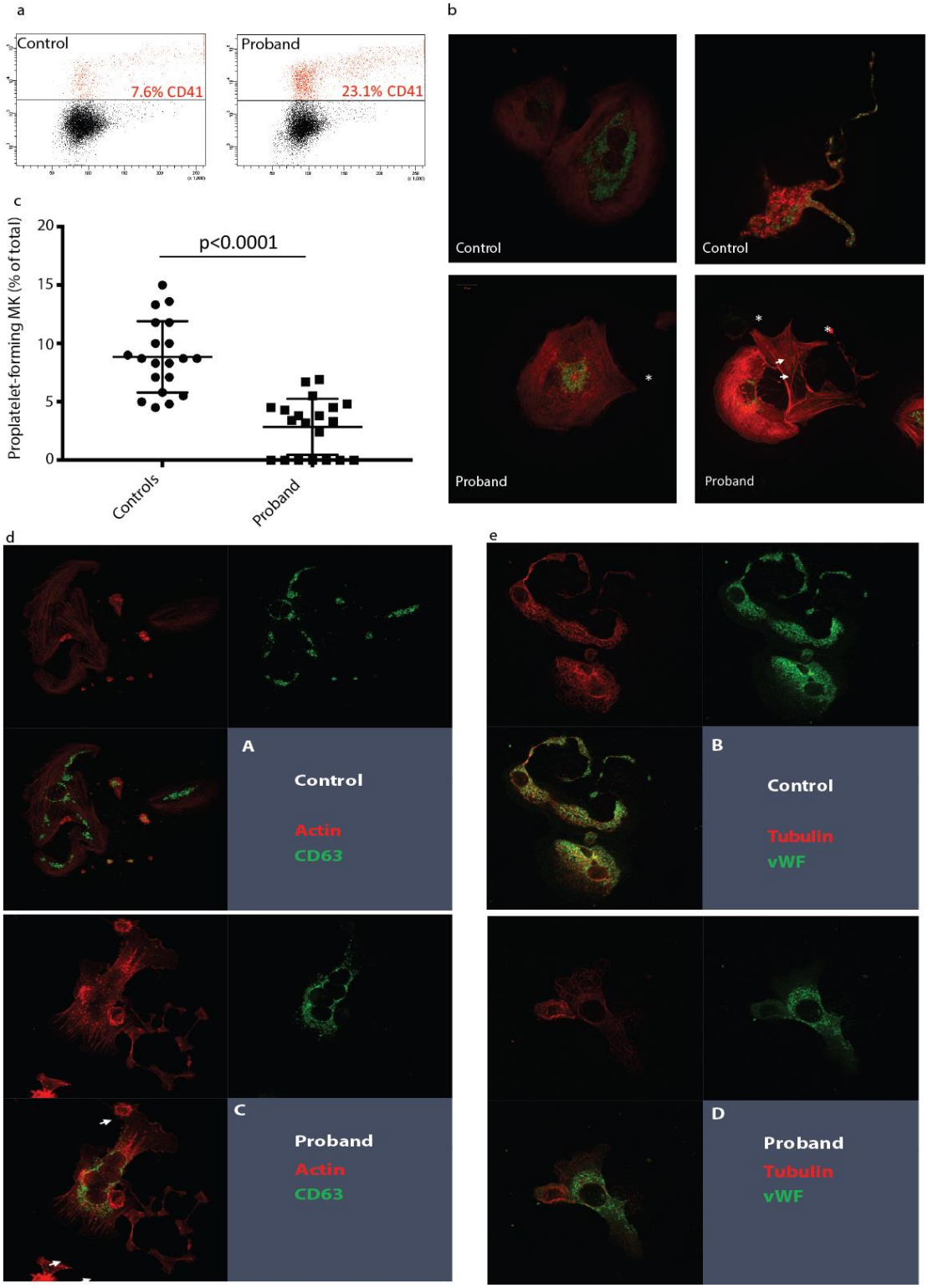


Figure 65 Liquid MK cultures from CD34+ cells from proband, performed by the University of Leuven
(a) Liquid MK cultures analysed at day 7 of differentiation using flow cytometry and staining for CD41 showed hyperproliferation of MK for the propositus.
(b) MK at day 11 from control and propositus spread on fibrinogen and stained for actin (red) and CD63 (green). MK for the propositus have irregular cytoskeletal structures with lamellipodia (*) and filopodia (arrow) and demonstrate defective CD63-positive granule trafficking to proplatelets.
(c) Proplatelet quantification at day 11 of differentiation. All MK with proplatelets and membrane budding were counted as positive as quantified for 20 images. Results were analysed by the unpaired Student's t-test.
(d) More images of MK at day 11 from control (A,B) and propositus (C,D) spread on fibrinogen and stained for the indicated markers. MK for the propositus have irregular cytoskeletal structures that resemble filopodia (arrow), do not form normal proplatelets and demonstrate defective granule trafficking.

4.11 Discussion

Variants in *KDSR* have recently been associated with a recessively inherited syndrome with significant skin pathology and thrombocytopenia [3, 4]. In this chapter we have described novel compound heterozygous variants in *KDSR* found in two individuals in a pedigree, associated with severe thrombocytopenia and at most a subtle dermatological phenotype. This expands the phenotypic spectrum of KDSR variants from severe skin problems with no apparent haematological involvement, to profound thrombocytopenia with very mild skin abnormalities. We have described an additional novel phenotype of severe juvenile myelofibrosis in the proband, however the affected sibling was too young to allow confirmation of this phenotype. Genotype-phenotype relationships were explored in view of the heterogeneity of presentation, however too few variants have been described as yet to allow a detailed examination.

Global metabolic profiling confirmed loss of function of *KDSR* in the proband. The enzyme substrate, KDS, is detectable in patient plasma and absent from the rest of the pedigree and controls. Unexpectedly, there was also a rise in the downstream metabolite DHS. These findings were confirmed on a separate mass spectrometry platform and raise the hypothesis that *KDSR* loss of function during de novo sphingolipid synthesis may be over-compensated by alternative mechanisms, though the source of the extra DHS remains unknown. DHS may arise from the recycling of relatively abundant sphingomyelins, a pathway previously shown to contribute to sphingolipid production downstream of DHS such as dihydroceramide, ceramide and sphingosine. However recycling was previously considered to contribute little if at all to free DHS production under normal conditions [188]. Post-translational modifications to sphingolipid enzymes may also explain the metabolic profile, and further work is required to explore this possibility.

Sphingolipids, in particular ceramide and its derivatives, have been shown to regulate apoptosis and as such have become new key players in the search for novel chemotherapeutic agents. In light of this, and following observation of cell fragility in culture we explored this phenotype *in vivo*. Lentiviral rescue of an iPSC line carrying *KDSR* variants led to improved growth and reduced apoptosis by TUNEL assay. Rescued iMK showed improved viability, yield and maturity. Wild-type expression rescue of iMK carrying *KDSR* variants showed transcriptional upregulation of *ASAH1* and downregulation of *CERS6*, and rescued iMK showed higher sphingosine and lower ceramide levels. These findings are in keeping with derangement of the 'sphingolipid rheostat,' a concept that describes the pro-apoptotic properties of ceramide, and anti-apoptotic, pro-growth properties of sphingosine-1-phosphate, which is derived from sphingosine. Wild-type expression rescue in iMK was also associated with differential expression of genes annotated with Gene Ontology terms related to growth. Kihara *et al.* provided further support for this phenotype by demonstration that knockout of the yeast *KDSR* analog, *Tsc10p* is associated with growth defects [189]. Interestingly DHS, under the trade name Safingol, has been shown to potentiate chemotherapy-induced apoptosis *in vivo* and *in vivo* [190], and is in clinical trials for treatment of

advanced malignancy (ClinicalTrials.gov identifier NCT01553071). This presents the intriguing hypothesis that the compensatory rise in DHS could contribute to the apoptosis phenotype. DHS has also been shown to inhibit protein kinase C (PKC) by competitive binding of the regulatory phorbol binding domain [190]. PKC has been shown in several studies to regulate the initiation of MK differentiation [191], proposing an alternative mechanism by which this metabolite might influence platelet count.

The next logical question is whether the apoptotic phenotype may cause or contribute to thrombocytopenia. Apoptosis is active in every cell type, yet the clinical phenotypes described to date only involve the skin/hair and blood. Considering chemotherapy as a model of the clinical features that accompany system-wide induction of apoptosis, the greatest side effects involve the blood, skin/hair, gastrointestinal and reproductive systems, which are the cells with highest turnover. Whilst no gastointestinal or reproductive symptoms have yet been reported in individuals carrying KDSR variants, few cases have been described, many of these cases are young, and *KDSR* has already shown itself to be capable of dramatic pleiotropy, so the full phenotypic spectrum is as yet unknown.

MK may be particularly vulnerable to dysregulated apoptosis. The idea that platelet release is a form of controlled apoptosis is a matter of debate. This concept was first suggested following the finding of thrombocytopenia in mice lacking pro-apoptotic proteins (BIM) or overexpressing pro-survival proteins (BCL2) [192]. Subsequent studies showed discrete, localised caspase activation in the cytoplasm of human $CD34^+$ -derived MK. Pharmacological caspase inhibition and retroviral overexpression of anti-apoptotic BCL2 impaired proplatelet formation [193]. Further support came from the report of a pedigree with a variant in *CYCS* presenting with isolated mild-to-moderate thrombocytopenia, enhanced apoptosis *in vivo* and abnormal platelet-like particle formation from patient $CD34^+$-derived MK [194, 195]. The CYCS story is particularly relevant as an example of how a disorder involving a protein known to be vital for the initiation of the intrinsic pathway of apoptosis and that is ubiquitously expressed can manifest as isolated thrombocytopenia. It has also been shown that MK-specific knockout of pro-survival protein BCL-X_L with an overall pro- apoptotic phenotype resulted in mice with severe thrombocytopenia and failed proplatelet formation, whilst knockout of pro-apoptotic BAK1 and BAX did not alter platelet formation [196].

We sought to functionally confirm *KDSR* as a regulator of thrombopoiesis and to describe the mechanism by which variants cause thrombocytopenia. Evidence has been shown that a lentiviral rescue model in iPSC with forward programming to iMK recapitulates and rescues the major metabolic phenotype of increased KDS. In iPSC DHS and several other sphingolipids also appeared less abundant in the rescued line. The pattern of metabolic changes was similar for KDS and DHS in iMK, and it is hard to take conclusions from these. We showed that rescue was associated with improved and increased proplatelet formation, increased iMK yield, maturity and size. We performed a limited, post-hoc analysis of PLP

formation to show that it was increased in the rescued line. A collaboration with the University of Leuven showed that *kdsr* MO knockdown in zebrafish is associated with thrombocytopenia. These findings must be interpreted with caution in view of a recent reverse genetic screening study in which CRISPR-induced mutants were unable to recapitulate the majority of MO-induced phenotypes tested [197]. The collaborators in Leuven also showed that MK derived from proband bone marrow $CD34^+$ stem cells were morphologically abnormal, showing a reduction in their ability to generate proplatelets *in vivo*, consistent with the iMK findings. Immunostaining for granule markers in these MK showed abnormal trafficking of granules. This was in-keeping with platelet EM from the propositus that demonstrated platelets with abnormal α- and δ- granules, agranular platelets and abnormal internal membrane complexes. However, proband MK were hyperplastic, at odds with our findings of increased yield for K^{resc} iMK. Repetition of this single experiment is planned, using a fresh peripheral blood sample.

These findings point to *KDSR* as a regulator of thrombopoiesis. However, mechanisms for the abnormal proplatelet formation remain unclear. It may be that *KDSR* variants mediate a cytoskeletal defect that impairs proplatelet formation and therefore granule trafficking. Cytoskeletal disorganisation has been shown to mediate several congenital thrombocytopenias caused by variants in *MYH9* [24]*, ACTN1* [25]*, FLNA* [26] *, TUBB1* [27], *DIAPH1* [6] and *WAS* [28], many of which are associated with abnormal proplatelet formation. Furthermore, the sphingolipids ceramide (downregulated in the propositus) and sphingosine-1-phosphate have been shown to regulate the actin cytoskeleton by interaction with adaptor molecules ezrin, radixin and moesin [198]. Whilst less is known about the bioactivity of early metabolites in sphingolipid synthesis that were most abnormal in the propositus, such as KDS, DHS, dihydroceramide and its derivatives, it is credible that derangement of the sphingolipid profile may mediate a cytoskeletal defect in this condition.

Further studies are required to definitively assess the hypotheses in this chapter. Repetition of forward programming using the rescue and empty vector lines is required. This is particularly important to assess PLP generation in a dedicated assay using a viability stain suitable for anucleate cells such as Calcein AM and double staining of CD41 and CD42, markers of true PLP. Confirmation of the apoptosis phenotype in iPSC could be performed using an alternative strategy such as activated caspase or cytochrome C detection by western blot, glutathione assays or mitochondrial membrane potential dyes. Large scale iMK generation could allow apoptosis assays to be performed in iMK (TUNEL TdT requires $1x10^6$ cells per condition per replicate). The gene-lists derived from RNA-seq and GO term enrichment analysis also require further exploration *in silico* and *in vivo*, for further investigation of the mechanism by which *KDSR* variants may influence growth, differentiation and the cytoskeleton.

However my findings will only be suitable for broader generalisation if further studies introduce biological variation, either by repetition of these studies using iPSC derived from an alternative affected pedigree, or using an alternative cellular model. A follow-up study could use site-directed mutagenesis

to introduce p.Arg154Trp and other variants described in the literature into the lentiviral vector used for this study, followed by transduction in iPSC and iMK generation for phenotyping. Alternatively, CRISPR-Cas9 knock-in could be revisited to introduce putative pathogenic variants. Strategies to improve efficiency would be essential, for example by testing more sgRNAs by surveyor assay prior to selection for use, testing several Cas9 plasmids to maximise transfection efficiency, or by using restriction fragment length polymorphisms (the p.Arg154Trp introduces a CviQI site in the reference sequence) to rapidly screen large numbers of clones at an early stage. For this study 17nt sgRNAs were used in light of reports at that time of reduced off target binding, however recent reports suggest 20nt sgRNAs with specific nucleotides at positions 3, 16, 20 and the variable nucleotide of the PAM sequence are more efficient[199, 200]. A plasmid carrying an alternative antibiotic selection that also carries the template for homology-directed repair, instead of ssODNs, may also improve efficiency. The choice of one or more hiPSC lines for this work would allow the comparison of cell phenotypes to established control datasets and thus improve the experimental design. The contribution of individual sphingolipid moieties to the phenotypes observed remains unknown, and supplementation studies in stock iPSC and iMK, beginning with synthetic KDS and DHS at physiological levels, could provide answers if phenotypes are recapitulated.

Chapter 5: ABCC4

5 ABCC4

5.1 Discovery of *ABCC4* as a candidate gene

High throughput sequencing data from cases recruited at RFH were initially examined to identify variants in genes known to influence BPD. Three sisters were identified that carry a novel heterozygous variant in the *P2RY12* gene, encoding the ADP receptor which is known to play a central role in ADP-induced platelet aggregation. Heterozygosity for *P2RY12* variants has been associated with minimal clinical bleeding symptoms despite abnormal aggregometry tracings [201]. Whilst a compatible aggregation defect was present in all three sisters, two out of three reported a lifelong history of excessive bleeding. The other sister, despite haemostatic challenges, reported no bleeding. This led to a search for a second hit, in *P2RY12* or in an alternative gene. Variants in genes identified by GWAS to influence platelet traits were reviewed and led to identification of a homozygous single base pair deletion leading to a frameshift in gene *ABCC4* in the sisters with bleeding symptoms (Figure 66). To our knowledge *ABCC4* variants have not been reported to cause human platelet pathology.

5.2 Aims

The primary objectives of this chapter are to explore the hypothesis that *ABCC4* variants cause the phenotype and to evaluate the contribution of the *P2RY12* variant.

I will:

- explore and expand the clinical and laboratory phenotype of the pedigree
- evaluate the variant effects in the context of the known structure and function of *ABCC4*
- explore the phenotype and mechanism of action using assays of patient platelets
- determine the metabolic profile of plasma samples from the pedigree to determine candidate mechanisms of action
- develop a cellular model to investigate the hypothesis using iPSC forward-programmed to iMK

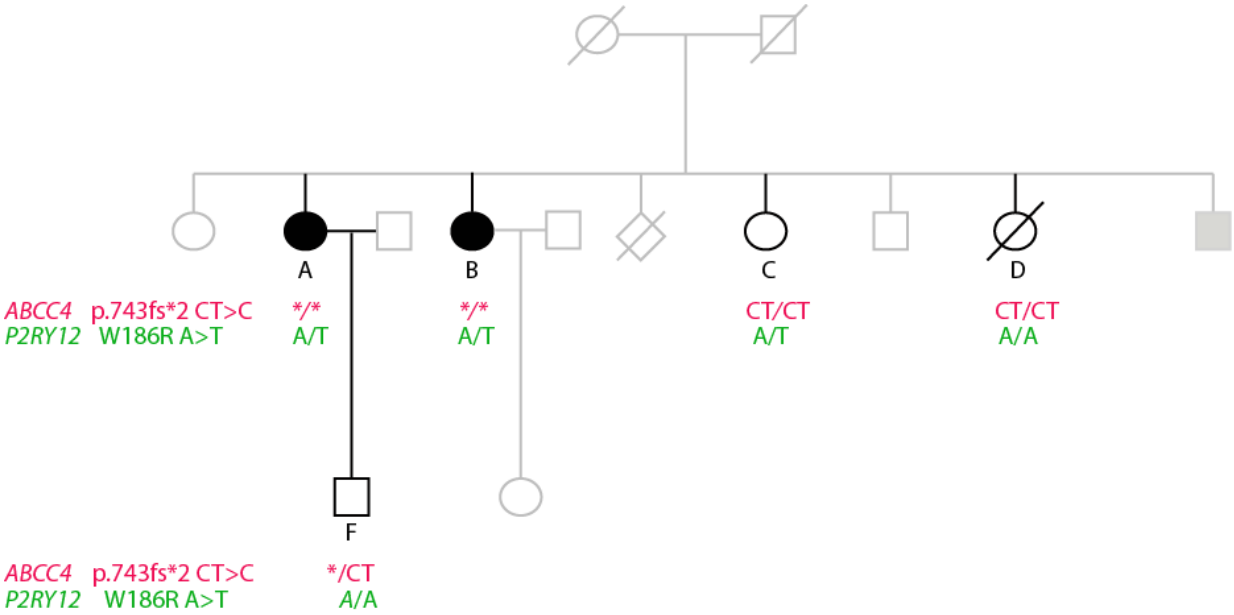


Figure 66 Pedigree tree showing pedigree members recruited, labelled A-F, with cases affected by bleeding shaded in black. Case D died of metastatic breast cancer during the course of the study.
Pink notation indicates *ABCC4* variant (* indicates stop codon, CT indicates wild type allele). Green notation indicates *P2RY12* variant. Cases with clinically significant bleeding were found to carry both the homozygous variant in *ABCC4* and the heterozygous variant in *P2RY12*.

5.3 Background

ABCC4 is located on the reverse strand of chromosome 13q32.1 and encodes a 1325 amino acid membrane-bound transporter glycoprotein. Like other members of the ABC (ATP-binding cassette) super-family, ABCC4 hydrolyses ATP to translocate substrates across a membrane, against a diffusion gradient [202]. ABCC4 was first reported as a mediator of drug resistance to nucleoside–based anti-retrovirals, hence its alternative name of multidrug resistance protein 4 (MRP4) [203]. It has broad substrate specificity, even amongst ABC transporters, including a multitude of drugs such as aspirin, chemotherapeutic agents and anti-virals. Its varied endogenous substrates include cAMP, cyclic guanine monophosphate (cGMP), ADP and eicosanoids, all of which regulate platelet activation [202].

5.3.1 Structure

ABCC4 has two membrane spanning domains, each comprising 6 transmembrane helices arranged longitudinally around a central pore, and two cytosolic nucleotide binding domains that are thought to be held in close proximity [202, 204](Figure 67a). The crystal structure of ABCC4 has not been reported, and structure-function relationships have been inferred from homology with the more extensively studied and crystallised bacterial ABC transporters and ABCB1 [205]. There may be at least two independent transport sites within the protein, as multiple compounds have been shown to be transported simultaneously [206].

5.3.2 Location

In the majority of cells ABCC4 is localised in the plasma membrane [202]. Its location in platelets is less clear. Jedlitschky *et al.* suggested ABCC4 was predominantly in dense granule membranes in resting platelets, with lower levels in the plasma membrane. They performed immunoblotting of platelet membrane fractions and showed that ABCC4 colocalised with mepacrine in fractions enriched in, but not specific for, dense granules. Platelet immunostaining also colocalised ABCC4 with mepacrine-storing, LAMP-2 positive structures [207]. A further study using the same antibody panel reported that ABCC4 was seen only in the plasma membrane in patients lacking dense granules [207, 208]. Mattiello *et al*., also showed ABCC4 predominantly in the dense granule membranes of healthy volunteers, and in the plasma membrane following coronary artery bypass grafting [209]. Conversely, a repository of platelet membrane proteins identified by mass spectrometry [210] and a mouse knockout study using plasma membrane biotinylation and super-resolution microscopy [211] showed the majority of ABCC4 in human and mouse platelet plasma membranes at rest. Another study by Oevermann *et al.,* showed ABCC4 in both granules and the plasma membrane of MK from bone marrow and MK grown in vitro from $CD34^+$ progenitors [212]. These discrepancies may be explained by regulation of the localisation of ABCC4, for example due to genetic variants or altered expression of interacting proteins; in HEK293T cells localisation of ABCC4 at the plasma membrane was abolished by missense mutation p.Thr796Met [213], p.Glu757Lys [214], and by interference with the PDZ-binding motif of ABCC4, through which it interacts with MPP1[215]. Recently siRNA knockdown of ABCC4 interactors SNX27, PSD95, and AP3B1

in a megakaryoblastic leukaemia line led to redistribution of MRP4 from the cytosol to the plasma membrane [216].

5.3.3 Function

Several lines of evidence support a significant role for *ABCC4* in megakaryopoiesis and/ or thrombopoiesis. *ABCC4* is highly and selectively expressed in MK and platelets compared with other haematopoietic cells (Figure 67b,c) [103, 212]. Promoter capture Hi-C experiments that use high-throughput sequencing to capture the three-dimensional structure of chromatin have shown that *ABCC4* has a highly connected promoter in MK, suggesting that transcription of *ABCC4* is highly regulated (Figure 67d, e, [217, 218], www.chicp.org, accessed 15.11.17). Two genome wide association studies (GWAS) of platelet traits identified a sentinel SNP *rs4148441* in *ABCC4* as strongly associated with platelet count (effect size 4.1 x 10^9, *p* value 6.76 × 10^{-12}) [108, 111]. The enrichment of known BPD genes in these GWAS studies suggests *ABCC4* is a promising candidate gene for BPD.

A wealth of targeted experiments provides further support for the overall hypothesis, despite some discordant findings between studies. Jedlitschky *et al.* postulated a role for ABCC4 in the uptake of ADP to dense granules. They showed ATP-dependent cGMP transport and ADP transport occurred in a platelet membrane fraction enriched in, but not specific for, dense granules. This transport was reduced by MK571, a pharmacological competitive antagonist of receptors ABCC4, ABCC1 and LTD4 [179]. The absence of platelet ABCC4 on immunostaining in two unrelated cases with selective ADP deficiency with otherwise normal dense granules was put forward as further evidence of this. No causal *ABCC4* variant was identified on exome sequencing [208].

Subsequently *ABCC4* knockout (KO) mice generated in one laboratory were reported in two independent studies [211, 219]. Cheepala *et al.* documented normal dense granule number, structure and function. *Abcc4*$^{-/-}$ was associated with increased tail-bleeding, impaired thrombus formation and a selective impairment in collagen-mediated platelet attachment and aggregation. Consistent with this, collagen receptor GPVI surface expression was reduced. Total intracellular cAMP was increased, and the authors concluded that ABCC4-mediated cAMP efflux regulated platelet function via altered GPVI surface expression [211], in keeping with previous reports that cAMP regulates GPVI surface expression and dimerization [220, 221]. Decouture *et al.* corroborated the bleeding phenotype in wild-type mice transfused with *Abcc4-/-* platelets, impaired attachment on collagen and normal dense granule structure, number and ADP content. However the aggregation profile was different, with defects observed in response to PAR4, U46619 and ADP but not collagen or convulxin (a specific GPVI agonist). In this study the ratio of cytosolic: dense granule cAMP was higher in KO than wild-type mice, but total platelet cAMP was similar in both conditions. Immunoblotting of whole platelet lysates showed faster VASP phosphorylation in KO platelets, suggestive of higher cytosolic cAMP; cytosolic cAMP-mediated activation of PKA causes phosphorylation of VASP. The authors concluded that ABCC4-mediated cAMP compartmentalization in dense granules regulates platelet function [219].

A number of lines of evidence support the hypothesis that ABCC4 regulates cAMP. Reduced expression and pharmacological inhibition of ABCC4 by MK571 increased intracellular cAMP in human, rat and mouse vascular smooth muscle cells [222-224], rat and mouse ventricular myocytes [225, 226] and human leukemia cell lines [227]. Borgognone and Pulcinelli used MK571 on human platelets to show reduced cAMP and cGMP uptake to dense granules, increased VASP-phosphorylation and enhance the inhibition of platelet aggregation caused by agents that elevate cAMP such as forskolin [228]. Lien *et al.* showed that MK571 inhibited collagen-induced aggregation and ATP-release, and delayed PFA-100 closure time in human platelets. Thrombus formation was also inhibited in a mouse mesenteric venule model [229]. Limited conclusions can be drawn from MK571 studies due to its lack of specificity, and further studies using recently identified specific inhibitors are warranted [230].

Jedlitschky *et al.* also observed upregulation of ABCC4 expression following treatment with agents that raise intracellular cAMP in several cell lines including a megakaryoblastic leukaemia line. Activation and inhibition of the EPAC and MAPK pathways (which are usually activated by cAMP) activated and inhibited this upregulation, respectively [231]. This suggests that *ABCC4* expression is dynamically regulated by intracellular cAMP levels via its downstream effectors, perhaps in order to provide its function of maintaining cAMP homeostasis. Genetic variants p.Gln187Trp [232] and p.Thr796Met [213] have also been found to regulate *ABCC4* expression in HEK293T cells. Studies of drug transport, also in HEK293T cells, have identified multiple genetic variants that alter ABCC4 function and may provide insights to structure-function relationships in platelets, including microRNA-124a and microRNA-506 [233], p.Gln487Glu and p.Val1071Ile [234], p.Gln187Trp and p.Gln487Glu [232], p.Arg998, p.Phe368 and p.Trp995 replacement [235] and p.Glu757Lys [214].

In summary, several observations support *ABCC4* as a candidate gene for BPD, including a strong association with platelet parameters in GWAS, selective, high expression in MK and platelets and bleeding phenotypes in knockout mice. Potential mechanisms of disease involving dense granule nucleotide storage and cAMP regulation have been presented and warrant further examination in the cases with high impact *ABCC4* variants identified by the NIHR BR-RD BPD study. These studies aim to inform the debate regarding the function of ABCC4 in platelets and is of value to the wider community because ABCC4 has become an important cardiovascular and chemotherapy drug target [204, 215, 236, 237].

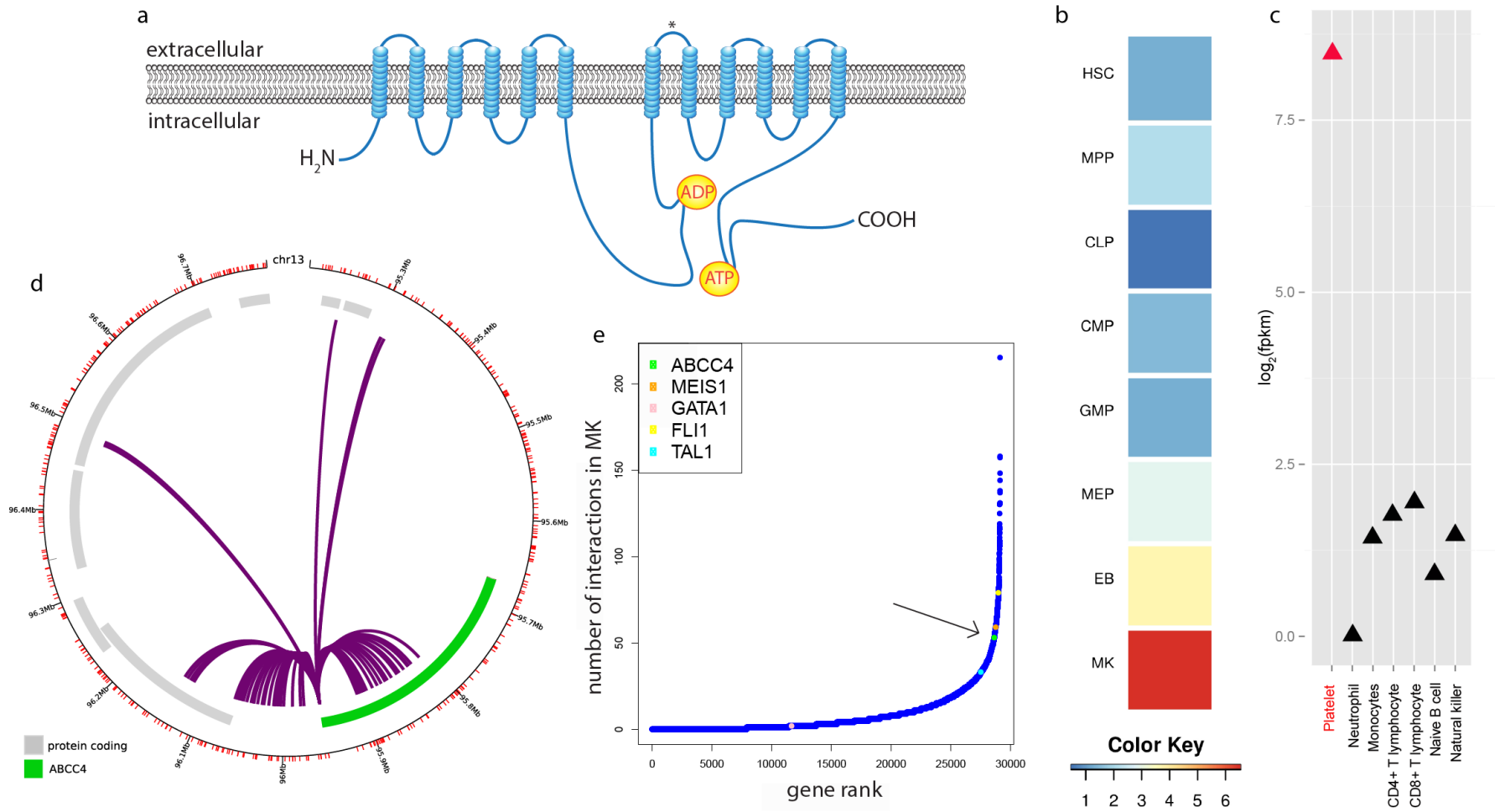


Figure 67 ABCC4 structure and function in megakaryopoiesis

a. Schematic of ABCC4 protein, adapted from Belleville-Rolland *et al.* [204]. Two sets of 6 transmembrane helices are followed by cytoplasmic nucleotide-binding domains that are held in close proximity.
b. Expression of *ABCC4* is higher in MK than other haematopoietic progenitors by RNA-seq. Normalised expression values expressed as log_2 of fragments per kilobase of transcript per million reads mapped (FPKM) [103]. Heatmap courtesy of Dr. Luigi Grassi, University of Cambridge. HSC, haematopoietic stem cell; MPP, multi-potent progenitor; CLP, common lymphoid progenitor; CMP, common myeloid progenitor; GMP, granulocytic-myeloid progenitor; MEP, megakaryocyte-erythrocyte progenitors; EB, erythroblast, MK, megakaryocyte.
c. Expression of ABCC4 is higher in platelets than other peripheral blood cells. Adapted from figure provided by Dr. Luigi Grassi using same units of scale and source as (b).
d. The *ABCC4* promoter is highly connected in megakaryocytes. CHICP plot showing interactions between the promoter of ABCC4 and the adjacent regions of the genome, as defined by promoter-capture HiC [217, 218] (www.chicp.org, accessed 15.11.17)
e. Data from (d) plotted with promoter capture HiC data from all other genes highlighting key regulatory genes in megakaryocytes, in order of the number of interactions. *ABCC4* (arrowed) lies on the steep part of the curve, suggesting that it is highly connected, and more so than *GATA1* and *TAL1*. Graph courtesy of Dr. Romina Petersen, University of Cambridge.

5.4 Pedigree and clinical characteristics

5.4.1 Clinical phenotype

5.4.1.1 Proband, case A (RFH_14I)

The proband is a 63 year old lady from the Seychelles (Figure 66). As a child she suffered recurrent epistaxis. She reports lifelong excessive bruising and bleeding following minor injury, for example requiring stitches following a minor knife injury that bled for several hours. Prior to the onset of menopause she suffered menorrhagia, bleeding for ~7 days with passage of clots, but did not seek intervention. At the age of seven years she underwent appendicectomy, and cannot recall bleeding but was kept in hospital for several weeks due to unknown complications. Following the birth of her son she suffered a postpartum haemorrhage. She was not transfused but returned to hospital following discharge with recurrent bleeding which continued for approximately a month, during which time she remained in hospital. She also suffers from migraine and bronchitis. Full blood count and blood film were unremarkable (Table 19). She reported no known consanguinity between her parents or previous generations. There is no known family history of excessive bleeding.

5.4.1.2 Affected sibling, case B (RFH_15G)

The proband's older sister is 64 years old and also suffered from lifelong easy bruising. She bled for over 24 hours following a tooth extraction at the age of ten and recalls her pillow saturated with blood. She returned to the dentist who intervened with 'gum soaks.' Prior to menopause she also reported menorrhagia, bleeding for 7 days with passage of clots, but sought no intervention. She has one child and reported no bleeding postpartum. She required stitches to a foot injury following prolonged bleeding. She had an appendicectomy and tonsillectomy in childhood and does not recall excessive bleeding during either operation. She has no other medical history of significance. Full blood count and blood film were also unremarkable (Table 19).

5.4.1.3 Other family members.

Unaffected sibling, case C (RFH_32E) is 59 years of age and reports no history of easy bruising, epistaxis, menorrhagia or bleeding following minor injury. A previous injury on broken glass did not bleed excessively and required no intervention. She has had a tonsillectomy, minor surgery to her foot and two dental extractions without haemostatic complications.

Unaffected sibling, case D (RAL_272A) was 55 years old at recruitment and reported no history of excessive bleeding or bruising, including following tonsillectomy and appendicectomy in her teens. She developed immune thrombocytopenia and Sjogren's syndrome in her third decade which was steroid responsive but recurrent thereafter. She underwent a lumpectomy for breast cancer and surgical management of a fracture of her foot in her fifth decade without excessive bruising or bleeding complications. At the time of review she had a platelet count of 18 x 10^9/L precluding extended platelet

phenotyping, but consented to genotyping. She sadly died following recurrence of breast cancer during the course of the study.

Case F (RFH_2738), son of case A, was 19 years old at recruitment and reported no history of bleeding but also no history of significant haemostatic challenges. An affected brother declined to participate in the study, and two unaffected siblings could not be contacted.

5.4.2 Clinical laboratory findings

Light transmission platelet aggregometry had been performed three times prior to recruitment on siblings A, B and C, and showed a consistent weak agonist response defect in all cases, with disaggregation following the primary phase response to low dose ADP stimulation and a reduced response to low dose epinephrine (Table 20). Aggregometry in case F showed a milder defect with disaggregation following primary phase with low dose ADP but response to epinephrine was normal. The only laboratory test to co-segregate with the bleeding phenotype in cases A and B was a prolonged PFA-100 closure time using the collagen/adrenaline cartridge.

5.5 *P2RY12R* variants co-segregate with the laboratory phenotype but not bleeding

A novel heterozygous missense *P2RY12R* variant was detected in all sequencing reads in cases A, B and C (3:151056078 A>T, p.Trp186Arg). Cases D and F were homozygous for the reference allele. In all cases the findings were confirmed by Sanger sequencing.

As stated, cases A, B and C showed a weak agonist response defect involving ADP and epinephrine on light transmission aggregometry. *P2RY12* encodes the ADP receptor on platelets and residue p.Trp186 has been shown to project into the ADP-binding pocket. Adjacent residue p.His187 has been shown to bind ADP. This region undergoes conformational change during receptor activation [201, 238]. The novel variant is not reported in the GnomAD dataset, however several other variants at this location are reported in heterozygosity, two of which lead to p.Trp186* at a minor allele frequency of $3.23x10^{-5}$ and $8.13x10^{-6}$ [95] (accessed 31.1.18).

These findings are consistent with reports that heterozygosity for *P2RY12R* variants that alter receptor function, including p.His187, usually mediate a laboratory phenotype of impaired aggregation with minimal clinical bleeding [201]. This is also consistent with the lack of clinical bleeding in case C, but does not explain bleeding in cases A and B.

5.6 Homozygosity for *ABCC4* variants co-segregates with clinical bleeding and is likely to be rare

A homozygous single base pair deletion was detected in all sequencing reads in cases A and B that leads to a frameshift in *ABCC4* (13:95815456 CT>C, p.Ser743fs*2, exon 18 of 35). Modelling of the consequences of the frameshift shows that a stop codon is introduced after two amino acids and repeatedly thereafter in exon 18 (Figure 68). This was absent in cases C and D, and heterozygous in case

F. Therefore homozygosity for this variant co-segregates with the bleeding phenotype and prolonged PFA-100 test. In all cases the findings were confirmed by Sanger sequencing.

The variant found in *ABCC4* is absent from the GnomAD database. There were no reported individuals with homozygous loss of function variants in *ABCC4* [95] (accessed 31.1.18) however, the under-representation of non-Caucasian ethnic groups reduces our confidence in the rarity of this variant. Haplotype mapping of NIHR BR-RD cases alongside data from the International HapMap Project performed by Dr. Ernest Turro showed that the Seychellois cases clustered close to Maasai people in Kenya and Africans in southwestern USA (Figure 69a, [99, 239]). The variant was also absent from 640 haplotypes in the African Sequencing project [99] (accessed 8.7.17). There was loss of heterozygosity in the pedigree indicating unknown consanguinity; both cases showed an approximately 60Mb stretch of homozygosity on the long arm of chromosome 13. This increases the likelihood that p.Ser743fs*3 is a pedigree-specific variant rather than a common Seychellois founder variant. However further work should include the collection of haplotype data from the Seychelles to establish the true frequency of this variant in this island population, preferably on-site.

Patient	A	B	C	D	F
Platelet count ($x10^9$/L)	199	336	291	18	228
Mean platelet volume (fL)	12.0	10.9	11.4	NA	NA
Immature platelet fraction (%)	4.3	1.7	2.5	17.8	NA
Platelet crit (%)	0.07	0.11	0.10	NA	NA
Haemoglobin (g/dL)	11.8	12.8	12.0	12.3	13.2
Mean red cell volume (fL)	82.2	81.6	82.2	82.9	69.8
White cell count ($x10^9$/L)	7.07	6.35	11.3	12.3	6.23
White cell differential	Unremarkable	Unremarkable	Unremarkable	Unremarkable	Unremarkable

Table 19 Sysmex full blood counts from cases A-F. Case D had immune-mediated thrombocytopenia in the context of multi-system autoimmune disease and disseminated malignancy.

5.7 The variant is present in platelets and MK and reduces ABCC4 protein expression

5.7.1 Platelet proteomics

Immunoblotting using three antibodies as described in section 2.4.4.10 showed no difference between cases A, B, C and unrelated controls. Therefore proteomic analysis by mass spectrometry was undertaken in collaboration with the Lamond laboratory at the University of Dundee as described in section 2.4.4.9. A reduction in *ABCC4* expression was shown in platelets from the affected siblings compared with eight controls (p = 0.002, unpaired, two-tailed Student's t-test, Figure 69d). There was no difference between cases and controls in expression of a range of other platelet proteins including GPVI, MYH9, GP1BB and P-selectin (Appendix 8.2.8), or a range of proteins involved in the regulation and downstream effects of cAMP (Table 22).

5.7.2 Platelet RNA-seq

RNA-seq studies of bone marrow-derived MK and platelets has shown that exon 18 is present in all major *ABCC4* isoforms (Figure 69 b, c), therefore this variant is predicted to alter ABCC4 function in both cell types [103]. RNA was extracted from platelets from case A, B and C. RNA-seq libraries were prepared and sequenced as described in section 2.5.2.1. Dr. Denis Seyres of the University of Cambridge performed read alignment using STAR [180] and featureCounts [181] software with respect to Ensembl v.70, of reference genome build GRCh37.

The variant was confirmed by RNA-seq of platelets in all reads, therefore this variant is predicted to alter ABCC4 function in platelets of affected cases. Dr. Seyres then used the DEXSeq package to determine relative exon usage for each transcript between cases and controls [240], which applies the χ^2 likelihood-ratio test to compute the p-value. Expression was lower in cases A and B compared to wild type sibling case C and five control samples run on the same sequencing lane. This was not significant at gene-level, but at exon- level usage of exon 2 was significantly lower in cases (p = 2.2 x 10-6, transcript ENST00000376887.4, Figure 70). This may be due to expression of this exon in only a subset of isoforms, allowing reads to be accurately assigned to transcripts, thus increasing statistical power.

Expression of transcripts and genes was quantified using MMSEQ and differential expression was assessed using MMDIFF by Dr. Ernest Turro of the University of Cambridge [69, 70]. This package derives estimates of gene-level expression and associated uncertainty and uses a Bayesian regression-based model to derive a posterior probability of differential expression (where a posterior probability greater than 0.5 was used to declare a transcript or gene differentially expressed. The size and directionality of the effect in terms of normalised expression values is denoted by η, the estimated log fold change.

The gene list was ranked by probability of differential expression and checked for the presence of a list of genes involved in the regulation of or downstream effects of cAMP in platelets. *ABCC4* was ranked #459, with a posterior probability of differential expression of 0.27, η -0.88, where a probability > 0.5 is

considered significant and η represents the direction and size (estimated log fold change) of the differential effect in cases compared with controls. Therefore by this method there was no overall differential expression of ABCC4 RNA expression in cases vs. controls. No other genes known to be involved in cAMP regulation were significantly differentially expressed, nor did they feature in the 500 top-ranked differentially expressed list (Table 21, Table 22). GPVI was the highest-ranking cAMP-related gene at #506, with a posterior probability of differential expression of 0.25, η+0.94. These findings suggest that *ABCC4* variant RNA is subject to minimal nonsense-mediated decay.

Patient	A	B	C	F (age 14yrs)	Normal range
PFA-100® Collagen/Adrenaline	**216s**	**219s**	174s	NA	<179s
PFA-100® Collagen/ADP	130s	134s	104s	NA	<134s
PAP-8E®: ADP	**Primary phase followed by disaggregation at 2µM and 3µM. Normal at 10µM**	**Primary phase followed by disaggregation at 2µM and 3µM. Normal at 5µM and 10µM**	**Primary phase followed by disaggregation at 2µM and 3µM. Normal at 5µM and 10µM**	**Primary phase followed by disaggregation at 2µM, normal at 3µM**	
PAP-8E®: Epinephrine	**Reduced at 2µM and 3µM, normal at 10µM**	**Reduced at 3µM, normal at 5µM and 10µM**	**Reduced at 3µM, normal at 5µM and 10µM**	Normal at 3µM	
PAP-8E®: Collagen (1µg/ml, 2µg/ml)	Normal	Normal	Normal	Normal	
PAP-8E®: Ristocetin (0.5mg/ml, 1.5mg/ml)	Normal	Normal	Normal	Normal	
PAP-8E®: Arachidonic acid (1.0mM)	Normal	Normal	Normal	Normal	
PAP-8E®: U46619 (1µM)	Normal	Normal	Normal	Normal	
ATP release: ADP	**Reduced at 5µM and 10µM**	**Reduced at 5µM and 10µM**	**Reduced at 5µM, normal at 10µM**	Normal at 5µM	
ATP release: Collagen (2µg/ml)	Normal	Normal	Normal	Normal	
ATP release: Thrombin (50µM)	Normal	Normal	Normal	Normal	
Nucleotide: ATP (amol/plt)	0.51	0.40	0.60	NA	0.30-0.88
Nucleotide: ADP (amol/plt)	0.39	0.65	0.33	NA	0.22-0.59
Nucleotide: ATP/ADP ratio	1.32	1.55	1.84	NA	0.86-2.26

Table 20 Platelet testing undertaken in the clinical laboratory at the Katherine Dormandy Haemostasis and Thrombosis Unit at Royal Free London NHS Foundation Trust.
These tests were performed at least three times on each case and representative results are shown. Abnormal results are shown in bold.

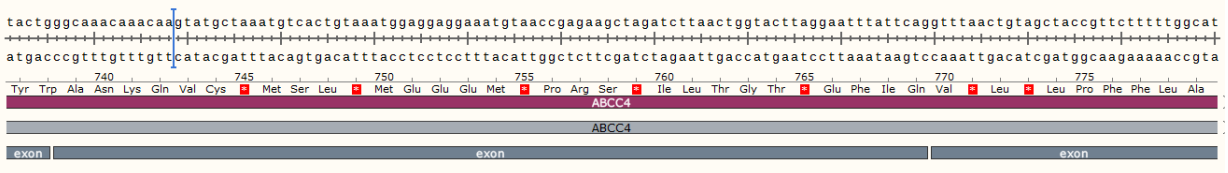


Figure 68 Modelling of the single base pair deletion in exon 18 of ABCC4.
The site of the 1bp deletion is indicated in blue. Following introduction of the deletion, the frameshift leads to a stop codon (indicated by *) after two amino acids and multiple subsequent stop codons in the sequence.

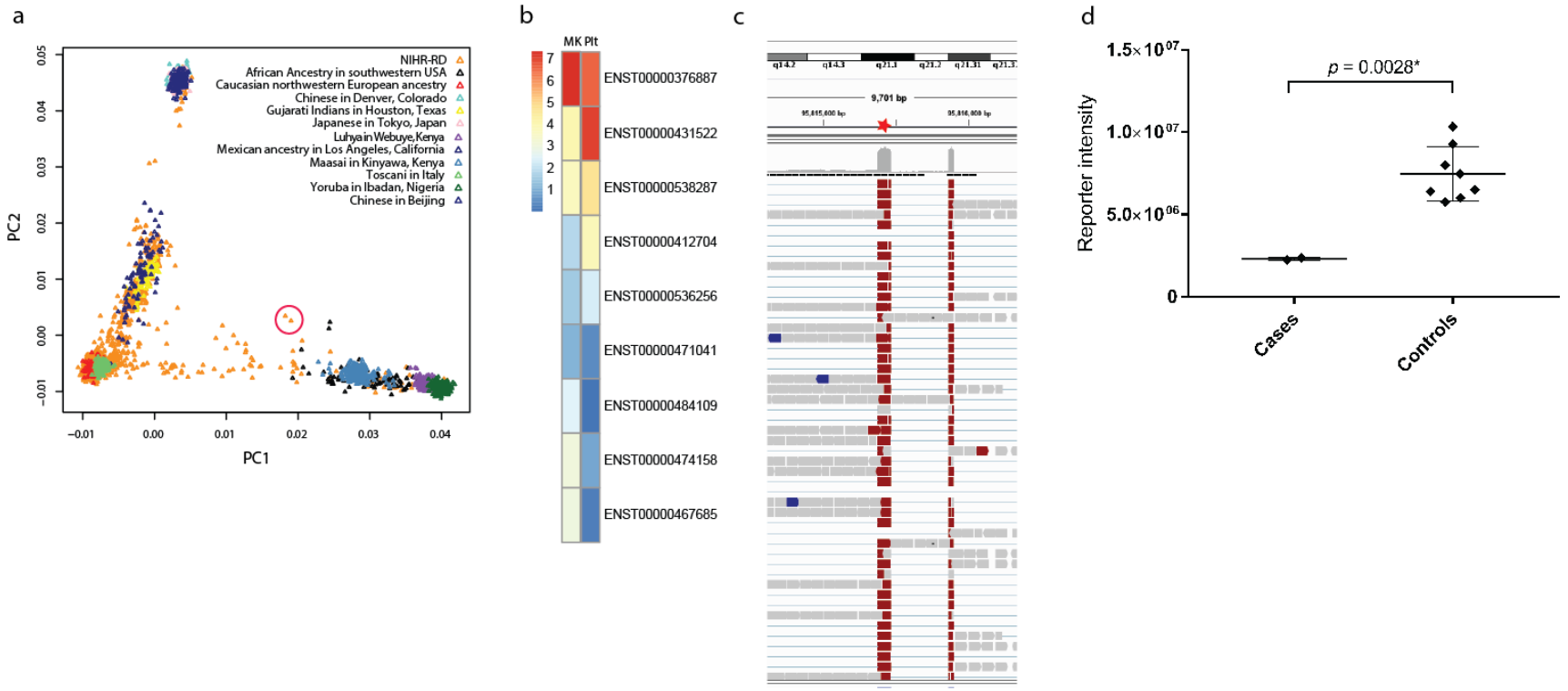


Figure 69 The *ABCC4* variant has unknown frequency in the Seychelles, is expressed in all major isoforms in platelets and MK and leads to reduced protein expression

a. Principal-component analysis plot of NIHR BR-RD cases as of June 2015, alongside data from the International HapMap Project [239]. Cases A and B are circled. The cases are from the Seychelles and are genotypically most similar to African Americans and Maasai Kenyans, but do not overlap these groups.
b. Main isoforms of *ABCC4* in platelets and MK. Normalised expression values expressed as $\log_2$ of fragments per kilobase of transcript per million reads mapped (FPKM) [103]. Heatmap courtesy of Dr. Luigi Grassi, University of Cambridge.
c. Single cell RNA sequencing of bone marrow MK shows presence of exon 18 (indicated by star) in all reads. Platelet RNA-seq from cases A and B also confirmed the presence of the variant in all reads.
d. Mass spectrometry of platelet protein pellets was performed by the Lamond laboratory at the University of Dundee. Expression of ABCC4 was significantly reduced in Case A and B compared to eight control samples on the same run (p = 0.002, unpaired Student's t-test).

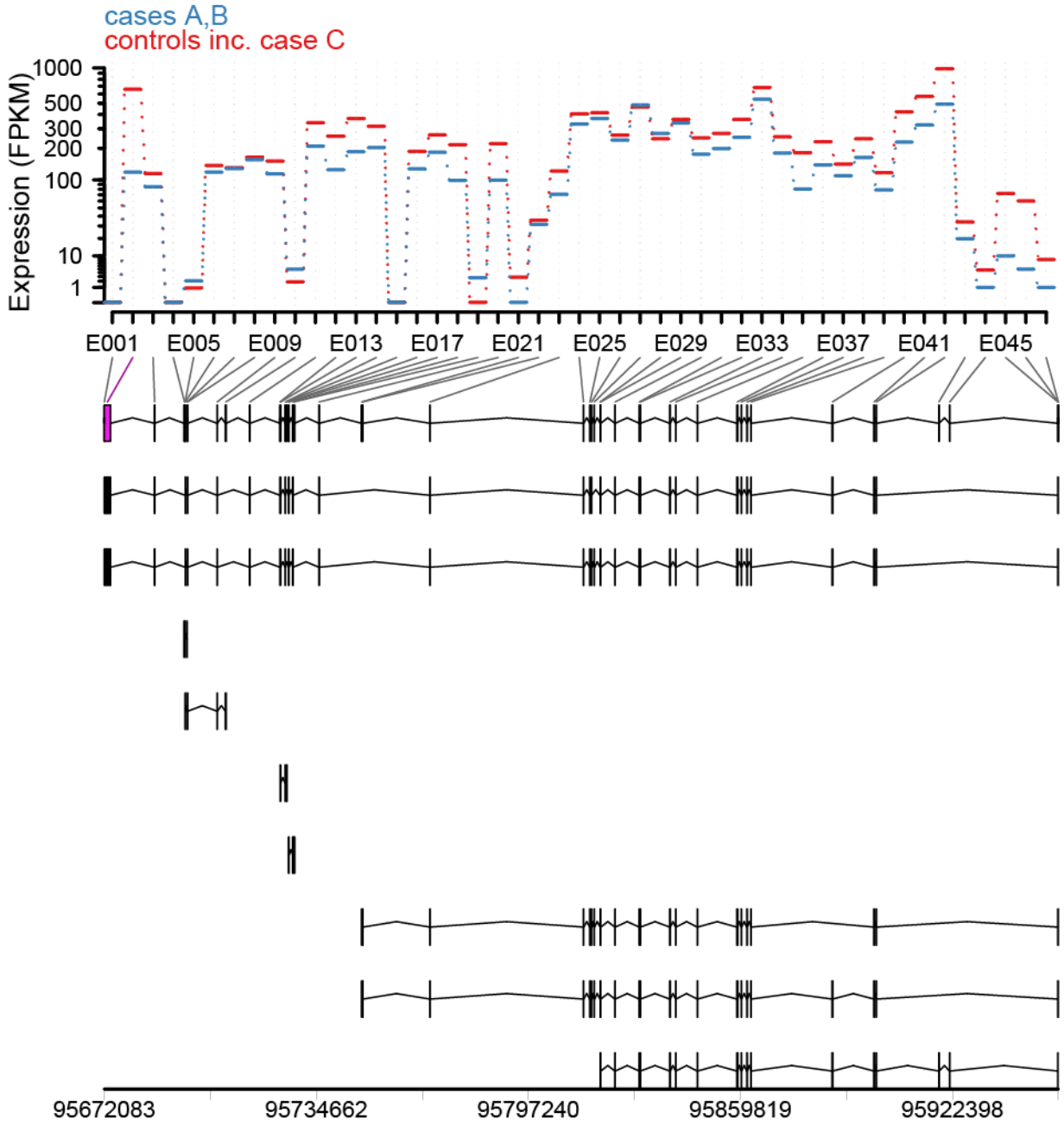


Figure 70 RNA-seq of platelets from cases A and B compared with 6 control samples including unaffected sibling (case C).
All samples were run on the same sequencing lane. Expression values (FPKM) are shown for each numbered exon (prefix E-). Figure courtesy of Dr Denis Seyres, who used the DEXSeq package to determine relative exon usage for each transcript between cases and controls [240], which applies the χ^2 likelihood-ratio test to compute the *p*-value. Cases A and B have lower expression across the board, but this only reached significance for exon 2, indicated in purple (corrected p = 2.2 x 10-6, transcript ENST00000376887.4)

gene_name	posterior_ probability	ɳ
ARID5A	0.97	-0.90
RILP	0.96	-0.83
BID	0.95	-1.12
ZNF606	0.87	-2.34
HIST1H3J	0.87	1.97
RPS17	0.85	3.56
RP5-837J1.2	0.81	2.33
RP11-182L21.2	0.80	-0.99
RP11-414K1.3	0.79	-1.83
KDM6B	0.77	-1.07
TMX4	0.76	-0.86
INPPL1	0.76	-0.64
TPST2	0.75	-0.72
CDR2L	0.75	-1.30
DDX47	0.74	1.07
RPS17L	0.73	-2.61
GTF2H2	0.73	1.86
MPHOSPH8	0.72	0.72
HMG20B	0.72	-0.68
COG3	0.71	1.15

Table 21 Top twenty ranked genes following differential gene expression analysis.
The posterior probability is a derivative of the Bayes factor. ɳ represents the size and the direction of the effect.

Differential expression by RNA-seq				Differential expression by proteomics	
gene_name	posterior_ probability	η	Rank	Signal detected?	*p* - value
ADCY3	0.12	-0.79	2588	Yes	0.38
ADCY5	0.08	0.56	42812	No	
ADCY6	0.15	-0.75	1459	Yes	0.11
PDE2A	0.07	-0.47	45394	No	
PDE3A	0.02	-0.15	55635	Yes	0.29
PRKACA	0.08	-0.40	41227	No	
PRKACB	0.02	0.15	55516	No	
PRKAR2A-AS1	0.15	-0.86	1489	No	
PRKAR2A	0.04	-0.27	52700	Yes	0.40
PRKAR1AP	0.10	0.18	8103	No	
PRKAR1A	0.02	-0.07	56282	Yes	0.24
PRKACG	0.09	-0.29	35768	No	
PRKAR2B	0.06	-0.37	48047	Yes	0.98
RAPGEF3	0.12	-0.69	2278	No	
RAPGEF4-AS1	0.09	-0.08	38098	No	
RAPGEF4	0.05	-0.15	49794	No	
GP6	0.25	-0.89	506	Yes	0.75
P2RY12	0.06	0.44	46239	No	
P2RY13	0.06	0.47	46144	No	
VASP	0.19	-0.62	921	Yes	0.38
FLNA	0.15	-0.85	1493	No	
CALD1	0.09	-0.60	40472	Yes	0.17
HSPB1	0.04	-0.32	50436	Yes	0.41
LASP1	0.24	-0.48	601	Yes	0.26

Table 22 Summary of RNA differential expression and proteomic analysis of genes in the cAMP pathway. Genes other than ABCC4 implicated in cAMP regulation and metabolism in platelets were not highly ranked in the RNA differential expression analysis and none achieved a posterior probability > 0.5.
The columns titled 'Proteomics' shows proteins that were detected by mass spectrometry. There was no difference between cases and controls for these proteins by unpaired two-tailed Student's t-test, assigning significance to values <0.05.

5.8 *ABCC4* loss of function does not associate with abnormal dense granules

As discussed, ABCC4 has been associated with disorders of dense granules and influence ADP uptake to the dense granule [207] Table 20 showed that whole-platelet ATP and ADP levels, quantified by luciferase assay, were within normal limits. Transmission scanning electron microscopy (TEMS) was performed to further explore this possibility. Samples for TEMS were processed as described in section 2.4.4.6. and Prof. Kathleen Freson provided support in interpretation of the images. Dense granules were plentiful in the *ABCC4* wild-type sibling C and *ABCC4* homozygous variant siblings A and B (Figure 71). These data show that *ABCC4* is not essential for dense granule formation or ADP uptake.

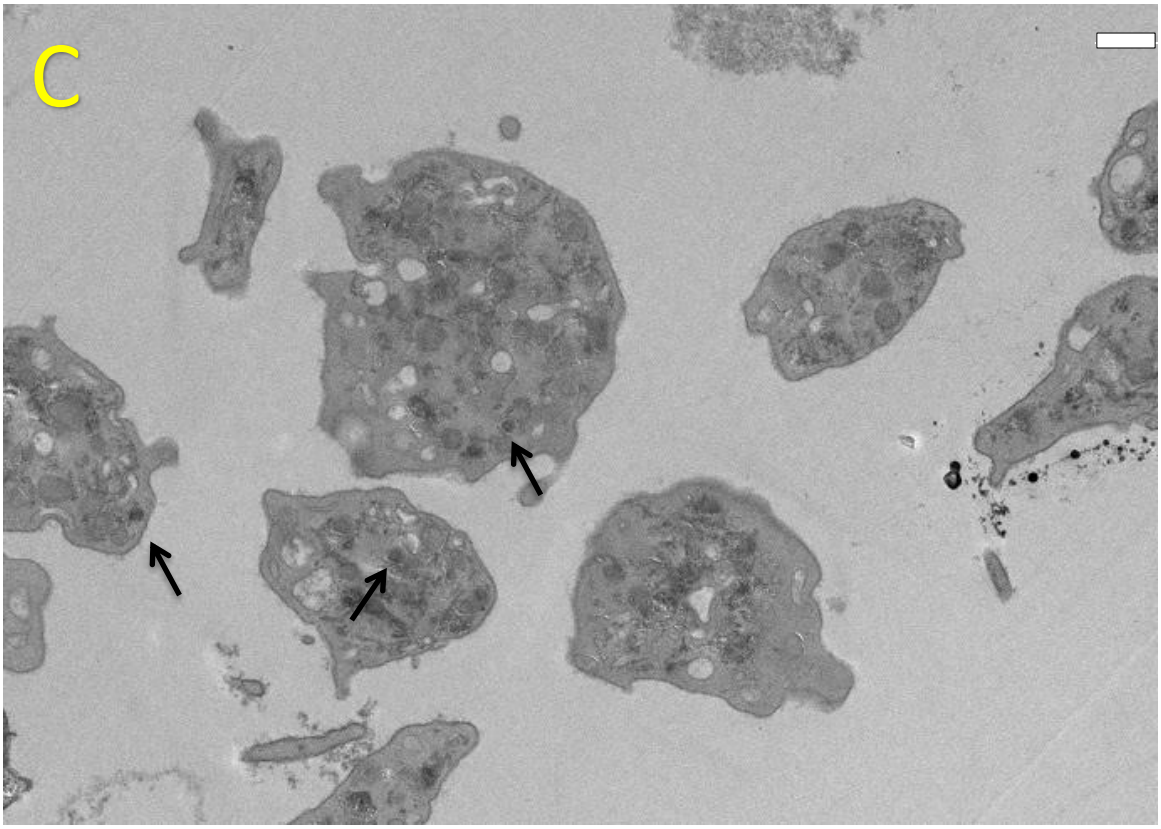
5.9 *ABCC4* and *P2RY12* variants are associated with increased intra-platelet cAMP

Reduced *ABCC4* expression has been associated with increased intracellular cAMP in platelets and other cells. We assessed this by cAMP ELISA as described in section 2.4.4.5. Blood was drawn on two occasions from cases A, B, C and controls. Each draw was tested twice to a total of four experiments. Each experiment was plated in triplicate and individual values are plotted in Figure 72a. Results were interpreted using the unpaired, two-tailed Student's t-test.

Baseline cAMP was higher in *ABCC4* homozygous cases A and B (mean 117 fmol/10^7 platelets, s.d 39.1) than *ABCC4* wild-type case C and four controls (mean 52 fmol/10^7 platelets, s.d 43, p <0.0001). There was no difference between cAMP levels in Case C (*ABCC4* wild-type, *P2RY12* heterozygous) and controls suggesting that the *P2RY12* variant does not influence unstimulated cAMP levels (p = 0.79, Figure 72b).

Blood was also drawn once and tested twice in triplicate from *ABCC4* heterozygous offspring case F and Figure 72c shows a small increase in cAMP level (mean 11.2 fmol/10^7 platelets, s.d 1.39) compared with the day control (mean 7.3 fmol/10^7 platelets, s.d 1.74, p =0.04). Case F does not report excessive bleeding but is a young man and has experienced few haemostatic challenges, further limiting the conclusions that can be drawn from this observation.

In order to assess whether the *P2RY12* variant contributes to the bleeding phenotype in this pedigree it is necessary to determine whether there is reduction of the normal ADP-P2RY12 mediated inhibition of cAMP synthesis during platelet activation. Platelets from case C (*ABCC4* wild-type, *P2RY12* heterozygous) were incubated with Iloprost as described in section 2.4.4.5 to maximise the intracellular cAMP levels, stimulated with ADP and then collected for cAMP ELISA. At 10μM ADP the cAMP response was blunted for case C (*ABCC4* wild-type, *P2RY12* heterozygous, mean 74 fmol/ 10^7 platelets, s.d 13.0) compared with the day control (mean 9 fmol/10^7 platelets, s.d 4.1, p<0.0001, Figure 72d). Therefore the *P2RY12* variant appears to damage the integrity of the P2RY12-mediated signalling pathway and contributes to the observed elevation in cAMP during platelet activation.



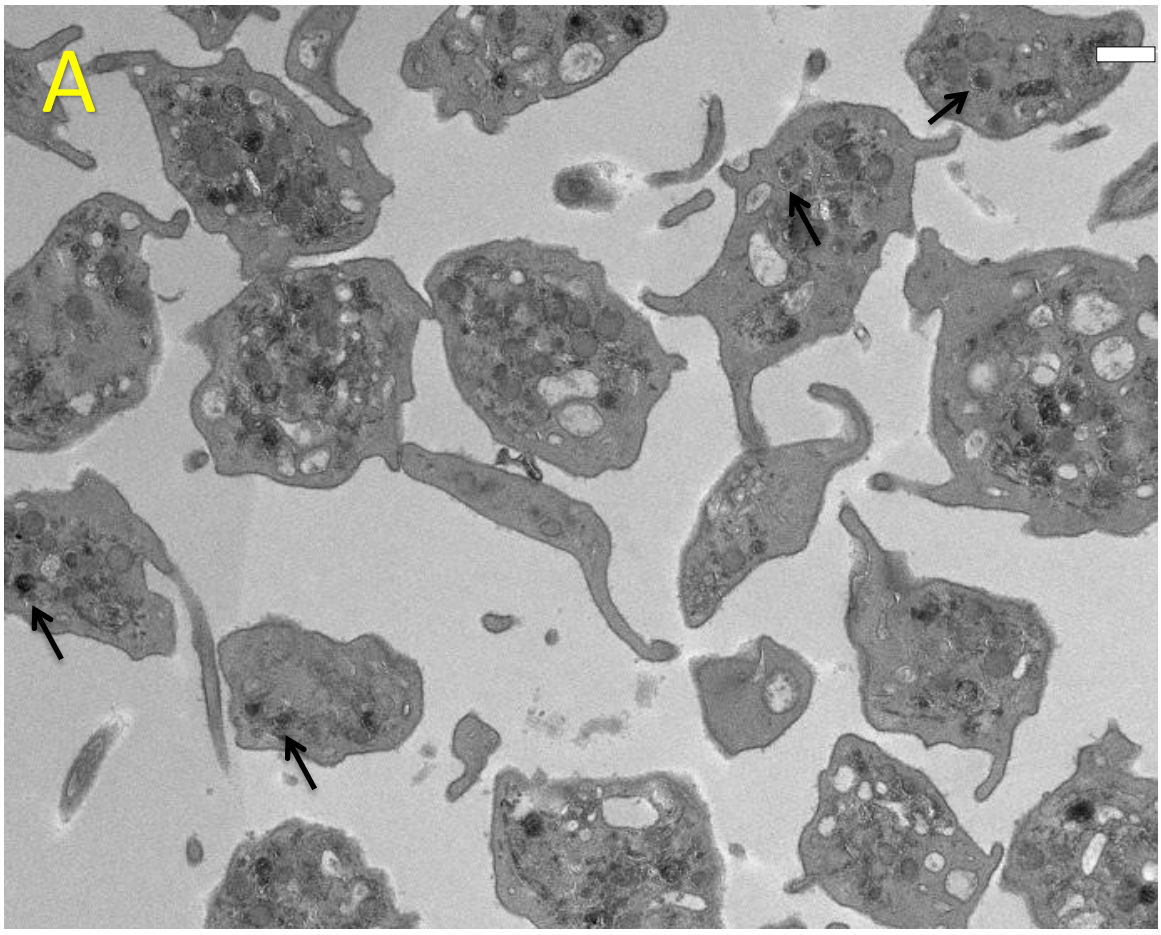


Figure 71 TEMS of platelets from case C (wild-type *ABCC4*) and A (homozygous *ABCC4* variant).
Dense granules are evident in both (arrows), despite some evidence of platelet activation in A. White scale bar indicates 0.5µm.

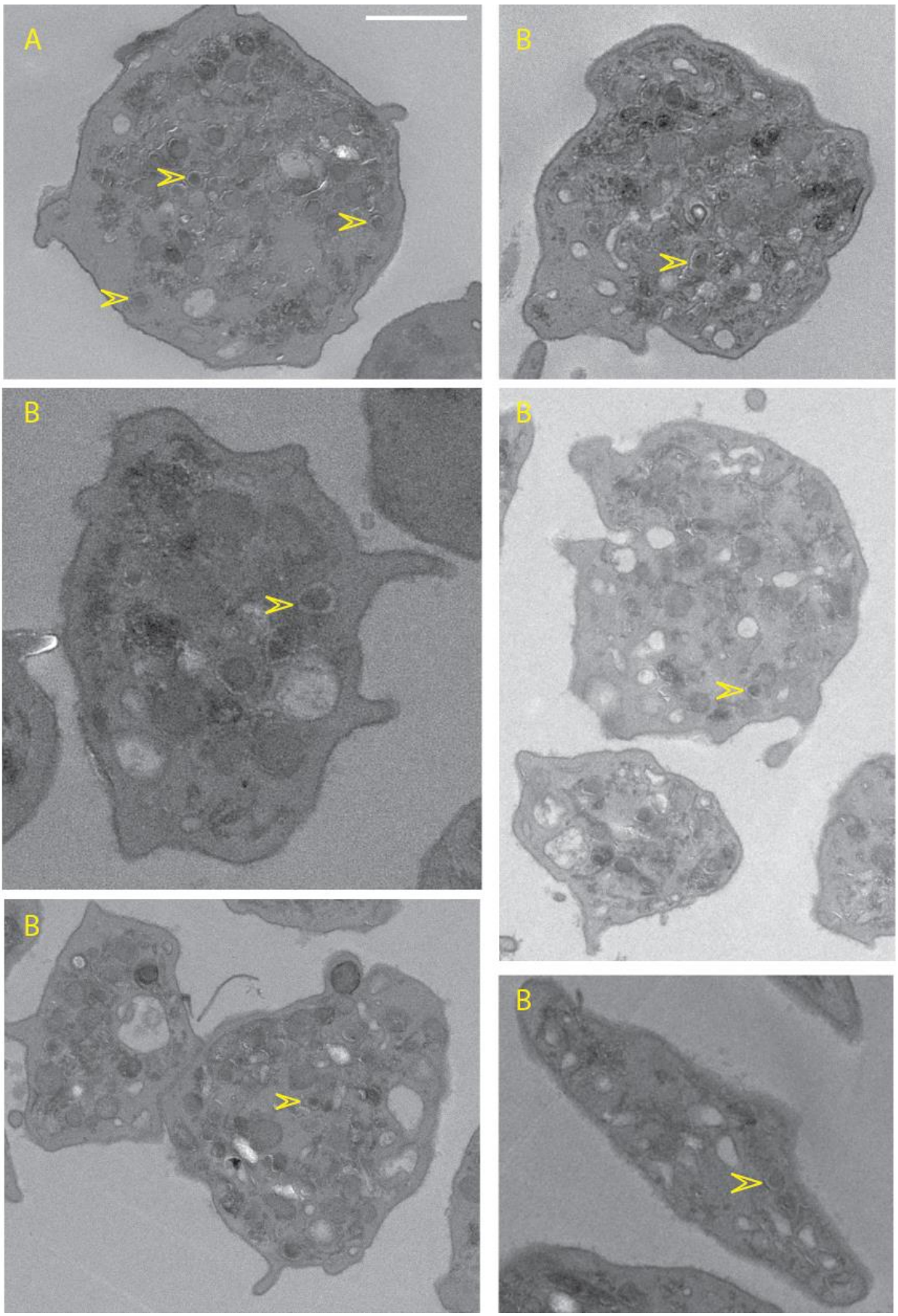


Figure 71 continued, arrows indicate dense granules in cases A and B (homozygous *ABCC4* variant). White scale bar indicates 0.5µm.

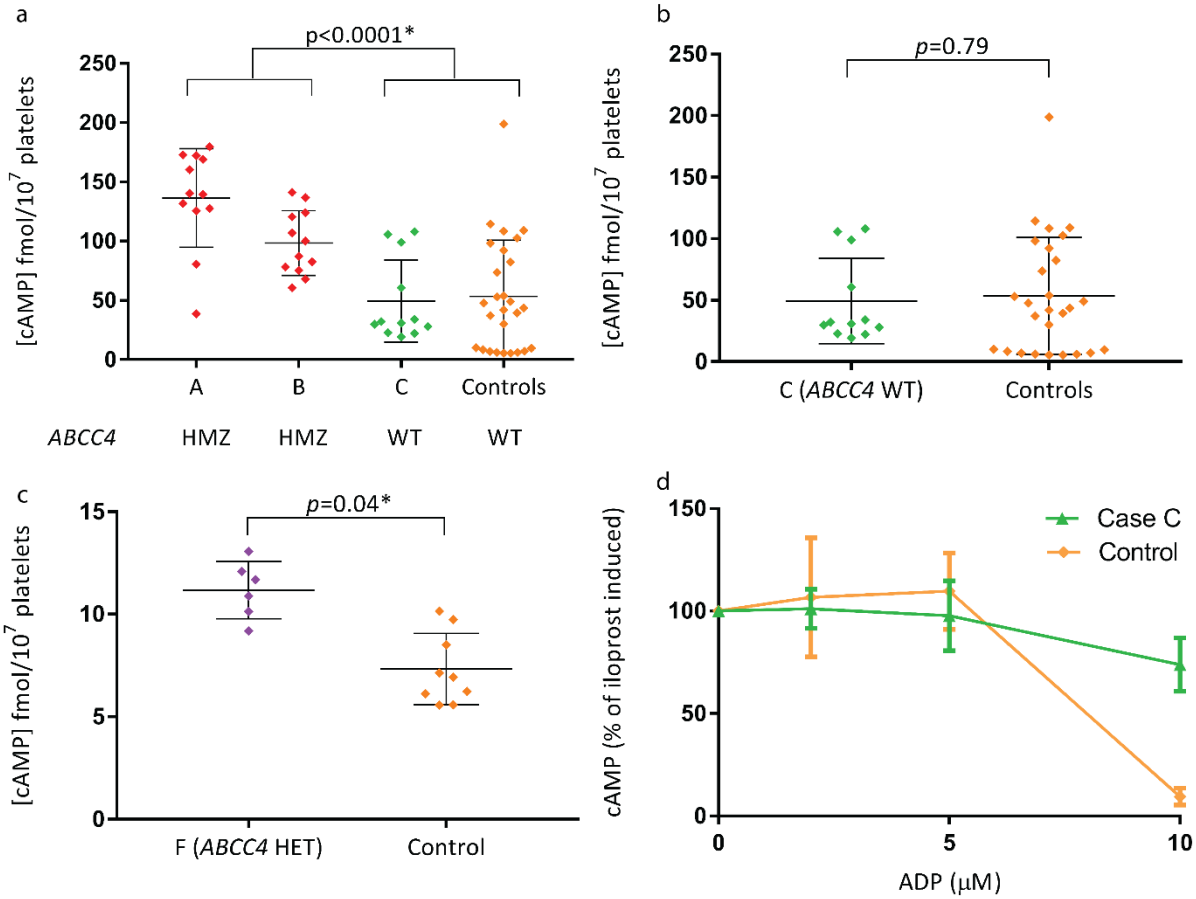


Figure 72 *ABCC4* and *P2RY12* variants influence intra-platelet cAMP by ELISA.
Results were interpreted using the unpaired, two-tailed Student's t-test.
a. cAMP was higher in *ABCC4* homozygous variant cases (A,B) than the *ABCC4* wild-type sibling (C) and three controls, p <0.0001*.
b. There was no difference between cAMP levels in Case C (*ABCC4* wild-type, *P2RY12* heterozygous) compared with controls suggesting that the *P2RY12* variant does not influence unstimulated cAMP levels.
c. There was a small, borderline significant difference between cAMP levels in Case F (*ABCC4* heterozygous, *P2RY12* wild-type) and control, suggesting that *ABCC4* haploinsufficiency may increase cAMP levels.
d. Platelets were incubated with iloprost to maximise the intracellular cAMP levels then stimulated with ADP to determine the ability of P2RY12-mediated signalling to lower intra-platelet cAMP. At 10μM ADP the response to ADP was significantly reduced for Case C *(P2RY12* heterozygous, *ABCC4* wild-type) compared to control (p <0.001). This implies that the *P2RY12* variant impairs inhibition of the cAMP pathway during platelet activation.

5.10 Assessment of plasma cAMP

Plasma samples were submitted to Metabolon for global metabolic profiling as described in section 2.5.6, and compared with 56 controls between the ages of 16 and 64 years of age without rare, moderate or high impact variants in *ABCC4,* that were recruited to the NIHR BR-RD BPD. Plasma cAMP was below the limit of detection in cases A and B but detectable in controls C and D. However, cAMP was also undetectable in four other controls: case CAV000152P and CAV000146L (both with mild thrombocytopenia and mucocutaneous bleeding, unrelated), CAV000157F (unclassified mucocutaneous bleeding) and 5D7000146Q (impaired platelet aggregation with collagen). Discrepant phenotypes suggests non-biological causes for the undetectable plasma cAMP in these cases; sample handling may be contributory, e.g. prolonged transit, as all four samples were shipped over some distance from Cardiff (prefix CAV-) or Newcastle (prefix 5D7). This could disproportionally reduce cAMP due to the action of phosphodiesterases. These findings limit the interpretation of this result and repetition of this assay and examination of the intra-cellular cAMP concentration should be undertaken to explore the Cardiff and Newcastle cases further.

ABCC4 is thought to transport a variety of endogenous substrates involved in steroid, bile acid, glutathione and uric acid metabolism. 42 compounds involved in these pathways were assessed (Appendix 8.2.8.) Oestrogen and progesterone-related steroids were excluded as the cohort was not matched for gender or menopausal status. None of these compounds were absent in cases A and B and present in day controls cases C and D. Multiple unpaired two-tailed Student's t-tests for these compounds using the Holm-Sidak method, where $p = 0.05$ as was considered significant did not detect any other differences between cases and controls. In the entire dataset of 954 named compounds there were no other compounds that were present in the day controls (siblings C and D) and absent in cases B and C.

5.11 Impaired GPVI expression and response to collagen did not co-segregate with *ABCC4* genotype

Mouse knockout experiments have generated conflicting results regarding the influence of ABCC4 on collagen-mediated platelet aggregation. One study showed a reduction in collagen-mediated aggregation and reduced surface-expression of GPVI [211]. Light transmission aggregometry and platelet nucleotide release studies performed in the clinical laboratory showed no impairment of collagen-mediated activation using horm collagen at 1 and 2µM concentrations (Table 20).

Light transmission aggregometry was repeated with assistance from Ms. Chantal Thys of the University of Leuven using lower concentrations of platelet agonist horm collagen as described in section 2.4.4.6. Cases A, B, C and F, regardless of genotype, all exhibited reduced aggregation than the day control (Figure 73a). Platelet aggregation was repeated following horm collagen stimulation, this time using a pre-incubation with iloprost to to stimulate intra-platelet cAMP, also as described in section 2.4.4.6. It was hypothesised that iloprost-induced inhibition of aggregation would be attenuated in cases A and B

due to elevated unstimulated cAMP. However inhibition of aggregation was marked in all pedigree cases compared with control, regardless of genotype (Figure 73b). Limited conclusions can be taken from this single experiment in isolation, using a single control.

In collaboration with Ms. Joana Batista of the Cell Function Team in the laboratory of Prof. Ouwehand at the University of Cambridge, platelet activation was further assessed by flow cytometry in response to CRP-XL (a specific GPVI agonist) and ADP. The results of cases A, B and C were compared with a bank of controls; case F was unavailable for testing. All cases had reduced platelet activation in response to ADP, consistent with the results of the clinical tests and the presence of the *P2RY12* variant in all cases (Figure 74a). There was no co-segregation of impaired platelet activation following stimulation with CRP-XL with *ABCC4* genotype; in fact all pedigree members exhibited impairment, which was most marked in case C (Figure 74b). Another recall is planned to confirm these findings.

Ms. Batista also tested GPVI expression in platelets from case A, B and C using a monomer- and dimer-specific antibody as described in section 2.4.4.4. There was no difference between cases A and B and the bank of controls. The cases were all confirmed to have the reference alleles for the 'PEALN' pentad of *GPVI* (Ser199Pro, Lys217Glu, Thr229Ala, Gln297Leu and His302Asn) which is associated with reduced GPVI expression and a lower functional response to CRP-XL [241].

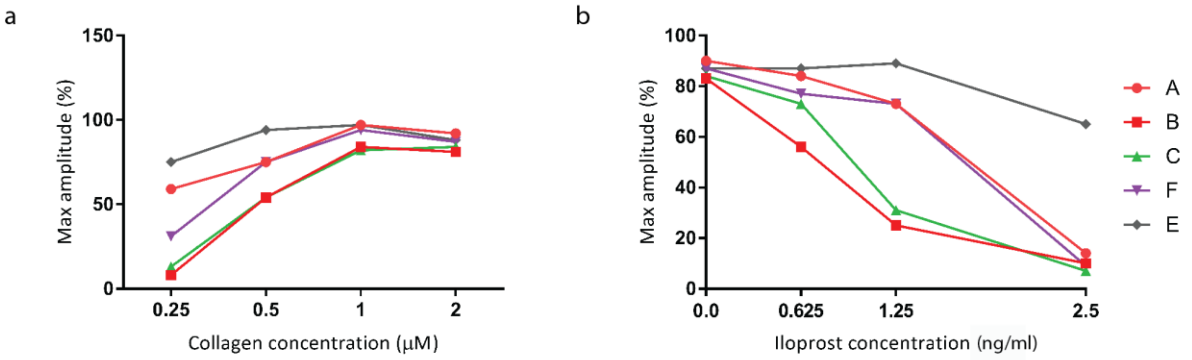


Figure 73 Light transmission aggregometry following stimulation with horm collagen.
Performed with assistance from Ms. Chantal Thys of the University of Leuven.

(a) All pedigree members regardless of genotype had reduced aggregation on stimulation with horm collagen than the control, most marked at the lowest concentrations.

(b) Inhibition of aggregation by Iloprost was also more marked in all pedigree members, regardless of genotype.

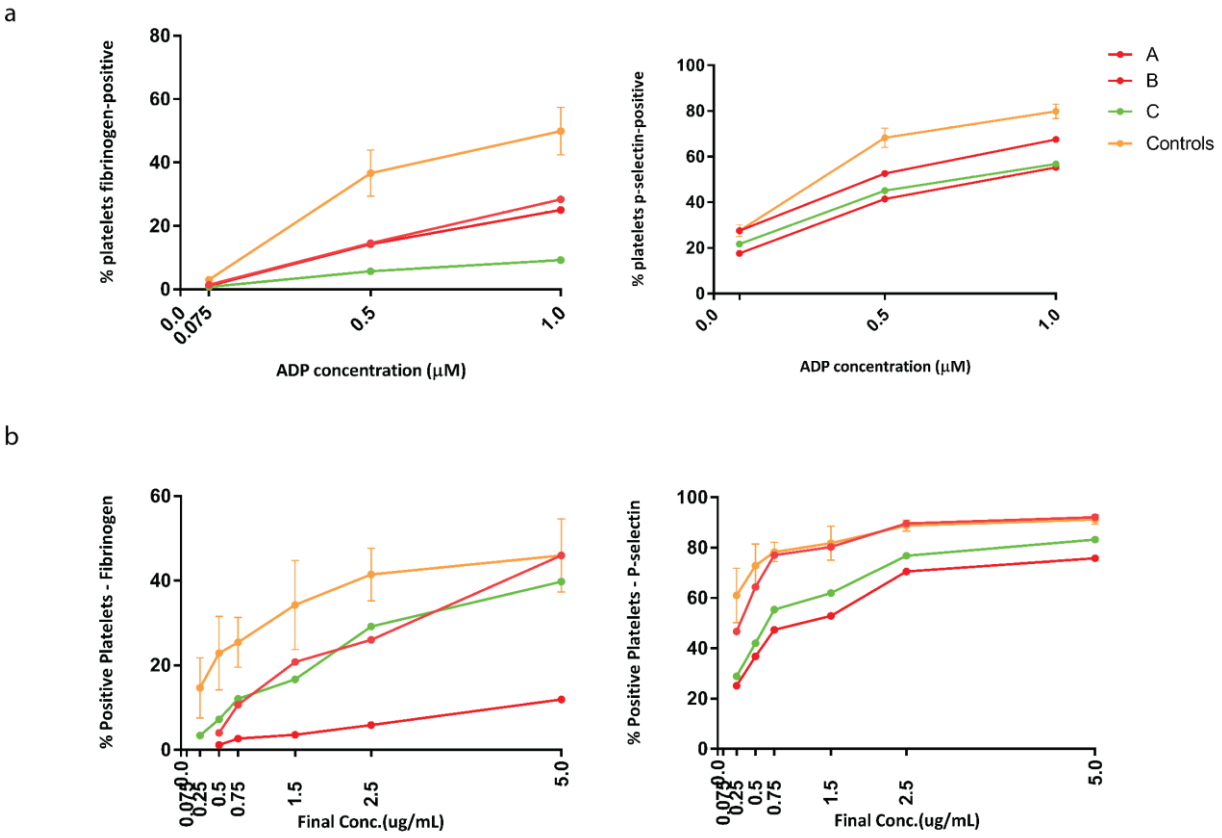


Figure 74 Platelet function by flow following stimulation with CRP-XL and ADP.
Performed in collaboration by Ms. Joana Batista of the Cell Function team at the University of Cambridge.

(a) Cases A, B and C had reduced platelet activation measured by surface expression of P-selectin and fibrinogen-binding following stimulation with ADP. This is consistent with the clinical testing and the presence of the *P2RY12* variant in all cases. Results were compared with a day control and previously obtained results for a bank of controls

(b) Reduced platelet activation following stimulation with CRP-XL to specifically activate the GPVI pathway did not co-segregate with *ABCC4* homozygous variant genotype (red); all pedigree members exhibited reduced P-selectin and fibrinogen-binding following stimulation with CRP-XL

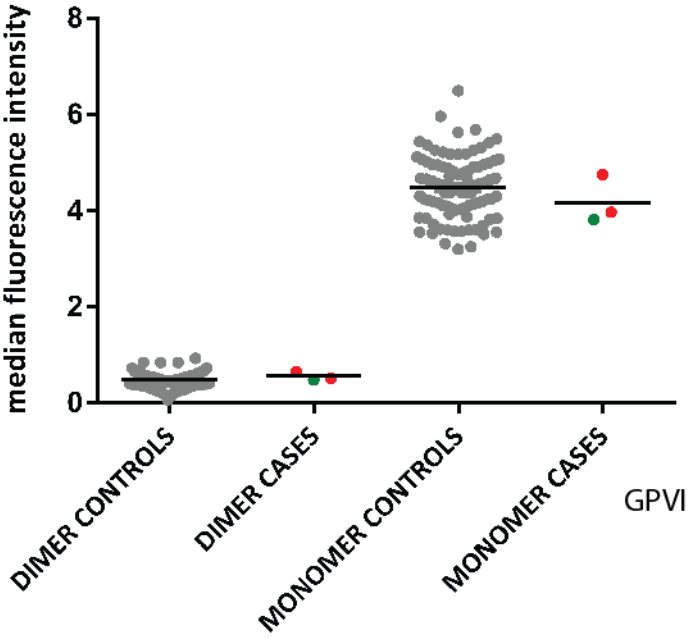


Figure 75 There was no significant difference between *ABCC4* homozygous variant cases (red) and controls in terms of platelet GPVI surface expression.
This was measured using a purchased monomeric antibody (HY101) and a dimer- specific antibody that was a gift from Dr. Stephanie Jung of the Biochemistry Department of the University of Cambridge. Case C (*ABCC4* wild-type) is shown in green.

5.12 ABCC4 contains regulatory variants that influence platelet traits

The following results were obtained in collaboration with Mr. Roman Kreuzhuber of the laboratory of Professor Willem Ouwehand at the University of Cambridge who performed the eQTL and MEIS1 analyses. Dr. Romina Petersen analysed the MK enhancer data. All variant locations are determined using Ensembl v.70, genome build GRCh37.

As previously discussed, *ABCC4* carries a GWAS signal for platelet count (PLT) and crit (PCT) [108, 111]. Variant 13:95901241 C/T (*rs4773860*, MAF 0.481) was associated with an increased platelet crit (PCT, $p = 2.9 \times 10^{-30}$) and variant 13:95899716 C/A (*rs4148435*, MAF 0.084) was associated with increased platelet count (PLT, $p = 9.1 \times 10^{-30}$).

Enhancers are DNA elements that activate transcription by interaction with gene promoters [242]. Enhancers may lie in or at considerable distance from the gene they regulate and are identified by high density of bound transcription factors and their characteristic chromatin modifications H3 lysine 27 acetylation (H3K27ac) and H3 lysine 4 monomethylation (H3K4me), as measured by chromatin immunoprecipitation and sequencing (ChIP-seq) [92]. Superenhancers are defined as regions containing multiple enhancers in close proximity.

An MK-specific superenhancer was identified by Dr. Petersen that overlaps *ABCC4* and interacts with its promoter. The GWAS variants lie in this superenhancer. This suggests that transcription factor binding in *ABCC4* regulates its expression in MK, and that allelic variation at the same locus alters platelet traits.

Expression quantitative trait loci (eQTL) are genetic loci at which allelic variation influences RNA expression. Platelet RNA expression datasets obtained by Dr. Kate Downes of the laboratory of Professor Willem Ouwehand at the University of Cambridge (156 samples) and Professor Michel Georges of the University of Liège (Correlated Expression and Disease Association Research (CEDAR) Project, 268 samples) were processed to allow comparability and examined for eQTL.

The strongest eQTL in the whole cohort was identified in *ABCC4* (*rs4148436*, 13:95899607 A/G intron variant, MAF 0.37 in Europeans, 0.38 in Africans [95]). The median gene expression at each genotype was modelled using linear regression to derive the β (slope) of 0.61, indicating that the AA genotype is associated with 2.33 fold lower expression than GG (Figure 76b). The *p*-value, under the null hypothesis that β = zero, was 1.7×10^{-22}. A statistical analysis using FINEMAP software [243] that examined the posterior probability of colocalisation between GWAS and eQTL variants gave $p = 0.73$, with $p > 0.05$ set as the threshold for significance. This indicates high linkage disequilibrium between eQTL *rs4148436* and the GWAS variant for PCT, suggesting that higher *ABCC4* expression drives a lower PCT ($\beta = -0.024$) and PLT ($\beta = -0.018$).

Furthermore, rs4148436 lies in a MEIS1 binding site. MEIS1 is an MK-specific homeobox transcription factor and is essential for normal megakaryopoiesis [21, 244, 245]. Assessment of in-house CHIP-seq data from a CHRF-288-11 cell line (acute megakaryoblastic leukaemia) that was heterozygous for *rs4148436* showed allele-specific binding of MEIS1 at that location (p = 2.27 x10^{-7}) [245]. The eQTL genotype associated with low *ABCC4* expression (A/A) was associated with high MEIS1 binding. This suggests a mechanism by which MEIS1 represses *ABCC4* expression in order to influence thrombopoiesis.

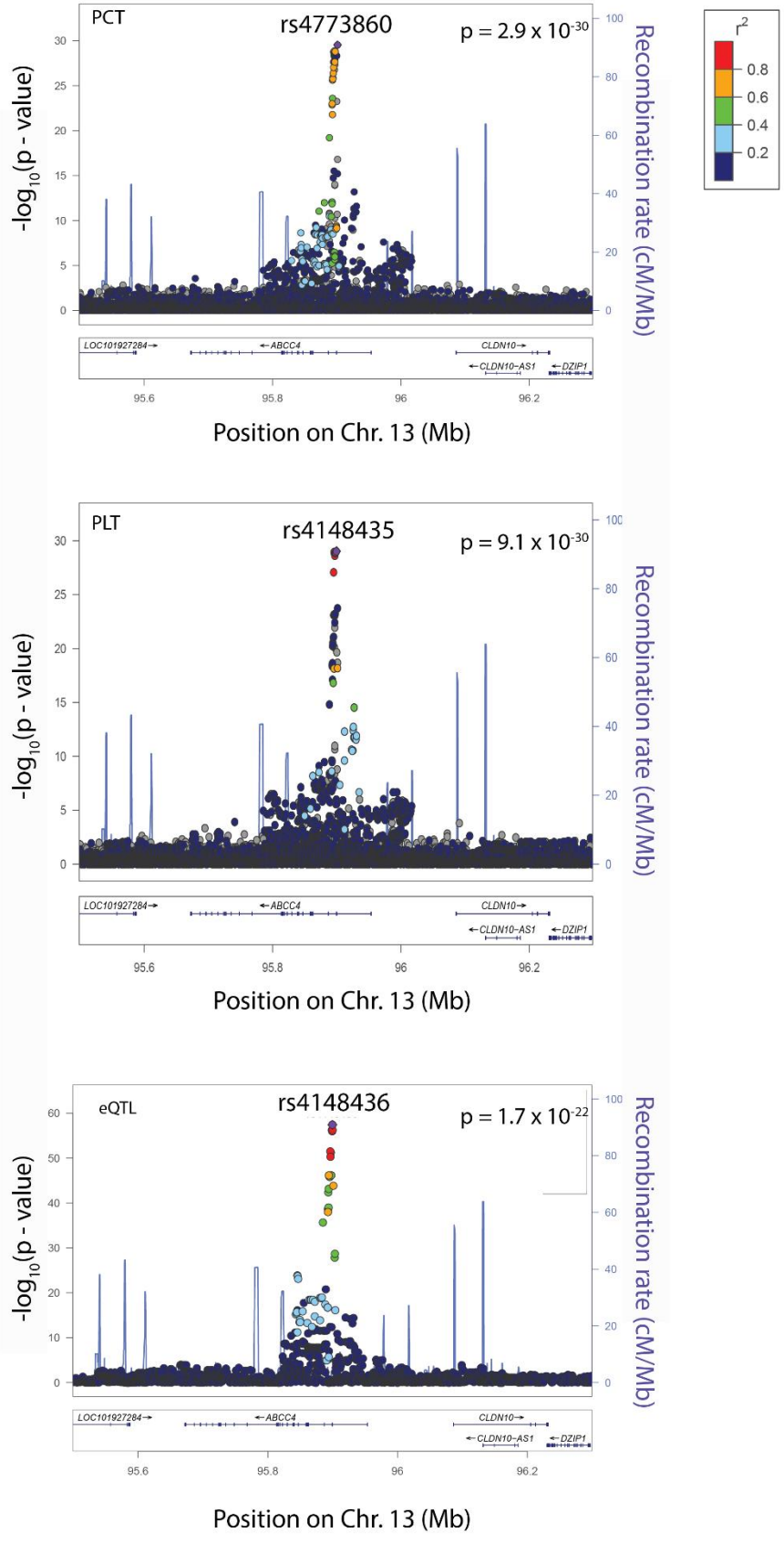


Figure 76 *ABCC4* contains a regulatory element that may influence thrombopoiesis
LocusZoom plots [246] showing lead GWAS SNPs for PCT and PLT [108, 111] and lead eQTL SNP. Data points show *p*-values of association between allelic variation at the loci shown and either platelet traits (upper, middle panel) or RNA expression (lower panel). Colours represent r^2, a measure of linkage disequilibrium from lead SNP. High r^2 suggests loci are co-inherited with low recombination in-between loci. Blue peaks show loci of recombination. The similar recombination pattern in all plots shows the proximity of these variants in the

genome. All variants lie within an MK-specific superenhancer. Plots courtesy of Dr. Roman Kreuzhuber of the University of Cambridge.

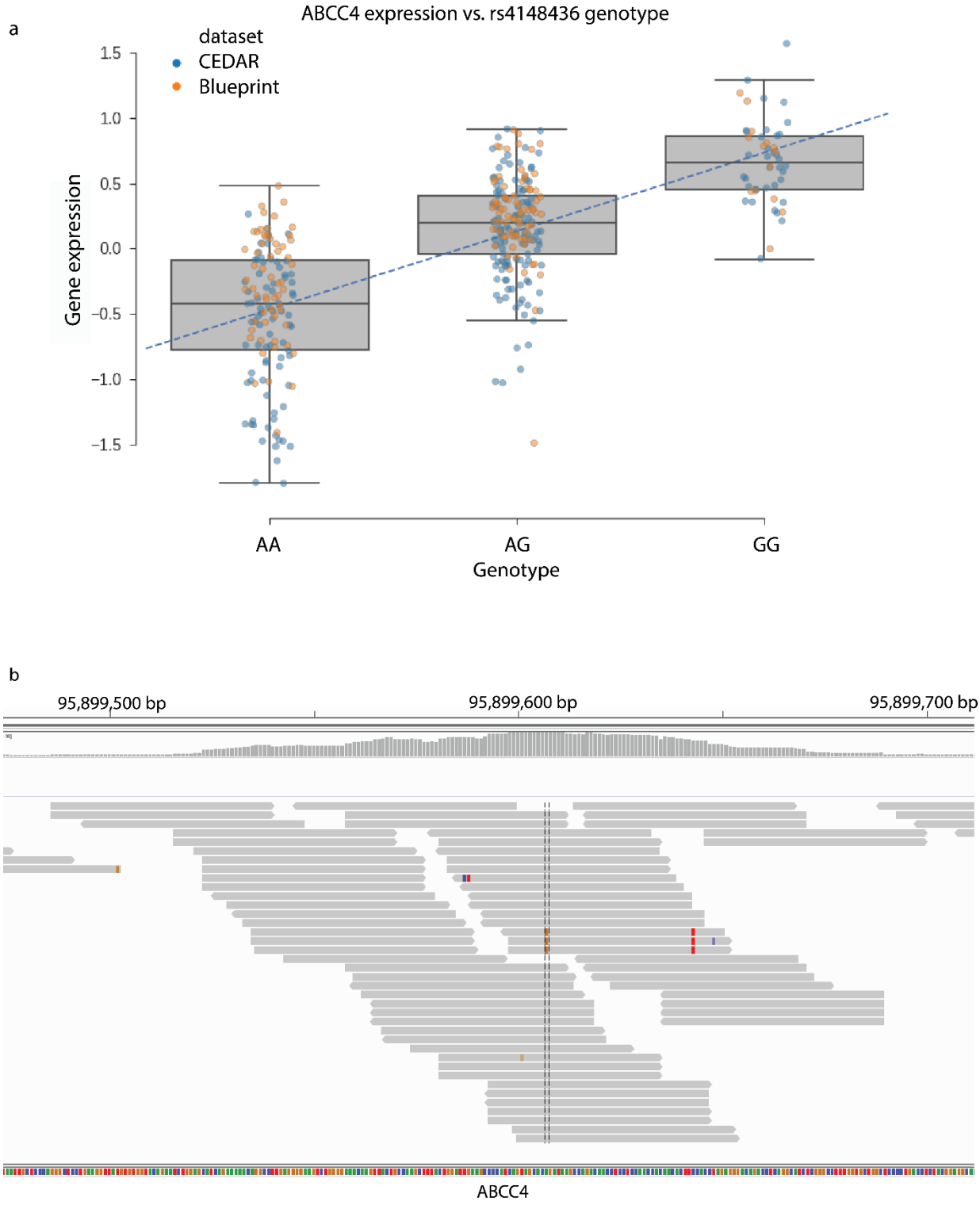


Figure 77 *ABCC4* contains an eQTL that shows allele-specific binding of transcription factor MEIS-1

(a) Boxplot showing differences in expression (log_2-transformed fluorescence intensities) between AA/ AG/ GG genotypes at ABCC4 eQTL. The median gene expression at each genotype was modelled using linear regression to derive the β (slope) of 0.61, indicating that the AA genotype is associated with 2.33 fold lower expression than GG. The *p*-value, under the null hypothesis that β = zero, was $1.7x10^{-22}$.

(b) CHIP-seq data from [245] show MEIS-1 binding (red) in only 3 reads carrying the A haplotype at rs148436 (eQTL lead SNP, indicated by black dotted line). There was additional binding of MEIS1 in the same reads at 13:95899643C/T.

5.13 Cellular modelling of ABCC4 function in iMK

5.13.1 Patient-derived iPSC generation

iPSC derived from cases A and B were generated as described in section 2.4.1.1. However forward programming was unsuccessful on three occasions.

5.13.2 CRISPR-dCas9 *ABCC4* promoter knockdown

This section describes the results of iMK phenotyping following CRISPR-targeted *ABCC4* promoter inhibition. The materials and methods are described in section 2.5.4. In brief, an iPSC line had been generated by Dr. John Lambourne and Miss Isabel Rosa of the University of Cambridge containing the stably integrated lentiviral Phage TRE dCas9-KRAB vector. The iPSC line (A1ATD-1) that was modified was known to forward program reliably and with high yield of viable $CD41^+$ iMK at day 20. The vector expresses dCas9, a catalytically dead Cas9 protein, resulting from engineered modifications that inactivate the RuvC and HNH nuclease domains [247]. dCas9 can be targeted to a specific DNA sequence by a compatible sgRNA, and binding has been shown to interfere with RNA polymerase binding and transcription factor binding efficiently and reversibly [247]. The dCas9 is fused to a Krüppel-associated box (KRAB) domain. KRAB domains have been shown to further enhance repression of transcription [247] by alteration of H-bonds in DNA, promoting H3K9 methylation (a repressive marker). dCas9-KRAB expression in this vector is under the control of a tetracycline-inducible (tet-on) element [141]. A constitutively expressed reverse tetracycline-controlled transactivator protein (rtTA) activates transcription at a tetracycline-response element (TRE) only when bound to tetracycline antibiotics such as doxycycline, which in turn leads to dCas9-KRAB expression (Figure 14). The dCas9-KRAB iPSC line was transduced with lentiviral particles containing twelve sgRNAs targeting the *ABCC4* promoter. Following puromycin selection for sgRNA-integrated cells, iPSC were forward programmed to iMK.

5.13.3 Generation of test and control conditions

As detailed above, the experiment was originally designed to capitalise on the tet-on system to compare sgRNA-targeted iMK in the presence and absence of doxycycline (+sgRNA/+dox vs. +sgRNA/-dox, respectively).

The +sgRNA/-dox condition did not forward program to iMK, with poor cell recovery following transduction. The +sgRNA/+dox condition and the original dCas9-KRAB line forward programmed well. There was significant ABCC4 knockdown (1,778-fold at iMK) by RTqPCR in the targeted line compared with the original dCas9-KRAB line (Figure 78a).

However analyses from these conditions are not presented in this thesis because the control line was not exposed to puromycin. Also, the original dCas9-KRAB iPSC line showed significant expression of dCas9 without doxycycline addition by RTqPCR and by immunoblotting (Figure 78c, d). This implies unpredictable, leaky expression of dCas9, which precluded further use of the +sgRNA/-dox control for subsequent experiments.

Therefore test conditions were altered to compare sgRNA-targeted iMK with non-targeted iMK that had been transduced with the Lentiguide-Puro empty vector, followed by culture in puromycin (+sgRNA/+dox, hereafter referred to as 'ABCC4 KD' vs. -sgRNA/+dox, hereafter referred to as 'DCAS9-EV,' respectively). These iPSC were forward programmed in three independent experiments. This time there was 5-fold reduced expression by RTqPCR between ABCC4 KD and DCAS9-EV at day 20 of forward programming (Figure 78b). Results were analysed by the paired, two-tailed Student's t-test.

5.13.4 iMK characteristics

There were no differences between ABCC4 KD and DCAS9-EV lines in terms of CD41 or CD42 expression or viability. There was a higher yield of viable $CD41^+$ iMK in the ABCC4-KD condition (mean 5.17 x 10^6, s.d. 3.4 x 10^5) vs. DCAS9-EV (mean 1.77 x 10^6, s.d. 9.6 x10^5, p = 0.01, Figure 79).

By cAMP ELISA there was a non-significant trend towards lower cAMP in the ABCC4-KD condition (mean 220 fmol/10^5 iMK, s.d. 50) vs. DCAS9-EV (mean 241fmol/10^5 iMK, s.d 41, p = 0.08, Figure 80). This is in contrast to findings in platelets, where lower ABCC4 expression was associated with higher intra-platelet cAMP.

A previous study of MK derived from mouse fetal livers showed that inhibition of cAMP-PKA signalling led to decreased MK ploidy and increased platelet release [248]. Cytospins were used to assess ploidy, however there were few polyploid cells (<1/1000) in either condition and there was no difference between conditions (Figure 81a). Proplatelet formation was assessed at 24 hours following plating of iMK on fibrinogen as described in section 2.4.2.2. Proplatelet release was infrequent and again there was no difference between conditions (Figure 81b).

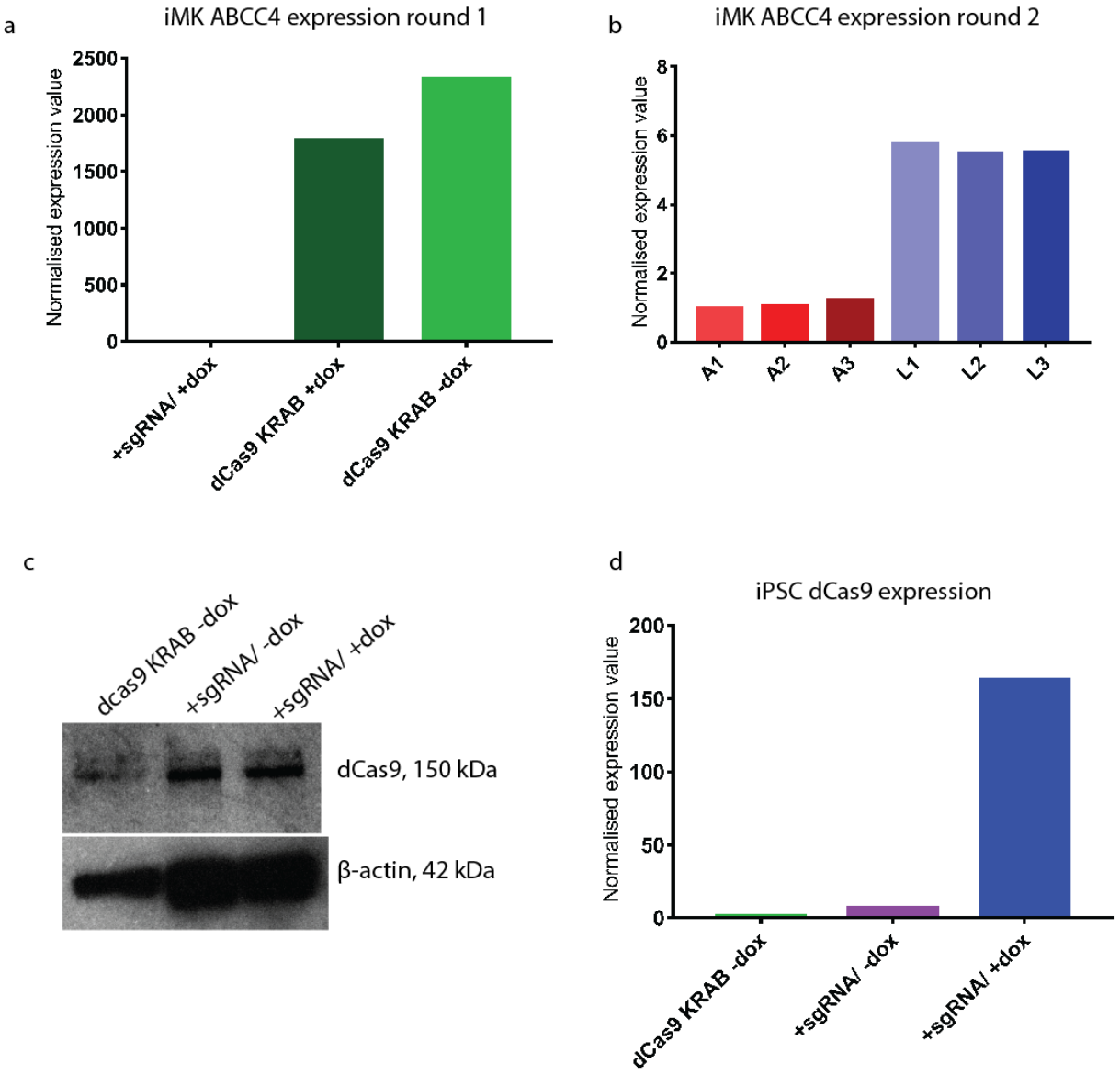


Figure 78 Generation of *ABCC4* knockdown iMK and control

An iPSC line generated by Dr. John Lambourne and Miss Isabel Rosa of the University of Cambridge containing the stably integrated lentiviral Phage TRE dCas9-KRAB vector was transduced with lentiviral particles containing twelve sgRNAs targeting the *ABCC4* promoter. Following puromycin selection for sgRNA-integrated cells, iPSC were forward programmed (FoP) to iMK.

a. Normalised expression values following RTqPCR of the first round of FoP to iMK. There was pronounced *ABCC4* knockdown (1778-fold reduced expression) in the +sgRNA/+dox condition compared with the baseline dCas9-KRAB line. The +sgRNA/–dox control failed to forward program. These cells were not used for further experiments.

b. The second round of iMK showed milder *ABCC4* knockdown (~5x reduced expression) in the +sgRNA/+dox conditions (A1-3) vs. the –sgRNA/+dox control (L1-3). Experiments were performed in triplicate, indicated by suffix 1-3.

c. Immunoblotting shows similar levels of expression of DCAS9 in the baseline DCAS9-KRAB line and sgRNA-transduced lines, with or without addition of doxycycline to the culture. Therefore the tet-on system was considered too leaky for use, and a protocol that compared the +sgRNA condition (ABCC4 KD) and the empty vector conferring puromycin resistance (DCAS9-EV) was chosen. Both conditions were incubated in the same antibiotics and media.

d. Following (c), RTqPCR also confirmed leaky expression of *DCAS9* in conditions with and without doxycycline supplementation, albeit at lower levels in the –dox lines.

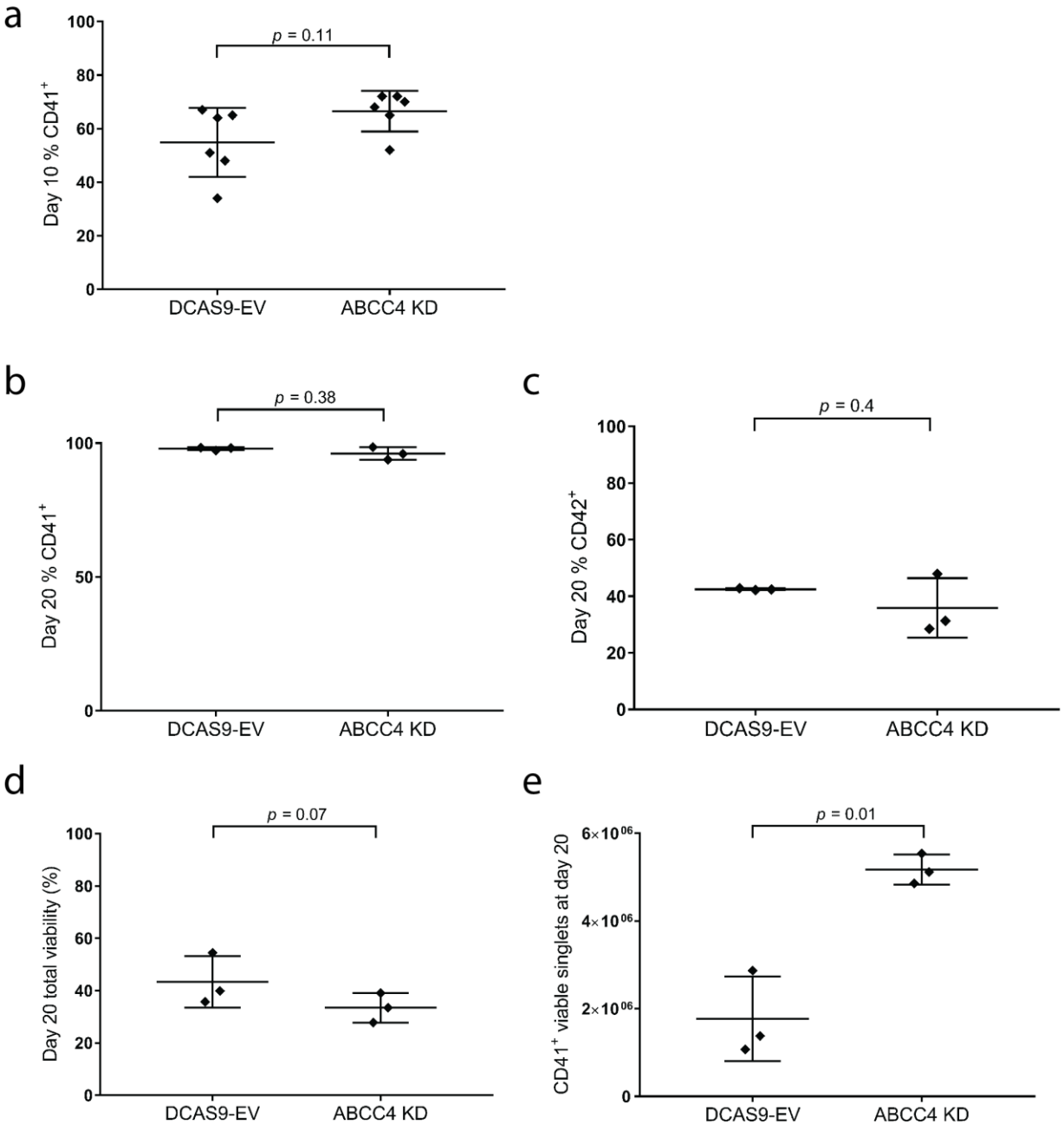


Figure 79 Comparison of iMK maturity and yield.
ABCC4 KD and DCAS9-EV iPSC were FoP to iMK in three experiments. Cells were collected at day 10 and day 20 and stained with antibodies to MK markers CD41 and CD42 and viability marker DAPI prior to flow cytometry. At Day 20 cells were also incubated with count beads to count viable cells. The effects were measured by paired, two-tailed Student's t-tests, $p<0.05$ was considered significant.
(a-d) There were no differences between ABCC4 KD and DCAS9-EV lines in terms of CD41 or CD42 expression, or viability.
(e) There was a higher yield of viable $CD41^+$ iMK in the ABCC4-KD condition (mean 5.17×10^6, s.d. 3.4×10^5) vs. DCAS9-EV (mean 1.77×10^6, s.d. 9.6×10^5, $p = 0.01$ by paired, two-tailed Student's t-test.

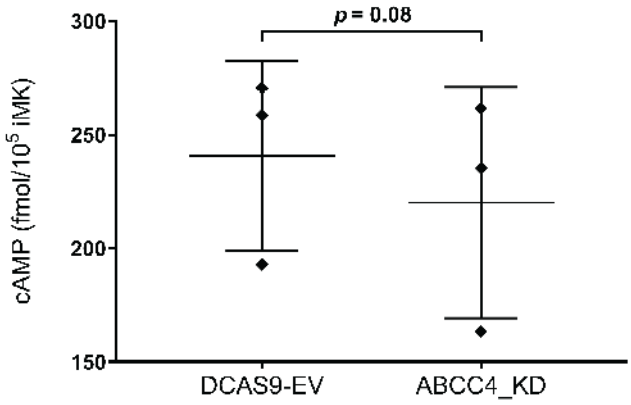


Figure 80 There was a non-significant trend towards lower cAMP in the ABCC4-KD condition by cAMP ELISA of iMK cell lysates.
In platelets lower ABCC4 expression was associated with higher intra-platelet cAMP. The results indicate that either ABCC4 has a different role in iMK, or that both *ABCC4* variants and *P2RY12* variants are required to elevate the cellular cAMP level.

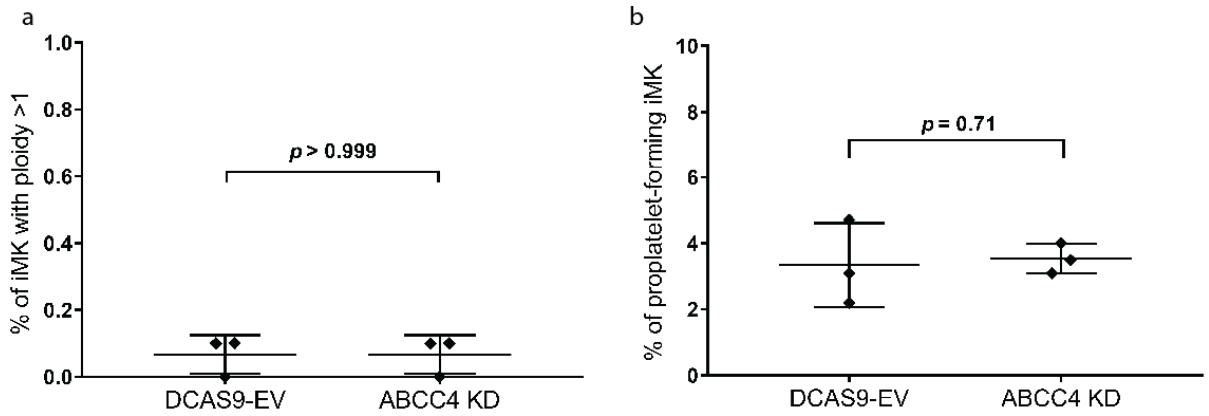


Figure 81 There was no difference between ABCC4 KD and DCAS9-EV iMK in terms of ploidy or proplatelet formation.
There were few polyploid cells and few proplatelet-forming iMK in both conditions which limits the conclusions that can be drawn from these experiments.

5.14 Discussion

In this chapter a pedigree is presented in whom I hypothesized that reduced expression of *ABCC4* is associated with a lifelong mild-to-moderate bleeding disorder. Several observations support *ABCC4* as a candidate BPD gene. These include its ability to transport major platelet signaling molecules cAMP, cGMP and ADP, a strong association with platelet parameters in GWAS, selective, high expression in MK and platelets, and a bleeding phenotype in knockout mice. In this study in collaboration with colleagues at the Universities of Cambridge and Liège, I have shown that *ABCC4* contains a strong platelet eQTL that colocalises with platelet GWAS loci, an MK superenhancer and a MEIS1 binding site. These observations provide further evidence for a role for *ABCC4* in the regulation of thrombopoiesis.

Platelets from cases A and B have increased levels of intracellular cAMP, providing a mechanism by which *ABCC4* variants may inhibit platelet activation to cause clinical bleeding. Plasma cAMP was undetectable in both cases by mass spectrometry; however four of the 56 controls also had undetectable levels, without rare high or moderate impact variants in *ABCC4*, limiting the conclusions from this experiment. Replication of these findings in other individuals with variants in *ABCC4*, who are yet to be identified, would be essential to establish confidence in this association. The intra-platelet cAMP defect is consistent with one of two published mouse *ABCC4* knockout reports, in which intra-platelet cAMP was elevated. However in mice there was also impaired collagen-mediated aggregation secondary to altered GPVI expression at the platelet surface. I have shown no difference in GPVI expression between human cases and controls using a monomeric and dimeric antibody. This discrepancy may represent genuine differences between human and mouse platelets or differences in experimental design; for example, differences between a mouse knockout and otherwise genetically identical controls are less susceptible to noise from variation than between human cases and genetically unrelated controls.

Pedigree members also carry a heterozygous missense *P2RY12* variant that, in keeping with the literature, does not appear to cause increased clinical bleeding or altered baseline levels of cAMP in the absence of *ABCC4* knockout, but does damage the integrity of P2RY12-mediated signalling and contributes to elevation in cAMP during platelet activation. These data suggest that the cause of bleeding in this pedigree is at the least digenic, and may be multi-factorial.

Platelet function phenotyping did not provide support for the hypothesis that *ABCC4* variants contribute to clinically evident platelet dysfunction via raised intracellular cAMP. Reduced collagen-induced platelet aggregation was shown in cases A and B (homozygous *ABCC4*), C (wild-type *ABCC4*) and F (heterozygous *ABCC4*). Iloprost-induced inhibition of aggregation was significantly impaired in all cases compared with a control. Low-dose collagen-induced expression of fibrinogen and p-selectin was also impaired in cases A, B and C (F is yet to be tested). There are several possible explanations for the failure of these phenotypes to co-segregate with the genotype. The *P2RY12* variant may mask the *in vivo* effect of *ABCC4*

loss of function. The observed disturbance of intra-platelet cAMP may not mediate platelet dysfunction. It is also feasible that the *P2RY12* variant is contributing to the clinical phenotype in cases A and B. Whilst the majority of heterozygous *P2RY12* variants are associated with a pure laboratory phenotype without clinical bleeding, dominant negative *P2RY12* variants have been reported [112] and the lack of clinical bleeding in cases C and F could be attributable to other factors. Case C is nulliparous and case F is a young male, therefore it is possible they have not received such significant haemostatic challenges and therefore report less bleeding. Importantly, platelet function tests using low-dose collagen (<1µg/ml) have been performed only once, therefore a third case recall is planned to confirm and expand these findings, including by assessment of platelet VASP-phosphorylation as a downstream surrogate marker of intra-platelet cAMP levels.

To further explore the role of *ABCC4* in platelets and MK, iPSC were derived from cases A and B, but did not forward program to iMK. Technique- and reagent-specific causes of this are unlikely as control lines were successfully reprogrammed. It is not possible to conclude from this failure to reprogram that *ABCC4* or *P2RY12* variants inhibit normal megakaryopoiesis because approximately 25% of lines from healthy donors also consistently fail to form iMK (*personal communication from Dr. C Ghevaert)*. Conversely, lines from donors with significant platelet and MK defects such as Gray Platelet Syndrome (*personal communication from Dr. J Guerrero*) and *KDSR* variants form iMK using standard protocols.

Subsequent cellular studies involved the phenotyping of iMK following CRISPR-targeted *ABCC4* promoter inhibition. An iPSC line (A1ATD-1) known to forward program reliably and with high yield of viable $CD41^{+}$ iMK at day 20 had been stably transduced by colleagues with a lentiviral vector expressing a catalytically dead Cas9 (dCas) protein fused to a Krüppel-associated box (KRAB) domain, under the control of a tet-on system. However significant expression of dCas9 in iMK in the the tet-off condition was observed, precluding the further use of this strategy. Whilst iPSC and iMK culture-media were xeno-free to avoid tetracycline contamination from cattle-farming, it transpired that communal use cytokine stocks for transduction of forward programming were initially made up in low concentrations of bovine serum albumin, therefore unintentional stimulation of the TRE may have occurred. Following a change in experimental protocol to compare +sgRNA/ -sgRNA conditions, reprogramming was repeated, however there was reduction in the level of transcriptional inhibition to only 5-fold. It is likely that silencing of the transgene had occurred, despite maintenance of cells in geneticin in the intervening period. Several groups have demonstrated an attenuation of transgene silencing by targeting transgenes to safe-harbor sites such as the AAVS1 locus in human iPSC, which could be a potential alternative approach for future work [249, 250].

Contrary to our findings in platelets, we showed a non-significant trend towards lower levels of cAMP in *ABCC4* knockdown iMK. The lack of significance may be attributable to the small size of the knockdown observed, the additional effect of the *P2RY12* variant *in vivo* that was not present *in vitro*, or a different

role for ABCC4 and cAMP in MK compared with platelets. Studies of *in vitro* models of megakaryopoiesis suggest that cAMP has a central, specialised role to play in the developing MK. Begonja *et al.* used *in vitro* MK derived from mouse fetal livers to show that intracellular cAMP rose dramatically during early MK differentiation, suggesting that cAMP may have a role to play at this developmental stage [248]. Their study also showed that cAMP inhibition reduced yield of iMK, ploidy and increased platelet release. In this thesis I have shown that *ABCC4* knockdown was associated with significantly higher iMK yield but no difference in ploidy or platelet release. However whilst the original A1ATD-1 line was chosen for its reliability and yield during reprogramming, it is established to be a poor producer of polyploid cells and proplatelets under culture conditions used in this experiment. Therefore the lack of positive events in either analysis suggests that this experiment is poorly designed to detect these differences. Recent alterations in culture conditions made in the Ghevaert laboratory have improved proplatelet formation in this line and could be adopted for experiments going forward.

Further work is required to establish the role of *ABCC4 in vivo* and *in vitro*. The true frequency of the *ABCC4* variants in the Seychelles is unknown and it remains possible that this is a founder mutation that has only caused symptoms due to coinheritance of a *P2RY12* variant that feeds into the same intracellular pathway. Other individuals may be identified as WGS becomes widespread, in whom the role of *ABCC4* in the absence of a *P2RY12* variant can be established. The cellular model described in this chapter also merits further investigation, to assess the transport of other endogenous molecules such as cGMP and ADP, to establish whether cAMP levels are truly contingent on *ABCC4* expression and to investigate downstream pathways by which alterations may affect platelet function and megakaryopoiesis. Avenues of investigation for future work that were insufficiently complete for inclusion in this thesis included the use of PKA-induced VASP-phosphorylation as a downstream surrogate marker of cAMP levels, the expression of GPVI on iMK (the HY101 antibody used to good effect in platelets gave poorly reproducible results in iMK), the use of specific inhibitors of ABCC4 Ceefourin 1 and 2 [230] in stock iMK lines and lentiviral expression of a tagged ABCC4 ORF to further explore cellular location in iMK and platelet-like particles. Several studies have shown that cAMP may influence calcium flux in the developing MK, providing a further potential area for investigation [251-253].

Chapter 6: Conclusions and future work

6 Conclusions and future work

The NIHR-BioResource Rare Diseases collaboration has resulted in the detection and feedback of new genetic diagnoses that have improved the detection and management of individuals with BPD and has provided new insights into normal platelet development and physiology. The non-coding space remains to a large extent a land unknown and work is ongoing to identify variants in known and novel BPD genes using innovative approaches in statistical genomics that interrogate this part of the genome.

This thesis reports on the role of KDSR in haematopoiesis. I have described novel clinical phenotypes, a characteristic metabolic signature and the discovery of *in vitro* phenotypes related to growth and apoptosis in iPSC, and yield, viability, maturity, size and proplatelet formation in iMK. Impaired proplatelet formation was replicated in $CD34^{+}$-derived MK and zebrafish morpholino studies, providing further support for an integral role of KDSR in thrombopoiesis.

Further work is required to confirm these findings, both to elaborate the mechanism of disease and enhance our understanding of the ubiquitous process of sphingolipid synthesis, which is essential for life. The confirmation of the platelet, myelofibrotic, metabolic and proplatelet phenotypes in the sister of the proband, who is too young for these analyses at the time of writing, and in other affected cases would provide significant support for the findings reported here. Multi-centre collaboration will be necessary to achieve this. The mechanism of the myelofibrotic phenotype has not been explored in this thesis and a mouse model would be of great benefit to this end. A tissue-specific or inducible knockout may be preferred as embryonic lethality is likely following a global *KDSR* knockout, in view of its integral role in *de novo* sphingolipid synthesis [254].

The use of alternative strategies to assess apoptosis and growth in the iPSC model described, such as activated caspase or cytochrome C detection by immunoblotting, glutathione assays or mitochondrial membrane potential dyes, would provide useful data. Repetition of forward programming on a larger scale to yield enough iMK to test growth and apoptosis phenotypes and confirm the PLP phenotype by a dedicated assay using a viability stain and multiple platelet-specific markers would also supply valuable information. However, the most convincing evidence that the *in vitro* phenotypes observed in this thesis are true would be replication in the presence of other reported pathogenic *KDSR* variants. This could be achieved by repetition of these studies using iPSC derived from an alternative affected pedigree, using site-directed mutagenesis of the lentiviral vector containing the wild-type open reading frame of *KDSR*, or by CRISPR-Cas9 knock-in gene-editing.

Assessment of the impact of individual sphingolipid moieties on phenotype may help clarify the specific mechanism by which *KDSR* variants act, for example by using spiking assays to imitate the metabolic profile observed in phenotypically normal iPSC and iMK during forward programming. These studies may also provide new information about feedback loops in the early stages of *de novo* sphingolipid synthesis, as may assessment of post-translational modifications of key enzymes in the pathway by

immunoblotting. The gene-lists derived from RNA-seq and GO term enrichment analysis also require further exploration *in silico* and *in vitro*, to further assess how *KDSR* variants influence growth, differentiation and the cytoskeleton.

This thesis also reports on the study of a novel candidate BPD gene *ABCC4*. *ABCC4* is shown to contain a strong platelet eQTL that colocalises with platelet GWAS loci, an MK superenhancer and a MEIS-1 binding site, supporting a role for ABCC4 in the regulation of thrombopoiesis. I have described the first pedigree to my knowledge in which *ABCC4* variants may be associated with human pathology. Coinheritance of homozygous loss of function variants in *ABCC4* and heterozygous variants in known BPD gene *P2RY12* cosegregate with a lifelong mild-to-moderate bleeding disorder. cAMP was dysregulated in the platelets of affected individuals. However, no downstream platelet function defect was identified that cosegregated with the bleeding phenotype, therefore it remains to be seen whether cAMP dysregulation associated with *ABCC4* variants truly alters platelet function to cause clinical bleeding. Future work should include the examination of the effect on downstream intermediates in the cAMP pathway in platelets, such as PKA-induced VASP-phosphorylation, to evaluate the impact of high cAMP. The use of specific inhibitors of ABCC4, Ceefourin 1 and 2 [230], in platelets during platelet function testing and prior to cAMP measurement could also provide useful data, allowing assessment of dose-response relationships.

CRISPR-Cas9-mediated inhibition of *ABCC4* expression was not associated with significant cAMP dysregulation in iMK in this study. This may be attributable to the small size of the knockdown observed or due to the additional effect of the *P2RY12* variant *in vivo*. The latter explanation illustrates the challenges of investigating a multigenic disorder. Replication of these experiments is warranted, ideally using several different sets of sgRNAs and achieving differing levels of knockdown to assess a dose-response relationship. Furthermore, recent alterations to the forward programming protocol, including the use of different culture media, have been shown to improve proplatelet formation and ploidy in the A1ATD-1 line used for this study (*personal communication from Dr C. Ghevaert*). This could improve the sensitivity of the proplatelet and ploidy experiments performed, and allow the determination of cAMP concentration in iMK-derived PLP to provide an improved *in vitro* model of ABCC4 function.

High throughput sequencing is expanding, producing seemingly endless data. Careful filtering, based on hypotheses about the nature of novel genes is essential to guide time-consuming, expensive, labour-intensive but necessary functional studies. Such filtering is not without hazard; variant prioritisation in the NIHR-BR RD project has been based on the presumption that similar phenotypes are associated with variants in the same gene, but I have shown that *KDSR* variants are associated with significant clinical pleiotropy. Variant discovery algorithms were primarily designed to detect monogenic disorders, though this thesis reports on a pedigree in whom bleeding is likely to be digenic and the associated investigative challenges. As a final word, it is my hope that this thesis has illustrated the benefits of collaborative

working and data-sharing, for the recruitment of cases with rare diseases, for variant prioritisation and for the experimental validation of hypothetical, statistical associations.

7 References

1. Quiroga, T., et al., *High prevalence of bleeders of unknown cause among patients with inherited mucocutaneous bleeding. A prospective study of 280 patients and 299 controls.* Haematologica, 2007. **92**(3): p. 357-65.
2. Hayward, C.P., et al., *Diagnostic utility of light transmission platelet aggregometry: results from a prospective study of individuals referred for bleeding disorder assessments.* J Thromb Haemost, 2009. **7**(4): p. 676-84.
3. Takeichi, T., et al., *Biallelic mutations in KDSR disrupt ceramide synthesis and result in a spectrum of keratinization disorders associated with thrombocytopenia.* J Invest Dermatol, 2017.
4. Boyden, L.M., et al., *Mutations in KDSR Cause Recessive Progressive Symmetric Erythrokeratoderma.* Am J Hum Genet, 2017. **100**(6): p. 978-984.
5. Turro, E., et al., *A dominant gain-of-function mutation in universal tyrosine kinase SRC causes thrombocytopenia, myelofibrosis, bleeding, and bone pathologies.* Sci Transl Med, 2016. **8**(328): p. 328ra30.
6. Stritt, S., et al., *A gain-of-function variant in DIAPH1 causes dominant macrothrombocytopenia and hearing loss.* Blood, 2016.
7. Pleines, I., et al., *Mutations in tropomyosin 4 underlie a rare form of human macrothrombocytopenia.* J Clin Invest, 2017. **127**(3): p. 814-829.
8. Sivapalaratnam, S., et al., *Rare variants in GP1BB are responsible for autosomal dominant macrothrombocytopenia.* Blood, 2017. **129**(4): p. 520-524.
9. Westbury, S.K., et al., *Expanded repertoire of RASGRP2 variants responsible for platelet dysfunction and severe bleeding.* Blood, 2017. **130**(8): p. 1026-1030.
10. Poggi, M., et al., *Germline variants in ETV6 underlie reduced platelet formation, platelet dysfunction and increased levels of circulating CD34+ progenitors.* Haematologica, 2017. **102**(2): p. 282-294.
11. Franco, A.T., A. Corken, and J. Ware, *Platelets at the interface of thrombosis, inflammation, and cancer.* Blood, 2015. **126**(5): p. 582-8.
12. Bolton-Maggs, P.H., et al., *A review of inherited platelet disorders with guidelines for their management on behalf of the UKHCDO.* Br J Haematol, 2006. **135**(5): p. 603-33.
13. Kaushansky, K., *The molecular mechanisms that control thrombopoiesis.* J Clin Invest, 2005. **115**(12): p. 3339-47.
14. Nakeff, A. and B. Maat, *Separation of megakaryocytes from mouse bone marrow by velocity sedimentation.* Blood, 1974. **43**(4): p. 591-5.
15. 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.
16. Woolthuis, C.M. and C.Y. Park, *Hematopoietic stem/progenitor cell commitment to the megakaryocyte lineage.* Blood, 2016. **127**(10): p. 1242-8.
17. Pang, W.W., et al., *Human bone marrow hematopoietic stem cells are increased in frequency and myeloid-biased with age.* Proc Natl Acad Sci U S A, 2011. **108**(50): p. 20012-7.
18. Rodriguez-Fraticelli, A.E., et al., *Clonal analysis of lineage fate in native haematopoiesis.* Nature, 2018. **553**(7687): p. 212-216.
19. Tijssen, M.R. and C. Ghevaert, *Transcription factors in late megakaryopoiesis and related platelet disorders.* J Thromb Haemost, 2013. **11**(4): p. 593-604.
20. Tiwari, S., et al., *A role for Rab27b in NF-E2-dependent pathways of platelet formation.* Blood, 2003. **102**(12): p. 3970-9.
21. Zeddies, S., et al., *MEIS1 regulates early erythroid and megakaryocytic cell fate.* Haematologica, 2014. **99**(10): p. 1555-64.
22. 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.
23. Richardson, J.L., et al., *Mechanisms of organelle transport and capture along proplatelets during platelet production.* Blood, 2005. **106**(13): p. 4066-75.
24. Seri, M., et al., *Mutations in MYH9 result in the May-Hegglin anomaly, and Fechtner and Sebastian syndromes. The May-Hegglin/Fechtner Syndrome Consortium.* Nat Genet, 2000. **26**(1): p. 103-5.
25. Kunishima, S., et al., *ACTN1 mutations cause congenital macrothrombocytopenia.* Am J Hum Genet, 2013. **92**(3): p. 431-8.
26. Nurden, P., et al., *Thrombocytopenia resulting from mutations in filamin A can be expressed as an isolated syndrome.* Blood, 2011. **118**(22): p. 5928-37.
27. Kunishima, S., et al., *Mutation of the beta1-tubulin gene associated with congenital macrothrombocytopenia affecting microtubule assembly.* Blood, 2009. **113**(2): p. 458-61.
28. Derry, J.M., H.D. Ochs, and U. Francke, *Isolation of a novel gene mutated in Wiskott-Aldrich syndrome.* Cell, 1994. **78**(4): p. 635-44.

29. 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.
30. 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.
31. 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.
32. Matsunaga, T., et al., *Potentiated activation of VLA-4 and VLA-5 accelerates proplatelet-like formation.* Ann Hematol, 2012. **91**(10): p. 1633-43.
33. Schachtner, H., et al., *Megakaryocytes assemble podosomes that degrade matrix and protrude through basement membrane.* Blood, 2013. **121**(13): p. 2542-52.
34. 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.
35. 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.
36. Junt, T., et al., *Dynamic visualization of thrombopoiesis within bone marrow.* Science, 2007. **317**(5845): p. 1767-70.
37. Schwertz, H., et al., *Anucleate platelets generate progeny.* Blood, 2010. **115**(18): p. 3801-9.
38. Thon, J.N., et al., *Cytoskeletal mechanics of proplatelet maturation and platelet release.* J Cell Biol, 2010. **191**(4): p. 861-74.
39. Bariana, T.K., et al., *Dawning of the age of genomics for platelet granule disorders: improving insight, diagnosis and management.* Br J Haematol, 2017. **176**(5): p. 705-720.
40. Thon, J.N. and J.E. Italiano, *Platelets: production, morphology and ultrastructure.* Handb Exp Pharmacol, 2012(210): p. 3-22.
41. Kamykowski, J., et al., *Quantitative immunofluorescence mapping reveals little functional coclustering of proteins within platelet alpha-granules.* Blood, 2011. **118**(5): p. 1370-3.
42. Jonnalagadda, D., L.T. Izu, and S.W. Whiteheart, *Platelet secretion is kinetically heterogeneous in an agonist-responsive manner.* Blood, 2012. **120**(26): p. 5209-16.
43. Min, S.H., et al., *Loss of PIKfyve in platelets causes a lysosomal disease leading to inflammation and thrombosis in mice.* Nat Commun, 2014. **5**: p. 4691.
44. Watson, S.P., *Platelet activation by extracellular matrix proteins in haemostasis and thrombosis.* Curr Pharm Des, 2009. **15**(12): p. 1358-72.
45. Dutta-Roy, A.K. and A.K. Sinha, *Purification and properties of prostaglandin E1/prostacyclin receptor of human blood platelets.* J Biol Chem, 1987. **262**(26): p. 12685-91.
46. Johnston-Cox, H.A., D. Yang, and K. Ravid, *Physiological implications of adenosine receptor-mediated platelet aggregation.* J Cell Physiol, 2011. **226**(1): p. 46-51.
47. Freson, K., et al., *PACAP and its receptor VPAC1 regulate megakaryocyte maturation: therapeutic implications.* Blood, 2008. **111**(4): p. 1885-93.
48. Dittrich, M., et al., *Platelet protein interactions: map, signaling components, and phosphorylation groundstate.* Arterioscler Thromb Vasc Biol, 2008. **28**(7): p. 1326-31.
49. Noe, L., et al., *Regulators of platelet cAMP levels: clinical and therapeutic implications.* Curr Med Chem, 2010. **17**(26): p. 2897-905.
50. Smolenski, A., *Novel roles of cAMP/cGMP-dependent signaling in platelets.* J Thromb Haemost, 2012. **10**(2): p. 167-76.
51. El-Daher, S.S., et al., *Distinct localization and function of (1,4,5)IP(3) receptor subtypes and the (1,3,4,5)IP(4) receptor GAP1(IP4BP) in highly purified human platelet membranes.* Blood, 2000. **95**(11): p. 3412-22.
52. Hoffmeister, M., et al., *Cyclic nucleotide-dependent protein kinases inhibit binding of 14-3-3 to the GTPase-activating protein Rap1GAP2 in platelets.* J Biol Chem, 2008. **283**(4): p. 2297-306.
53. Bodnar, R.J., et al., *Regulation of glycoprotein Ib-IX-von Willebrand factor interaction by cAMP-dependent protein kinase-mediated phosphorylation at Ser 166 of glycoprotein Ib(beta).* J Biol Chem, 2002. **277**(49): p. 47080-7.
54. Butt, E., et al., *cAMP- and cGMP-dependent protein kinase phosphorylation sites of the focal adhesion vasodilator-stimulated phosphoprotein (VASP) in vitro and in intact human platelets.* J Biol Chem, 1994. **269**(20): p. 14509-17.
55. Butt, E., et al., *Actin binding of human LIM and SH3 protein is regulated by cGMP- and cAMP-dependent protein kinase phosphorylation on serine 146.* J Biol Chem, 2003. **278**(18): p. 15601-7.
56. Watson, S.P., R.T. McConnell, and E.G. Lapetina, *The rapid formation of inositol phosphates in human platelets by thrombin is inhibited by prostacyclin.* J Biol Chem, 1984. **259**(21): p. 13199-203.

57. Zhang, W. and R.W. Colman, *Thrombin regulates intracellular cyclic AMP concentration in human platelets through phosphorylation/activation of phosphodiesterase 3A.* Blood, 2007. **110**(5): p. 1475-82.
58. Noe, L., et al., *Platelet Gs hypofunction and abnormal morphology resulting from a heterozygous RGS2 mutation.* J Thromb Haemost, 2010. **8**(7): p. 1594-603.
59. Gresele, P., et al., *Diagnosis of suspected inherited platelet function disorders: results of a worldwide survey.* J Thromb Haemost, 2014. **12**(9): p. 1562-9.
60. UKHCDO, *UK National Haemophilia Database Bleeding Disorder Statistics for 2011-2012.* 2012.
61. Cines, D.B., et al., *Congenital and acquired thrombocytopenia.* Hematology Am Soc Hematol Educ Program, 2004: p. 390-406.
62. Quiroga, T., et al., *Template bleeding time and PFA-100 have low sensitivity to screen patients with hereditary mucocutaneous hemorrhages: comparative study in 148 patients.* J Thromb Haemost, 2004. **2**(6): p. 892-8.
63. Cattaneo, M., et al., *Results of a worldwide survey on the assessment of platelet function by light transmission aggregometry: a report from the platelet physiology subcommittee of the SSC of the ISTH.* J Thromb Haemost, 2009. **7**(6): p. 1029.
64. Cattaneo, M., et al., *Recommendations for the Standardization of Light Transmission Aggregometry: A Consensus of the Working Party from the Platelet Physiology Subcommittee of SSC/ISTH.* J Thromb Haemost, 2013.
65. Gresele, P., *Diagnosis of inherited platelet function disorders: guidance from the SSC of the ISTH.* J Thromb Haemost, 2015. **13**(2): p. 314-22.
66. Simeoni, I., et al., *A comprehensive high-throughput sequencing test for the diagnosis of inherited bleeding, thrombotic and platelet disorders.* Blood, 2016.
67. Bastida, J.M., et al., *Design and application of a 23-gene panel by next-generation sequencing for inherited coagulation bleeding disorders.* Haemophilia, 2016. **22**(4): p. 590-7.
68. Langmead, B., et al., *Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.* Genome Biol, 2009. **10**(3): p. R25.
69. Turro, E., et al., *Haplotype and isoform specific expression estimation using multi-mapping RNA-seq reads.* Genome Biol, 2011. **12**(2): p. R13.
70. Turro, E., W.J. Astle, and S. Tavare, *Flexible analysis of RNA-seq data using mixed effects models.* Bioinformatics, 2014. **30**(2): p. 180-8.
71. Lentaigne, C., et al., *Inherited platelet disorders: toward DNA-based diagnosis.* Blood, 2016. **127**(23): p. 2814-23.
72. Westbury, S.K., et al., *Human phenotype ontology annotation and cluster analysis to unravel genetic defects in 707 cases with unexplained bleeding and platelet disorders.* Genome Med, 2015. **7**(1): p. 36.
73. Peplow, M., *The 100 000 Genomes Project.* BMJ, 2016. **353**: p. i1757.
74. Rath, A., et al., *Representation of rare diseases in health information systems: the Orphanet approach to serve a wide range of end users.* Hum Mutat, 2012. **33**(5): p. 803-8.
75. Robinson, P.N. and S. Mundlos, *The human phenotype ontology.* Clin Genet, 2010. **77**(6): p. 525-34.
76. Mauer, A.C., et al., *Impact of sex, age, race, ethnicity and aspirin use on bleeding symptoms in healthy adults.* J Thromb Haemost, 2011. **9**(1): p. 100-8.
77. Rodeghiero, F., et al., *The discriminant power of bleeding history for the diagnosis of type 1 von Willebrand disease: an international, multicenter study.* J Thromb Haemost, 2005. **3**(12): p. 2619-26.
78. Tosetto, A., et al., *A quantitative analysis of bleeding symptoms in type 1 von Willebrand disease: results from a multicenter European study (MCMDM-1 VWD).* J Thromb Haemost, 2006. **4**(4): p. 766-73.
79. Bowman, M., et al., *Generation and validation of the Condensed MCMDM-1VWD Bleeding Questionnaire for von Willebrand disease.* J Thromb Haemost, 2008. **6**(12): p. 2062-6.
80. Bowman, M., et al., *Evaluation of the diagnostic utility for von Willebrand disease of a pediatric bleeding questionnaire.* J Thromb Haemost, 2009. **7**(8): p. 1418-21.
81. Marcus, P.D., et al., *The power of a standardized bleeding score in diagnosing paediatric type 1 von Willebrand's disease and platelet function defects.* Haemophilia, 2011. **17**(2): p. 223-7.
82. Azzam, H.A., et al., *The condensed MCMDM-1 VWD bleeding questionnaire as a predictor of bleeding disorders in women with unexplained menorrhagia.* Blood Coagul Fibrinolysis, 2012. **23**(4): p. 311-5.
83. Tosetto, A., et al., *Prospective evaluation of the clinical utility of quantitative bleeding severity assessment in patients referred for hemostatic evaluation.* J Thromb Haemost, 2011. **9**(6): p. 1143-8.
84. Lowe, G.C., M. Lordkipanidze, and S.P. Watson, *Utility of the ISTH Bleeding Assessment Tool in predicting platelet defects in participants with suspected inherited platelet function disorders.* J Thromb Haemost, 2013.
85. Gunay-Aygun, M., et al., *Gray platelet syndrome: natural history of a large patient cohort and locus assignment to chromosome 3p.* Blood, 2010. **116**(23): p. 4990-5001.

86. Petersen, B.S., et al., *Opportunities and challenges of whole-genome and -exome sequencing.* BMC Genet, 2017. **18**(1): p. 14.
87. Albers, C.A., et al., *Exome sequencing identifies NBEAL2 as the causative gene for gray platelet syndrome.* Nat Genet, 2011. **43**(8): p. 735-7.
88. Albers, C.A., et al., *Compound inheritance of a low-frequency regulatory SNP and a rare null mutation in exon-junction complex subunit RBM8A causes TAR syndrome.* Nat Genet, 2012. **44**(4): p. 435-9, S1-2.
89. Klopocki, E., et al., *Complex inheritance pattern resembling autosomal recessive inheritance involving a microdeletion in thrombocytopenia-absent radius syndrome.* Am J Hum Genet, 2007. **80**(2): p. 232-40.
90. Robinson, P.N., P. Krawitz, and S. Mundlos, *Strategies for exome and genome sequence data analysis in disease-gene discovery projects.* Clin Genet, 2011. **80**(2): p. 127-32.
91. Ecker, S., et al., *Genome-wide analysis of differential transcriptional and epigenetic variability across human immune cell types.* Genome Biol, 2017. **18**(1): p. 18.
92. Petersen, R., et al., *Platelet function is modified by common sequence variation in megakaryocyte super enhancers.* Nat Commun, 2017. **8**: p. 16058.
93. Belkadi, A., et al., *Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants.* Proc Natl Acad Sci U S A, 2015. **112**(17): p. 5473-8.
94. Barseghyan, H., et al., *Next-generation mapping: a novel approach for detection of pathogenic structural variants with a potential utility in clinical diagnosis.* Genome Med, 2017. **9**(1): p. 90.
95. Lek, M., et al., *Analysis of protein-coding genetic variation in 60,706 humans.* Nature, 2016. **536**(7616): p. 285-91.
96. Walter, K., et al., *The UK10K project identifies rare variants in health and disease.* Nature, 2015. **526**(7571): p. 82-90.
97. Auton, A., et al., *A global reference for human genetic variation.* Nature, 2015. **526**(7571): p. 68-74.
98. Sudlow, C., et al., *UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age.* PLoS Med, 2015. **12**(3): p. e1001779.
99. Gurdasani, D., et al., *The African Genome Variation Project shapes medical genetics in Africa.* Nature, 2015. **517**(7534): p. 327-32.
100. Narasimhan, V.M., et al., *Health and population effects of rare gene knockouts in adult humans with related parents.* Science, 2016. **352**(6284): p. 474-7.
101. Kimchi-Sarfaty, C., et al., *A "silent" polymorphism in the MDR1 gene changes substrate specificity.* Science, 2007. **315**(5811): p. 525-8.
102. Kircher, M., et al., *A general framework for estimating the relative pathogenicity of human genetic variants.* Nat Genet, 2014. **46**(3): p. 310-5.
103. Chen, L., et al., *Transcriptional diversity during lineage commitment of human blood progenitors.* Science, 2014. **345**(6204): p. 1251033.
104. Noris, P., et al., *Mutations in ANKRD26 are responsible for a frequent form of inherited thrombocytopenia: analysis of 78 patients from 21 families.* Blood, 2011. **117**(24): p. 6673-80.
105. Kundaje, A., et al., *Integrative analysis of 111 reference human epigenomes.* Nature, 2015. **518**(7539): p. 317-30.
106. Encode_Project_Consortium, *An integrated encyclopedia of DNA elements in the human genome.* Nature, 2012. **489**(7414): p. 57-74.
107. 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.
108. Gieger, C., et al., *New gene functions in megakaryopoiesis and platelet formation.* Nature, 2011. **480**(7376): p. 201-8.
109. van der Harst, P., et al., *Seventy-five genetic loci influencing the human red blood cell.* Nature, 2012. **492**(7429): p. 369-75.
110. Moore, C., et al., *The INTERVAL trial to determine whether intervals between blood donations can be safely and acceptably decreased to optimise blood supply: study protocol for a randomised controlled trial.* Trials, 2014. **15**: p. 363.
111. 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.
112. Takahashi, K., et al., *Induction of pluripotent stem cells from adult human fibroblasts by defined factors.* Cell, 2007. **131**(5): p. 861-72.
113. Cox, J.L. and A. Rizzino, *Induced pluripotent stem cells: what lies beyond the paradigm shift.* Exp Biol Med (Maywood), 2010. **235**(2): p. 148-58.
114. Lister, R., et al., *Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells.* Nature, 2011. **471**(7336): p. 68-73.
115. Kilpinen, H., et al., *Common genetic variation drives molecular heterogeneity in human iPSCs.* Nature, 2017. **546**(7658): p. 370-375.

116. Hanna, J.H., K. Saha, and R. Jaenisch, *Pluripotency and cellular reprogramming: facts, hypotheses, unresolved issues.* Cell, 2010. **143**(4): p. 508-25.
117. Hussein, S.M., et al., *Copy number variation and selection during reprogramming to pluripotency.* Nature, 2011. **471**(7336): p. 58-62.
118. Gore, A., et al., *Somatic coding mutations in human induced pluripotent stem cells.* Nature, 2011. **471**(7336): p. 63-7.
119. Vining, K.H. and D.J. Mooney, *Mechanical forces direct stem cell behaviour in development and regeneration.* Nat Rev Mol Cell Biol, 2017. **18**(12): p. 728-742.
120. Mills, J.A., et al., *Hematopoietic differentiation of pluripotent stem cells in culture.* Methods Mol Biol, 2014. **1185**: p. 181-94.
121. Vodyanik, M.A., et al., *Human embryonic stem cell-derived CD34+ cells: efficient production in the coculture with OP9 stromal cells and analysis of lymphohematopoietic potential.* Blood, 2005. **105**(2): p. 617-26.
122. Borst, S., et al., *Induced Pluripotent Stem Cell-Derived Megakaryocytes and Platelets for Disease Modeling and Future Clinical Applications.* Arterioscler Thromb Vasc Biol, 2017. **37**(11): p. 2007-2013.
123. 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.
124. Horvath, P. and R. Barrangou, *CRISPR/Cas, the immune system of bacteria and archaea.* Science, 2010. **327**(5962): p. 167-70.
125. Jinek, M., et al., *A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.* Science, 2012. **337**(6096): p. 816-21.
126. Ran, F.A., et al., *Genome engineering using the CRISPR-Cas9 system.* Nat Protoc, 2013. **8**(11): p. 2281-308.
127. Ran, F.A., et al., *Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity.* Cell, 2013. **154**(6): p. 1380-9.
128. Harrison, P., et al., *Guidelines for the laboratory investigation of heritable disorders of platelet function.* Br J Haematol, 2011. **155**(1): p. 30-44.
129. Greene, D., S. Richardson, and E. Turro, *A Fast Association Test for Identifying Pathogenic Variants Involved in Rare Diseases.* Am J Hum Genet, 2017. **101**(1): p. 104-114.
130. Berman, H.M., et al., *The Protein Data Bank.* Nucleic Acids Res, 2000. **28**(1): p. 235-42.
131. Smith, C.L. and J.T. Eppig, *The mammalian phenotype ontology: enabling robust annotation and comparative analysis.* Wiley Interdiscip Rev Syst Biol Med, 2009. **1**(3): p. 390-9.
132. Jupe, S., et al., *Reactome - a curated knowledgebase of biological pathways: megakaryocytes and platelets.* J Thromb Haemost, 2012. **10**(11): p. 2399-402.
133. Yu, J., et al., *Induced pluripotent stem cell lines derived from human somatic cells.* Science, 2007. **318**(5858): p. 1917-20.
134. Schindelin, J., et al., *Fiji: an open-source platform for biological-image analysis.* Nat Methods, 2012. **9**(7): p. 676-82.
135. 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.
136. Zhou, Y., et al., *High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells.* Nature, 2014. **509**(7501): p. 487-91.
137. Bjerrum, J.T., et al., *Integration of transcriptomics and metabonomics: improving diagnostics, biomarker identification and phenotyping in ulcerative colitis.* Metabolomics, 2014. **10**(2): p. 280-290.
138. Sanjana, N.E., O. Shalem, and F. Zhang, *Improved vectors and genome-wide libraries for CRISPR screening.* Nat Methods, 2014. **11**(8): p. 783-784.
139. Shalem, O., et al., *Genome-scale CRISPR-Cas9 knockout screening in human cells.* Science, 2014. **343**(6166): p. 84-87.
140. Guschin, D.Y., et al., *A rapid and general assay for monitoring endogenous gene modification.* Methods Mol Biol, 2010. **649**: p. 247-56.
141. Gossen, M. and H. Bujard, *Tight control of gene expression in mammalian cells by tetracycline-responsive promoters.* Proc Natl Acad Sci U S A, 1992. **89**(12): p. 5547-51.
142. Matsui, H., et al., *Delivery of full-length factor VIII using a piggyBac transposon vector to correct a mouse model of hemophilia A.* PLoS One, 2014. **9**(8): p. e104957.
143. Long, T., et al., *Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites.* Nat Genet, 2017. **49**(4): p. 568-578.
144. Koulman, A., et al., *High-resolution extracted ion chromatography, a new tool for metabolomics and lipidomics using a second-generation orbitrap mass spectrometer.* Rapid Commun Mass Spectrom, 2009. **23**(10): p. 1411-8.

145. Lu, L., et al., *An Unbiased Lipidomics Approach Identifies Early Second Trimester Lipids Predictive of Maternal Glycemic Traits and Gestational Diabetes Mellitus.* Diabetes Care, 2016. **39**(12): p. 2232-2239.
146. Lin, H.F., et al., *Analysis of thrombocyte development in CD41-GFP transgenic zebrafish.* Blood, 2005. **106**(12): p. 3803-10.
147. Louwette, S., et al., *Regulator of G-protein signaling 18 controls megakaryopoiesis and the cilia-mediated vertebrate mechanosensory system.* FASEB J, 2012. **26**(5): p. 2125-36.
148. Louwette, S., et al., *NPC1 defect results in abnormal platelet formation and function: studies in Niemann-Pick disease type C1 patients and zebrafish.* Hum Mol Genet, 2013. **22**(1): p. 61-73.
149. Adams, D., et al., *BLUEPRINT to decode the epigenetic signature written in blood.* Nat Biotechnol, 2012. **30**(3): p. 224-6.
150. Matthijs G, A.M., Bauer P, Corveleyn A, Eck S, Feenstra I, Race V, Scheffer H, Sistermans E, Souche E, Sturm M, Weiss M, Yntema H, *EuroGentest guidelines for diagnostic next generation sequencing.* 2014.
151. Hirata, T., et al., *Arg60 to Leu mutation of the human thromboxane A2 receptor in a dominantly inherited bleeding disorder.* J Clin Invest, 1994. **94**(4): p. 1662-7.
152. Mumford, A.D., et al., *A novel thromboxane A2 receptor D304N variant that abrogates ligand binding in a patient with a bleeding diathesis.* Blood, 2010. **115**(2): p. 363-9.
153. Nisar, S.P., et al., *A novel thromboxane A2 receptor N42S variant results in reduced surface expression and platelet dysfunction.* Thromb Haemost, 2014. **111**(5): p. 923-32.
154. Mumford, A.D., et al., *Platelet dysfunction associated with the novel Trp29Cys thromboxane A(2) receptor variant.* J Thromb Haemost, 2013. **11**(3): p. 547-54.
155. Kirschner, L.S., et al., *Mutations of the gene encoding the protein kinase A type I-alpha regulatory subunit in patients with the Carney complex.* Nat Genet, 2000. **26**(1): p. 89-92.
156. Mayer, L., et al., *The BEACH-domain containing protein, Nbeal2, interacts with Dock7, Sec16a and Vac14.* Blood, 2017.
157. Simeoni, I., et al., *A high-throughput sequencing test for diagnosing inherited bleeding, thrombotic, and platelet disorders.* Blood, 2016. **127**(23): p. 2791-803.
158. Robinson, P.N., et al., *The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease.* Am J Hum Genet, 2008. **83**(5): p. 610-5.
159. Gueguen, P., et al., *A missense mutation in the alpha-actinin 1 gene (ACTN1) is the cause of autosomal dominant macrothrombocytopenia in a large French family.* PLoS One, 2013. **8**(9): p. e74728.
160. Tosetto, A., et al., *A comparison between two semi-quantitative bleeding scales for the diagnosis and assessment of bleeding severity in type 1 von Willebrand disease.* Haemophilia, 2011. **17**(1): p. 165-6.
161. Kihara, A. and Y. Igarashi, *FVT-1 is a mammalian 3-ketodihydrosphingosine reductase with an active site that faces the cytosolic side of the endoplasmic reticulum membrane.* J Biol Chem, 2004. **279**(47): p. 49243-50.
162. Beeler, T., et al., *The Saccharomyces cerevisiae TSC10/YBR265w gene encoding 3-ketosphinganine reductase is identified in a screen for temperature-sensitive suppressors of the Ca2+-sensitive csg2Delta mutant.* J Biol Chem, 1998. **273**(46): p. 30688-94.
163. Sims, K.J., et al., *Yeast sphingolipid metabolism: clues and connections.* Biochem Cell Biol, 2004. **82**(1): p. 45-61.
164. Hannun, Y.A., C. Luberto, and K.M. Argraves, *Enzymes of sphingolipid metabolism: from modular to integrative signaling.* Biochemistry, 2001. **40**(16): p. 4893-903.
165. Munzer, P., et al., *Sphingosine kinase 1 (Sphk1) negatively regulates platelet activation and thrombus formation.* Am J Physiol Cell Physiol, 2014. **307**(10): p. C920-7.
166. Munzer, P., et al., *Acid sphingomyelinase regulates platelet cell membrane scrambling, secretion, and thrombus formation.* Arterioscler Thromb Vasc Biol, 2014. **34**(1): p. 61-71.
167. Zhang, L., et al., *Sphingosine kinase 2 (Sphk2) regulates platelet biogenesis by providing intracellular sphingosine 1-phosphate (S1P).* Blood, 2013. **122**(5): p. 791-802.
168. Thomas, A.S., A. Mehta, and D.A. Hughes, *Gaucher disease: haematological presentations and complications.* Br J Haematol, 2014. **165**(4): p. 427-40.
169. Tani, M., et al., *Mechanisms of sphingosine and sphingosine 1-phosphate generation in human platelets.* J Lipid Res, 2005. **46**(11): p. 2458-67.
170. Rimokh, R., et al., *FVT-1, a novel human transcription unit affected by variant translocation t(2;18)(p11;q21) of follicular lymphoma.* Blood, 1993. **81**(1): p. 136-42.
171. Fan, W., et al., *Distinct subsets of primary effusion lymphoma can be identified based on their cellular gene expression profile and viral association.* J Virol, 2005. **79**(2): p. 1244-51.
172. Czuchlewski, D.R., et al., *Expression of the follicular lymphoma variant translocation 1 gene in diffuse large B-cell lymphoma correlates with subtype and clinical outcome.* Am J Clin Pathol, 2008. **130**(6): p. 957-62.

173. Krebs, S., et al., *A missense mutation in the 3-ketodihydrosphingosine reductase FVT1 as candidate causal mutation for bovine spinal muscular atrophy.* Proc Natl Acad Sci U S A, 2007. **104**(16): p. 6746-51.
174. Rego, A., et al., *The yeast model system as a tool towards the understanding of apoptosis regulation by sphingolipids.* FEMS Yeast Res, 2014. **14**(1): p. 160-78.
175. Ungari, M., et al., *[LAT (linker for activation of T cells): a useful marker for megakaryocyte evaluation on bone marrow biopsies].* Pathologica, 2002. **94**(6): p. 325-30.
176. Gupta, S.D., et al., *Tsc10p and FVT1: topologically distinct short-chain reductases required for long-chain base synthesis in yeast and mammals.* J Lipid Res, 2009. **50**(8): p. 1630-40.
177. Yang, L., et al., *Optimization of scarless human stem cell genome editing.* Nucleic Acids Res, 2013. **41**(19): p. 9049-61.
178. Trevino, A.E. and F. Zhang, *Genome editing using Cas9 nickases.* Methods Enzymol, 2014. **546**: p. 161-74.
179. Newton, J., et al., *Revisiting the sphingolipid rheostat: Evolving concepts in cancer therapy.* Exp Cell Res, 2015. **333**(2): p. 195-200.
180. Dobin, A., et al., *STAR: ultrafast universal RNA-seq aligner.* Bioinformatics, 2013. **29**(1): p. 15-21.
181. Liao, Y., G.K. Smyth, and W. Shi, *featureCounts: an efficient general purpose program for assigning sequence reads to genomic features.* Bioinformatics, 2014. **30**(7): p. 923-30.
182. Ashburner, M., et al., *Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.* Nat Genet, 2000. **25**(1): p. 25-9.
183. Carbon, S., et al., *AmiGO: online access to ontology and annotation data.* Bioinformatics, 2009. **25**(2): p. 288-9.
184. D'Andrea, D., et al., *FIDEA: a server for the functional interpretation of differential expression analysis.* Nucleic Acids Res, 2013. **41**(Web Server issue): p. W84-8.
185. Kim, K.I. and M.A. van de Wiel, *Effects of dependence in high-dimensional multiple testing problems.* BMC Bioinformatics, 2008. **9**: p. 114.
186. *Expansion of the Gene Ontology knowledgebase and resources.* Nucleic Acids Res, 2017. **45**(D1): p. D331-D338.
187. Bill, B.R., et al., *A primer for morpholino use in zebrafish.* Zebrafish, 2009. **6**(1): p. 69-77.
188. Kitatani, K., J. Idkowiak-Baldys, and Y.A. Hannun, *The sphingolipid salvage pathway in ceramide metabolism and signaling.* Cell Signal, 2008. **20**(6): p. 1010-8.
189. !!! INVALID CITATION !!!
190. Noda, T., et al., *Induction of apoptosis of detached oral squamous cell carcinoma cells by safingol. Possible role of Bim, focal adhesion kinase and endonuclease G.* Apoptosis, 2009. **14**(3): p. 287-97.
191. Racke, F.K., et al., *A potential role for protein kinase C-epsilon in regulating megakaryocytic lineage commitment.* J Biol Chem, 2001. **276**(1): p. 522-8.
192. Bouillet, P., et al., *Proapoptotic Bcl-2 relative Bim required for certain apoptotic responses, leukocyte homeostasis, and to preclude autoimmunity.* Science, 1999. **286**(5445): p. 1735-8.
193. De Botton, S., et al., *Platelet formation is the consequence of caspase activation within megakaryocytes.* Blood, 2002. **100**(4): p. 1310-7.
194. Morison, I.M., et al., *A mutation of human cytochrome c enhances the intrinsic apoptotic pathway but causes only thrombocytopenia.* Nat Genet, 2008. **40**(4): p. 387-9.
195. De Rocco, D., et al., *Mutations of cytochrome c identified in patients with thrombocytopenia THC4 affect both apoptosis and cellular bioenergetics.* Biochim Biophys Acta, 2014. **1842**(2): p. 269-74.
196. Josefsson, E.C., et al., *Megakaryocytes possess a functional intrinsic apoptosis pathway that must be restrained to survive and produce platelets.* J Exp Med, 2011. **208**(10): p. 2017-31.
197. Kok, F.O., et al., *Reverse genetic screening reveals poor correlation between morpholino-induced and mutant phenotypes in zebrafish.* Dev Cell, 2015. **32**(1): p. 97-108.
198. Adada, M., et al., *Sphingolipid regulation of ezrin, radixin, and moesin proteins family: implications for cell dynamics.* Biochim Biophys Acta, 2014. **1841**(5): p. 727-37.
199. Zhang, J.P., et al., *Different Effects of sgRNA Length on CRISPR-mediated Gene Knockout Efficiency.* Sci Rep, 2016. **6**: p. 28566.
200. Doench, J.G., et al., *Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation.* Nat Biotechnol, 2014. **32**(12): p. 1262-7.
201. Watson, S., et al., *Phenotypic approaches to gene mapping in platelet function disorders - identification of new variant of P2Y12, TxA2 and GPVI receptors.* Hamostaseologie, 2010. **30**(1): p. 29-38.
202. Russel, F.G., J.B. Koenderink, and R. Masereeuw, *Multidrug resistance protein 4 (MRP4/ABCC4): a versatile efflux transporter for drugs and signalling molecules.* Trends Pharmacol Sci, 2008. **29**(4): p. 200-7.
203. Schuetz, J.D., et al., *MRP4: A previously unidentified factor in resistance to nucleoside-based antiviral drugs.* Nat Med, 1999. **5**(9): p. 1048-51.

204. Belleville-Rolland, T., et al., *MRP4 (ABCC4) as a potential pharmacologic target for cardiovascular disease.* Pharmacol Res, 2016. **107**: p. 381-389.
205. Ravna, A.W. and G. Sager, *Molecular model of the outward facing state of the human multidrug resistance protein 4 (MRP4/ABCC4).* Bioorg Med Chem Lett, 2008. **18**(12): p. 3481-3.
206. Van Aubel, R.A., et al., *Human organic anion transporter MRP4 (ABCC4) is an efflux pump for the purine end metabolite urate with multiple allosteric substrate binding sites.* Am J Physiol Renal Physiol, 2005. **288**(2): p. F327-33.
207. Jedlitschky, G., et al., *The nucleotide transporter MRP4 (ABCC4) is highly expressed in human platelets and present in dense granules, indicating a role in mediator storage.* Blood, 2004. **104**(12): p. 3603-10.
208. Jedlitschky, G., et al., *Role of MRP4 (ABCC4) in platelet adenine nucleotide-storage: evidence from patients with delta-storage pool deficiencies.* Am J Pathol, 2010. **176**(3): p. 1097-103.
209. Mattiello, T., et al., *Aspirin extrusion from human platelets through multidrug resistance protein-4-mediated transport: evidence of a reduced drug action in patients after coronary artery bypass grafting.* J Am Coll Cardiol, 2011. **58**(7): p. 752-61.
210. Lewandrowski, U., et al., *Platelet membrane proteomics: a novel repository for functional research.* Blood, 2009. **114**(1): p. e10-9.
211. Cheepala, S.B., et al., *ABCC4, a plasma membrane transporter modulating platelet aggregation.* Blood, 2015.
212. Oevermann, L., et al., *Hematopoietic stem cell differentiation affects expression and function of MRP4 (ABCC4), a transport protein for signaling molecules and drugs.* Int J Cancer, 2009. **124**(10): p. 2303-11.
213. Cheepala, S.B., et al., *Crucial role for phylogenetically conserved cytoplasmic loop 3 in ABCC4 protein expression.* J Biol Chem, 2013. **288**(31): p. 22207-18.
214. Krishnamurthy, P., et al., *Transporter-mediated protection against thiopurine-induced hematopoietic toxicity.* Cancer Res, 2008. **68**(13): p. 4983-9.
215. Pitre, A., et al., *An unexpected protein interaction promotes drug resistance in leukemia.* Nat Commun, 2017. **8**(1): p. 1547.
216. Schaletzki, Y., et al., *Several adaptor proteins promote intracellular localisation of the transporter MRP4/ABCC4 in platelets and haematopoietic cells.* Thromb Haemost, 2017. **117**(1): p. 105-115.
217. Schofield, E.C., et al., *CHiCP: a web-based tool for the integrative and interactive visualization of promoter capture Hi-C datasets.* Bioinformatics, 2016. **32**(16): p. 2511-3.
218. Javierre, B.M., et al., *Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters.* Cell, 2016. **167**(5): p. 1369-1384 e19.
219. Decouture, B., et al., *Impaired platelet activation and cAMP homeostasis in MRP4-deficient mice.* Blood, 2015. **126**(15): p. 1823-30.
220. Loyau, S., et al., *Platelet glycoprotein VI dimerization, an active process inducing receptor competence, is an indicator of platelet reactivity.* Arterioscler Thromb Vasc Biol, 2012. **32**(3): p. 778-85.
221. Takayama, H., et al., *A novel antiplatelet antibody therapy that induces cAMP-dependent endocytosis of the GPVI/Fc receptor gamma-chain complex.* J Clin Invest, 2008. **118**(5): p. 1785-95.
222. Sassi, Y., et al., *Multidrug resistance-associated protein 4 regulates cAMP-dependent signaling pathways and controls human and rat SMC proliferation.* J Clin Invest, 2008. **118**(8): p. 2747-57.
223. Hara, Y., et al., *Inhibition of MRP4 prevents and reverses pulmonary hypertension in mice.* J Clin Invest, 2011. **121**(7): p. 2888-97.
224. Cheng, D., J. Ren, and E.K. Jackson, *Multidrug resistance protein 4 mediates cAMP efflux from rat preglomerular vascular smooth muscle cells.* Clin Exp Pharmacol Physiol, 2010. **37**(2): p. 205-7.
225. Sassi, Y., et al., *Regulation of cAMP homeostasis by the efflux protein MRP4 in cardiac myocytes.* FASEB J, 2012. **26**(3): p. 1009-17.
226. Sellers, Z.M., et al., *MRP4 and CFTR in the regulation of cAMP and beta-adrenergic contraction in cardiac myocytes.* Eur J Pharmacol, 2012. **681**(1-3): p. 80-7.
227. Copsel, S., et al., *Multidrug resistance protein 4 (MRP4/ABCC4) regulates cAMP cellular levels and controls human leukemia cell proliferation and differentiation.* J Biol Chem, 2011. **286**(9): p. 6979-88.
228. Borgognone, A. and F.M. Pulcinelli, *Reduction of cAMP and cGMP inhibitory effects in human platelets by MRP4-mediated transport.* Thromb Haemost, 2012. **108**(5): p. 955-62.
229. Lien, L.M., et al., *Multidrug resistance protein 4 (MRP4/ABCC4) regulates thrombus formation in vitro and in vivo.* Eur J Pharmacol, 2014. **737**: p. 159-67.
230. Cheung, L., et al., *High-throughput screening identifies Ceefourin 1 and Ceefourin 2 as highly selective inhibitors of multidrug resistance protein 4 (MRP4).* Biochem Pharmacol, 2014. **91**(1): p. 97-108.
231. Broderdorf, S., et al., *cAMP regulates expression of the cyclic nucleotide transporter MRP4 (ABCC4) through the EPAC pathway.* Pharmacogenet Genomics, 2014. **24**(10): p. 522-6.
232. Abla, N., et al., *The human multidrug resistance protein 4 (MRP4, ABCC4): functional analysis of a highly polymorphic gene.* J Pharmacol Exp Ther, 2008. **325**(3): p. 859-68.

233. Markova, S.M. and D.L. Kroetz, *ABCC4 is regulated by microRNA-124a and microRNA-506.* Biochem Pharmacol, 2014. **87**(3): p. 515-22.
234. Kelly, L., et al., *Functional hot spots in human ATP-binding cassette transporter nucleotide binding domains.* Protein Sci, 2010. **19**(11): p. 2110-21.
235. Wittgen, H.G., et al., *Phenylalanine 368 of multidrug resistance-associated protein 4 (MRP4/ABCC4) plays a crucial role in substrate-specific transport activity.* Biochem Pharmacol, 2012. **84**(3): p. 366-73.
236. Yaneff, A., et al., *MRP4/ABCC4 as a new therapeutic target: meta-analysis to determine cAMP binding sites as a tool for drug design.* Curr Med Chem, 2017.
237. Perez, D.R., et al., *Cyclic AMP efflux inhibitors as potential therapeutic agents for leukemia.* Oncotarget, 2016. **7**(23): p. 33960-82.
238. Lecchi, A., et al., *Identification of a new dysfunctional platelet P2Y12 receptor variant associated with bleeding diathesis.* Blood, 2015. **125**(6): p. 1006-13.
239. Frazer, K.A., et al., *A second generation human haplotype map of over 3.1 million SNPs.* Nature, 2007. **449**(7164): p. 851-61.
240. Anders, S., A. Reyes, and W. Huber, *Detecting differential usage of exons from RNA-seq data.* Genome Res, 2012. **22**(10): p. 2008-17.
241. Joutsi-Korhonen, L., et al., *The low-frequency allele of the platelet collagen signaling receptor glycoprotein VI is associated with reduced functional responses and expression.* Blood, 2003. **101**(11): p. 4372-9.
242. Pott, S. and J.D. Lieb, *What are super-enhancers?* Nat Genet, 2015. **47**(1): p. 8-12.
243. Benner, C., et al., *FINEMAP: efficient variable selection using summary data from genome-wide association studies.* Bioinformatics, 2016. **32**(10): p. 1493-501.
244. Cvejic, A., et al., *The role of meis1 in primitive and definitive hematopoiesis during zebrafish development.* Haematologica, 2011. **96**(2): p. 190-8.
245. Nurnberg, S.T., et al., *A GWAS sequence variant for platelet volume marks an alternative DNM3 promoter in megakaryocytes near a MEIS1 binding site.* Blood, 2012. **120**(24): p. 4859-68.
246. Pruim, R.J., et al., *LocusZoom: regional visualization of genome-wide association scan results.* Bioinformatics, 2010. **26**(18): p. 2336-7.
247. Qi, L.S., et al., *Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression.* Cell, 2013. **152**(5): p. 1173-83.
248. Begonja, A.J., et al., *Differential roles of cAMP and cGMP in megakaryocyte maturation and platelet biogenesis.* Exp Hematol, 2013. **41**(1): p. 91-101 e4.
249. Smith, J.R., et al., *Robust, persistent transgene expression in human embryonic stem cells is achieved with AAVS1-targeted integration.* Stem Cells, 2008. **26**(2): p. 496-504.
250. Hong, S.G., et al., *Rhesus iPSC Safe Harbor Gene-Editing Platform for Stable Expression of Transgenes in Differentiated Cells of All Germ Layers.* Mol Ther, 2017. **25**(1): p. 44-53.
251. den Dekker, E., et al., *Development of platelet inhibition by cAMP during megakaryocytopoiesis.* J Biol Chem, 2002. **277**(32): p. 29321-9.
252. den Dekker, E., et al., *Biogenesis of G-protein mediated calcium signaling in human megakaryocytes.* Thromb Haemost, 2001. **86**(4): p. 1106-13.
253. den Dekker, E., et al., *Cyclic AMP raises intracellular Ca(2+) in human megakaryocytes independent of protein kinase A.* Arterioscler Thromb Vasc Biol, 2002. **22**(1): p. 179-86.
254. Meinders, M., et al., *Repercussion of Megakaryocyte-Specific Gata1 Loss on Megakaryopoiesis and the Hematopoietic Precursor Compartment.* PLoS One, 2016. **11**(5): p. e0154342.

8 Appendix

8.1 Supplementary methods

8.1.1 Suppliers

Addgene	Massachusetts, U.S.A
Agilent	California, U.S.A
Alpha Diagnostic	San Antonio, U.S.A
BD Biosciences	San Jose, California, U.S.A
Beckman Coulter	California, U.S.A
Becton Dickinson	Oxford, U.K
Bristol Institute for Transfusion Science	Bristol, UK
Cambridge Advanced Imaging Centre	Cambridge, U.K
Cambridge Institute for Medical Research Central Kitchens, CIMRCK	Cambridge, U.K
Cellgenix	Freiburg, Germany
Cell Signalling	Massachusets, U.S.A
Clontech	California, U.S.A
Diagenode	Liége, Belgium
Diagnostica Stago S.A.S	Asnières sur Seine Cedex, France
GE Healthcare	Little Chalfont, U.K.
Integrated DNA Technologies	Iowa, U.S.A
Invivogen	California, U.S.A
Jackson Laboratory	Maine, U.S.A
Lonza	Basel, Switzerland
New England Biolabs (NEB)	Massachusetts, U.S.A
Promega	Wisconsin, U.S.A
Qiagen	Hilden, Germany
R&D systems	Minneapolis, U.S.A
Roche	Basel, Switzerland
Santa Cruz Biotechnology	Texas, U.S.A
Sigma	Missouri, U.S.A
Takeda	Osaka, Japan
Tocris Biosciences	Bristol, U.K
Thermo Fisher Scientific	Massachusetts, U.S.A
Vectalys	Toulouse, France
Zymo Research	Irvine, California, U.S.A

8.1.2 Images of online phenotyping database.

BRIDGE

ThromboGenomics

HPO Build: 1702

NHS National Institute for Health Research

Cambridge Biomedical Research Centre

General Information Entry Page

rfh000015g

Age At Presentation: [?] 13	**Gender:** [?] F
Ethnicity: [?] Sri Lanka/Seychelles/French Caucasian	
Disorder Type: [?] Both Bleeding and Platelet	
Bleeding Disorder Severity: Mild	
Mode of inheritance: [?] Unknown	
Results of molecular assays: [?]	
Syndrome: [?]	
Family History: [?] Yes ◉ No ○	**Family Recruited:** [?] Yes ◉ No ○
Consanguinity in the family?: [?] Unknown	**Person is related to Index case ID:** [?] rfh000032e, rfh
Family History Details: [?] 2 sisters (recruited), 1 brother (not recruited) all with weak agonist defect. 1 of the sisters has excessive bleeding, the other is asymptomatic. The brother has excessive bleeding.	

BRIDGE

ThromboGenomics

HPO Build: 1702

NHS National Institute for Health Research

Cambridge Biomedical Research Centre

Bleeding Score

rfh000015g

For more information about the Condensed MCMDM-1 VWD scoring system please click here
If you do not have a Condensed MCMDM-1 VWD Bleeding Score please select 'Bleeding Phenotype for patients without Bleeding score' from the Next Page at the bottom of the screen

Epistaxis: [?] 0	**Cutaneous:** [?] 1
Bleeding from minor wounds: [?] 1	**Oral Cavity:** [?] 0
Tooth Extraction: [?] 3	**Surgery:** [?] 0
Menorrhagia: [?] 3	**Postpartum Haemorrhage:** [?] 2
Muscle Haematomas: [?] 0	**Haemarthrosis:** [?] 0
CNS Bleeding: [?] 0	**GI Bleeding:** [?] 0
Total: [?] 10	

BRIDGE

ThromboGenomics

NHS National Institute for Health Research

Cambridge Biomedical Research Centre

Bleeding Full Blood Count Entry Page

rfh000015g

HGB: 11.4 g/dl	**WBC:** 5.53 x10e9/L
PLT: 300 x10e9/L	**MPV:** fl
Macrothrombocytopenia: Yes ○ No ◉	
RBC: 4.4 x10e12/L	**HCT:** 0.365 Ratio
MCV: 83 fl	**MCH:** 25.9 pg
MCHC: 31.2 g/dl	**RDW:** PC
Neut: 3.47 x10e9/L	**Lymph:** 1.3 x10e9/L
Mono: 0.36 x10e9/L	**Eo:** 0.36 x10e9/L
BASO: 0.04 x10e9/L	**Machine Type:** Sysmex

BRIDGE

ThromboGenomics

NHS National Institute for Health Research

Cambridge Biomedical Research Centre

Coagulation

rfh000015g

APTT: 36.1 Seconds	**Prothrombin Time:** 16 Seconds
Fibrinogen: g/L	**Vwf Ag:** 83 U/dl
Vwf Rcof: 99 U/dl	**Factor VIII:** 84 U/dl
Factor II: U/ml	**Factor V:** U/ml
Factor VII: U/ml	**Factor IX:** U/ml
Factor X: U/ml	**Factor XI:** U/ml
Factor XII: U/ml	**Factor XIII:** U/ml

Platelet Function

rfh000015g

PFA100 ADP Collagen: 134 Seconds prolonged	
PFA100 Epinephrine Collagen: 219 Seconds prolonged	
Aggregometry ADP: abnormal	**Aggregometry Epinephrine:** normal
Spontaneous aggregation: Yes ○ No ○ Dont Know ◉	
Aggregometry Collagen: normal	**Aggregometry Ristocetin:** normal
Aggregometry Arachidonic Acid: normal	**Aggregometry Thromboxane Analogue:** normal
Aggregometry Trap:	**ATP Secretion Stimulated with ADP:** Abnormal
ATP Secretion Stimulated with Collagen: Normal	**Flow Cytometry CD41 GPIIB:**
Flow Cytometry CD42a GPIX:	**Flow Cytometry CD42b GPIB alpha:**
Flow Cytometry CD61 GPIIIA:	**Flow Cytometry CD62p Pselectin:**
Nucleotide ADP: 0.26 ADP uMol/Plt Normal	
Nucleotide ATP: 0.4 ATP uMol/Plt Normal	
Nucleotide ATP ADP Ratio: 1.55 Ratio Normal	

BRIDGE

ThromboGenomics

NHS National Institute for Health Research

Cambridge Biomedical Research Centre

Microscopy

rfh000015g

EM Available: Yes ◉ No ○
EM Results: Normal
Bone Marrow Available: Yes ○ No ◉
Bone Marrow Results:
Microscopy confirms large platelets: Yes ○ No ◉
Other Morphological Abnormalities RBC: Yes ○ No ◉
Other Morphological Abnormalities WBC: Yes ○ No ◉
Other Morphological Abnormalities PLT: Yes ○ No ◉
Other morphological abnormalities on microscopy: None

BRIDGE

ThromboGenomics

HPO Build: 1702

NHS National Institute for Health Research

Cambridge Biomedical Research Centre

Overall Clinical Interpretation

rfh000015g

Overall Clinical Interpretation: [?]

Bleeding disorder with weak agonist defect. Strong family history of bleeding.

Patient HPO Entry

Ask for presence of HPO term in other family members: ☑

Tree Selector (click to switch)

- Growth abnormality
- Abnormality of the integument
- Abnormality of the voice
- Abnormality of the cardiovascular system
- Abnormality of blood and blood-forming tissues
  - Abnormality of thrombocytes
    - Abnormal platelet function
      - Impaired platelet adhesion
      - Abnormal platelet membrane protein expression
      - Abnormal platelet granule secretion
      - Abnormal platelet aggregation
        - Impaired platelet aggregation
          - ☑ Impaired ADP-induced platelet aggregation
          - Impaired epinephrine-induced platelet aggregation
          - Impaired collagen-induced platelet aggregation
          - Impaired arachidonic acid-induced platelet aggregation

Click on the heading 'HPO ID' to show a list of HPO ID's which can be copied

HPO ID	HPO Entry Leaf	Family	Action
HP:0004866	Impaired ADP-induced platelet aggregation	Yes	Remove View
HP:0000554	Uveitis	No	Remove View
HP:0011108	Recurrent sinusitis	No	Remove View
HP:0001892	Abnormal bleeding	Yes	Remove View
HP:0011889	Bleeding with minor or no trauma	undetermined	Remove View
HP:0001933	Subcutaneous hemorrhage	undetermined	Remove View
HP:0000132	Menorrhagia	undetermined	Remove View
HP:0011891	Post-partum hemorrhage	undetermined	Remove View
HP:0006298	Prolonged bleeding after dental extraction	undetermined	Remove View
HP:0000118	Phenotypic abnormality	undetermined	Remove View
HP:0004866	Impaired ADP-induced platelet aggregation	undetermined	Remove View

Exit 12 Entries

8.1.3 Example pages from GeneDocs

KDSR

tadbir.bariana@nhs.net / IP: 10.248.201.247

gene: KDSR, gnomad: 0.001, cadd: 10, include: moderate-high

Main page Site help Version

Help Highlight project: None ▾ Only haploid, homozygous or multi-heterozygous variants

Include variants by consequence: no gene consequence frameshift_variant large_deletion missense_variant splice_acceptor_variant stop_gained

Gene	CHR	POS	REF	ALT	Samples	Transcript	Consequence	CADD	gnomAD	HGMD
KDSR	18	60676331		**3.0Mb deletion**	D015836		large_deletion 18:60676331-63722983 (canvas)			
KDSR	18	60999064	C	**T**	E012479, F012302	ENST00000406396	missense_variant	35.00	0.0000578	
KDSR	18	60999065	G	**A**	F010583, C002295*	ENST00000406396	missense_variant	35.00	0.0000244	
KDSR	18	61002504	C	**G**	K002581	ENST00000406396	missense_variant	25.50		
KDSR	18	61002504	C	**T**	M007309, E002644 [BMPR2], F006325, C005266*	ENST00000406396	missense_variant	29.00	0.0000122	
KDSR	18	61006104	G	**A**	D015460	ENST00000406396	stop_gained	47.00	0.0000261	
KDSR	18	61006117	C	**T**	S009225	ENST00000406396	splice_acceptor_variant	27.50	0.0000043	
KDSR	18	61011696	C	**A**	C009007	ENST00000406396	missense_variant	23.70		
KDSR	18	61018197	A	**G**	H011042 [ACTN1]	ENST00000406396	missense_variant	28.80		
KDSR	18	61018218	A	**G**	E009288, R014083, A007217, R010589, R008850, A007327, A014166 [TUBB1], E004142, F000869, A014093, G005021 [SETD5], C006531*, C006715*, D015807	ENST00000406396	missense_variant	25.10	0.0004185	
KDSR	18	61018264	C	**T**	R014537	ENST00000406396	missense_variant	34.00	0.0000041	
KDSR	18	61018282	C	**T**	C005295*	ENST00000406396	missense_variant	34.00	0.0000543	
KDSR	18	61018313	C	**T**	F008094	ENST00000406396	splice_acceptor_variant	26.90		
KDSR	18	61022447	C	**T**	W000221	ENST00000406396	missense_variant	21.30	0.0000253	
KDSR	18	61022463	C	**CA**	T010487	ENST00000406396	stop_gained&frameshift_variant			
KDSR	18	61022741	A	**C**	A007374	ENST00000406396	missense_variant	20.50	0.0000123	
KDSR	18	61022759	T	**C**	A009944*, E009395, S012613*, D012788	ENST00000406396	missense_variant	22.50	0.0000369	
KDSR	18	61026986	A	**G**	E002575	ENST00000406396	missense_variant	22.60	0.0000203	
KDSR	18	61030077	C	**T**	F006360, F000954	ENST00000406396	missense_variant	35.00	0.0000397	
KDSR	18	61034258	CG	**C**	C002858	ENST00000406396	frameshift_variant			
KDSR	18	61034285	T	**C**	C003255, C000572	ENST00000406396	missense_variant	11.74		
KDSR	18	61034291	G	**C**	F010731*	ENST00000406396	missense_variant	11.62	0.0000082	
KDSR	18	61034300	T	**A**	A009778 [RUNX1]	ENST00000406396	missense_variant	14.75		
KDSR	18	61034344	A	**T**	C003538*	ENST00000406396	missense_variant	23.70	0.0000088	

8.2 Supplementary results

8.2.1 78 HPO terms were added to the HPO library by our working group to allow coding of BPD cases

HPO code	HPO label
HP:0001973	Autoimmune thrombocytopenia
HP:0003540	Impaired platelet aggregation
HP:0005527	Reduced kininogen activity
HP:0005528	Bone marrow hypocellularity
HP:0005561	Abnormality of bone marrow cell morphology
HP:0008148	Impaired epinephrine-induced platelet aggregation
HP:0011870	Impaired arachidonic acid-induced platelet aggregation
HP:0011871	Impaired ristocetin-induced platelet aggregation
HP:0011872	Impaired thrombin-induced platelet aggregation
HP:0011877	Increased mean platelet volume
HP:0011878	Abnormal platelet membrane protein expression
HP:0011879	Decreased platelet glycoprotein Ib-IX-V
HP:0011881	Decreased platelet glycoprotein VI
HP:0011882	Decreased platelet P2Y12 receptor
HP:0011883	Abnormal platelet granules
HP:0011894	Impaired thromboxane A2 agonist-induced platelet aggregation
HP:0012483	Abnormal alpha granules
HP:0012484	Abnormal dense granules
HP:0012491	Abnormal dense tubular system
HP:0004813	Post-transfusion thrombocytopenia
HP:0002584	Intestinal bleeding
HP:0007420	Spontaneous hematomas
HP:0007902	Vitreous haemorrhage
HP:0011884	Abnormal umbilical stump bleeding
HP:0011885	Haemorrhage of the eye
HP:0011888	Bleeding requiring red cell transfusion
HP:0011890	Prolonged bleeding following procedure
HP:0011891	Post-partum haemorrhage
HP:0011895	Anaemia due to reduced life span of red cells
HP:0011896	Subconjunctival haemorrhage
HP:0011897	Neutrophilia
HP:0011898	Abnormality of circulating fibrinogen
HP:0011899	Hyperfibrinogenaemia
HP:0011901	Dysfibrinogenemia
HP:0012130	Abnormality of cells of the erythroid lineage
HP:0012131	Abnormal number of erythroid precursors
HP:0012132	Erythroid hyperplasia
HP:0012133	Erythroid hypoplasia
HP:0012135	Abnormality of cells of the granulocytic lineage

HP:0100827	Lymphocytosis
HP:0012146	Abnormality of von Willebrand factor
HP:0012147	Reduced quantity of von Willebrand factor
HP:0012148	Multiple lineage myelodysplasia
HP:0012149	Lineage myelodysplasia
HP:0012150	Single lineage myelodysplasia
HP:0008320	Impaired collagen-induced platelet aggregation
HP:0011873	Abnormal platelet count
HP:0011875	Abnormal platelet morphology
HP:0011889	Bleeding with no or minor trauma
HP:0011869	Abnormal platelet function
HP:0011874	Heparin-induced thrombocytopenia
HP:0011876	Abnormal platelet volume
HP:0011880	Acute disseminated intravascular coagulation
HP:0011886	Hyphema
HP:0011887	Choroid haemorrhage
HP:0011892	Vitamin K deficiency
HP:0011893	Abnormal leukocyte count
HP:0011900	Hypofibrinogenemia
HP:0011902	Abnormal haemoglobin
HP:0011903	Haemoglobin H
HP:0011904	Persistence of haemoglobin F
HP:0011905	Reduced haemoglobin A
HP:0011906	Reduced beta/alpha synthesis ratio
HP:0011907	Reduced alpha/beta synthesis ratio
HP:0011908	Unilateral radial aplasia
HP:0011909	Flattened metacarpal heads
HP:0011910	Shortening of all phalanges of fingers
HP:0011911	Abnormality of metocarpophalangeal joint
HP:0003010	Prolonged bleeding time
HP:0012524	Abnormal platelet shape
HP:0012525	Abnormal alpha granule distribution
HP:0012526	Absence of alpha granules
HP:0012527	Abnormal alpha granule content
HP:0012528	Abnormal number of alpha granules
HP:0012529	Abnormal dense granule content
HP:0012530	Abnormal number of dense granules
HP:0011460	Embryonal onset
HP:0011461	Fetal onset

8.2.2 Raw counts for sphingomyelins and glycosphingolipids measured on the Metabolon platform

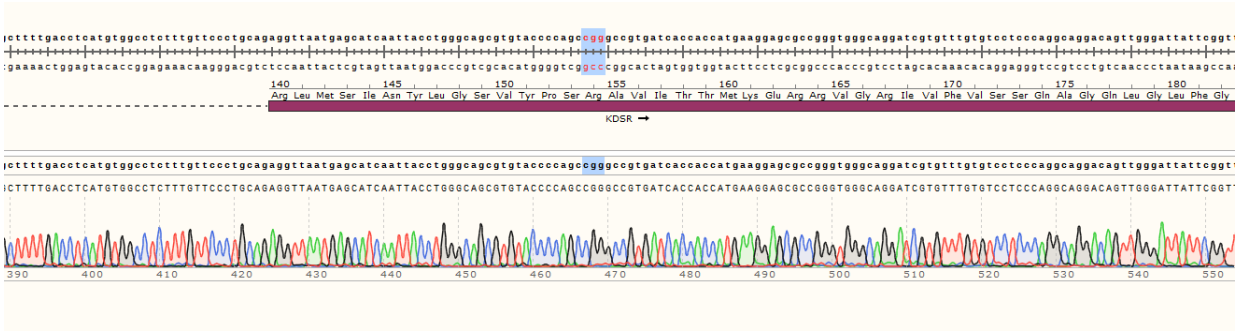
	IPD123	IPD123-mother	IPD123-father	IPD123-brother	median	s.d
palmitoyl sphingomyelin (d18:1/16:0)	3.5E+08	2.8E+08	3.3E+08	3.1E+08	3.1E+08	4.4E+07
sphingomyelin (d18:1/14:0, d16:1/16:0)*	3.6E+07	4.6E+07	3.8E+07	3.1E+07	4.2E+07	1.1E+07
sphingomyelin (d18:1/15:0, d16:1/17:0)*	1.9E+07	2.3E+07	2.1E+07	2.0E+07	2.5E+07	6.1E+06
sphingomyelin (d18:1/17:0, d17:1/18:0, d19:1/16:0)	9.9E+06	1.1E+07	1.0E+07	1.4E+07	1.2E+07	3.1E+06
sphingomyelin (d18:1/18:1, d18:2/18:0)	2.1E+07	2.1E+07	2.1E+07	2.4E+07	2.6E+07	5.2E+06
sphingomyelin (d18:1/20:0, d16:1/22:0)*	4.1E+07	7.0E+07	7.4E+07	6.4E+07	7.1E+07	1.6E+07
sphingomyelin (d18:1/20:1, d18:2/20:0)*	1.1E+07	1.4E+07	1.2E+07	1.5E+07	1.6E+07	3.5E+06
sphingomyelin (d18:1/21:0, d17:1/22:0, d16:1/23:0)*	1.9E+07	3.6E+07	2.6E+07	2.8E+07	3.2E+07	9.9E+06
sphingomyelin (d18:1/22:1, d18:2/22:0, d16:1/24:1)*	8.2E+07	1.1E+08	1.0E+08	8.8E+07	1.0E+08	2.0E+07
sphingomyelin (d18:1/24:1, d18:2/24:0)*	3.7E+08	3.1E+08	3.9E+08	3.3E+08	3.0E+08	5.3E+07
sphingomyelin (d18:2/14:0, d18:1/14:1)*	3.4E+06	6.4E+06	3.6E+06	2.8E+06	4.4E+06	2.0E+06
sphingomyelin (d18:2/16:0, d18:1/16:1)*	7.0E+07	7.2E+07	6.3E+07	6.3E+07	6.4E+07	1.1E+07
sphingomyelin (d18:2/23:0, d18:1/23:1, d17:1/24:1)*	6.5E+07	6.8E+07	5.2E+07	6.6E+07	6.6E+07	1.5E+07
sphingomyelin (d18:2/24:1, d18:1/24:2)*	1.3E+08	1.1E+08	1.2E+08	1.2E+08	1.2E+08	2.2E+07
stearoyl sphingomyelin (d18:1/18:0)	4.3E+07	4.6E+07	5.7E+07	6.5E+07	6.1E+07	1.3E+07
tricosanoyl sphingomyelin (d18:1/23:0)*	7.6E+07	8.1E+07	7.3E+07	1.0E+08	7.7E+07	1.9E+07
glycosyl-N-palmitoyl-sphingosine	6.5E+06	3.7E+06	4.9E+06	5.4E+06	4.1E+06	1.1E+06
glycosyl-N-stearoyl-sphingosine	4.5E+05	5.6E+05	4.4E+05	8.0E+05	5.1E+05	1.5E+05
lactosyl-N-palmitoyl-sphingosine	5.5E+06	4.7E+06	7.0E+06	6.4E+06	5.7E+06	1.4E+06

8.2.3 Z-scores for sphingomyelins and glycosphingolipids measured on the Metabolon platform.

COMPOUND	IPD123	IPD123-mother	IPD123-father	IPD123-brother
palmitoyl-sphingomyelin-d18-1-16-0-	0.9	-0.7	0.4	-0.1
stearoyl-sphingomyelin-d18-1-18-0-	-1.7	-1.3	-0.4	0.3
behenoyl-sphingomyelin-d18-1-22-0-	-0.5	1.1	1.6	0.4
tricosanoyl-sphingomyelin-d18-1-23-0-	0.0	0.3	-0.1	1.3
sphingomyelin-d18-1-14-0-d16-1-16-0-	-0.6	0.4	-0.4	-1.2
sphingomyelin-d18-2-14-0-d18-1-14-1-	-0.6	0.9	-0.5	-1.1
sphingomyelin-d18-1-15-0-d16-1-17-0-	-1.1	-0.3	-0.8	-1.0
sphingomyelin-d18-2-16-0-d18-1-16-1-	0.5	0.7	-0.2	-0.1
sphingomyelin-d18-1-17-0-d17-1-18-0-d19-1-16-0-	-0.7	-0.2	-0.7	0.7
sphingomyelin-d18-1-18-1-d18-2-18-0-	-1.0	-1.1	-1.1	-0.5
sphingomyelin-d18-1-20-0-d16-1-22-0-	-2.5	0.0	0.2	-0.5
sphingomyelin-d18-1-20-1-d18-2-20-0-	-1.7	-0.7	-1.2	-0.3
sphingomyelin-d18-1-21-0-d17-1-22-0-d16-1-23-0-	-1.7	0.4	-0.6	-0.4
sphingomyelin-d18-1-22-1-d18-2-22-0-d16-1-24-1-	-1.2	0.4	0.0	-0.8
sphingomyelin-d18-2-23-0-d18-1-23-1-d17-1-24-1-	-0.1	0.1	-1.1	0.0
sphingomyelin-d18-1-24-1-d18-2-24-0-	1.2	0.1	1.5	0.6
sphingomyelin-d18-2-24-1-d18-1-24-2-	0.5	-0.3	0.2	0.1
glycosyl-N-palmitoyl-sphingosine	1.9	-0.4	0.7	1.1
glycosyl-N-stearoyl-sphingosine	-0.4	0.4	-0.4	1.6
lactosyl-N-palmitoyl-sphingosine	-0.2	-0.9	0.9	0.5

8.2.4 Confirmation of reference sequence at the targeted region in S4 iPSC.

Top: reference sequence. Maroon bar: exon. Lower panel: aligned Sanger sequencing of amplicon showing identical sequence.

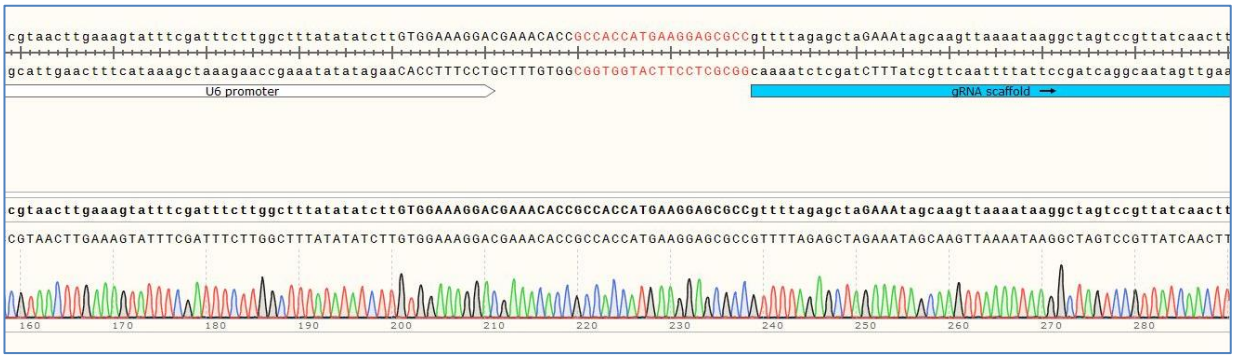


8.2.5 Confirmation by Sanger sequencing of sgRNA integration

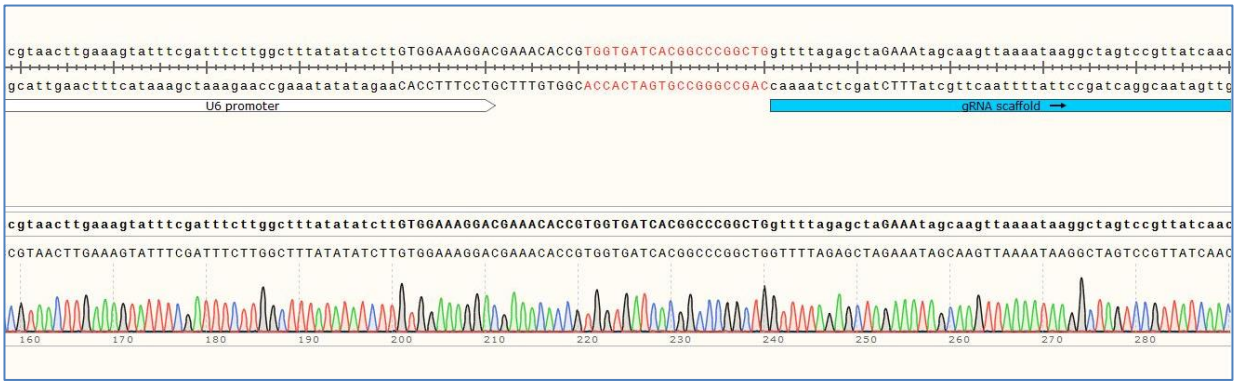
For each Snapgene image, upper panel shows sequence of pSpCas9n (BB)-2A GFP plasmid containing named sgRNA. Lower panel shows Sanger sequencing of amplicon, confirming appropriate ligation of sgRNA. Continued overleaf for sgRNA 66, 67 and 70.

sgRNA52:

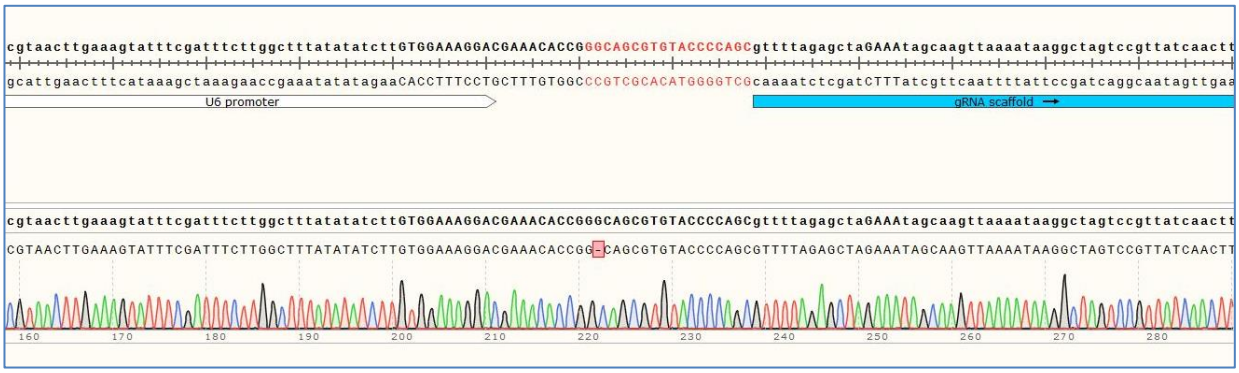
sgRNA57:



sgRNA64:



sgRNA66:



sgRNA67:

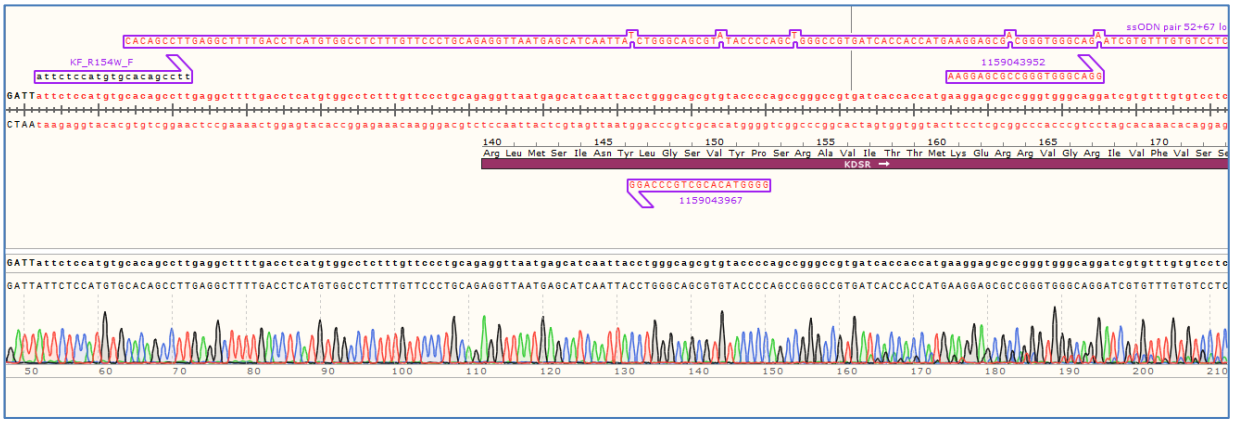
sgRNA70:

8.2.6 pGEM-T Easy identified in-frame indels following *KDSR* CRISPR Cas9n

Figure 82 All pGEM-T Easy colonies from condition 1, iPSC clone 6 showed a 3bp insertion:

Figure 83 Condition 1, iPSC clone 7 showed heterozygosity for a 33bp deletion:

Figure 84 Condition 1, iPSC clone 18 showed the normal amplicon (upper panel) and a probable 12bp deletion (lower panel):



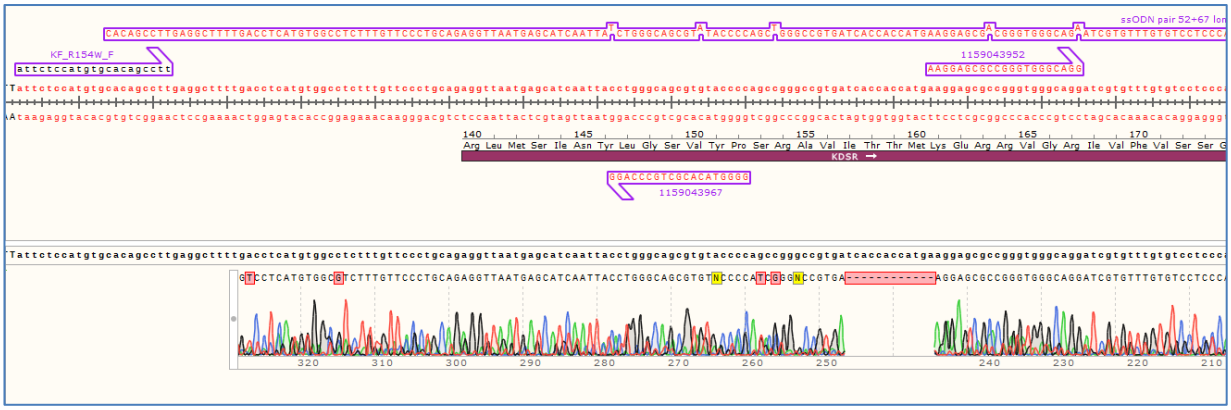


Figure 85 Condition 2, iPSC clone 11 showed the normal amplicon (lower tracing) as well as a messy trace suggestive either of technical artefact, or that more than two colonies had been picked and amplified:

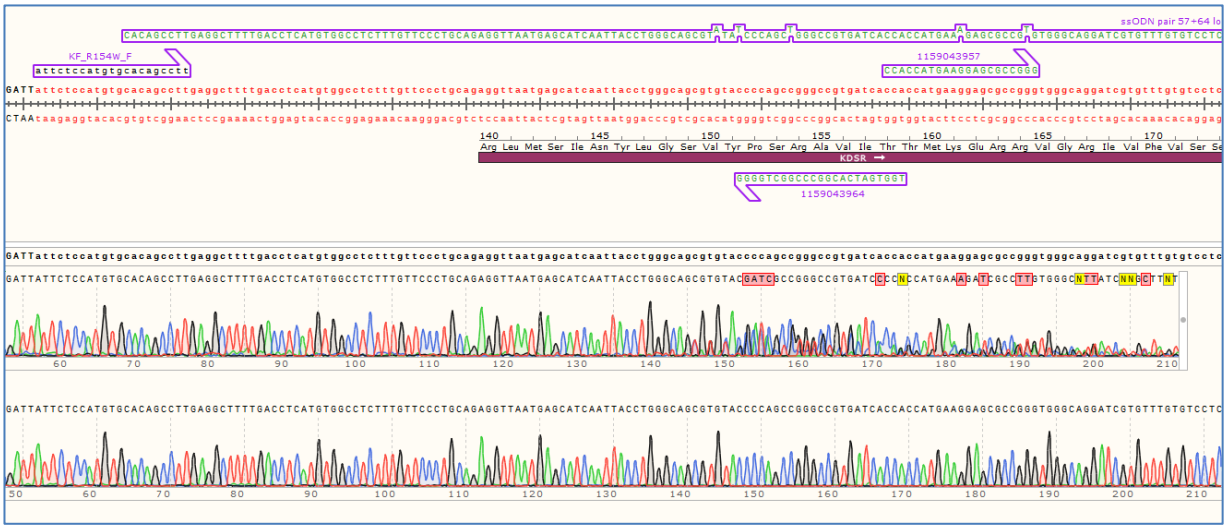
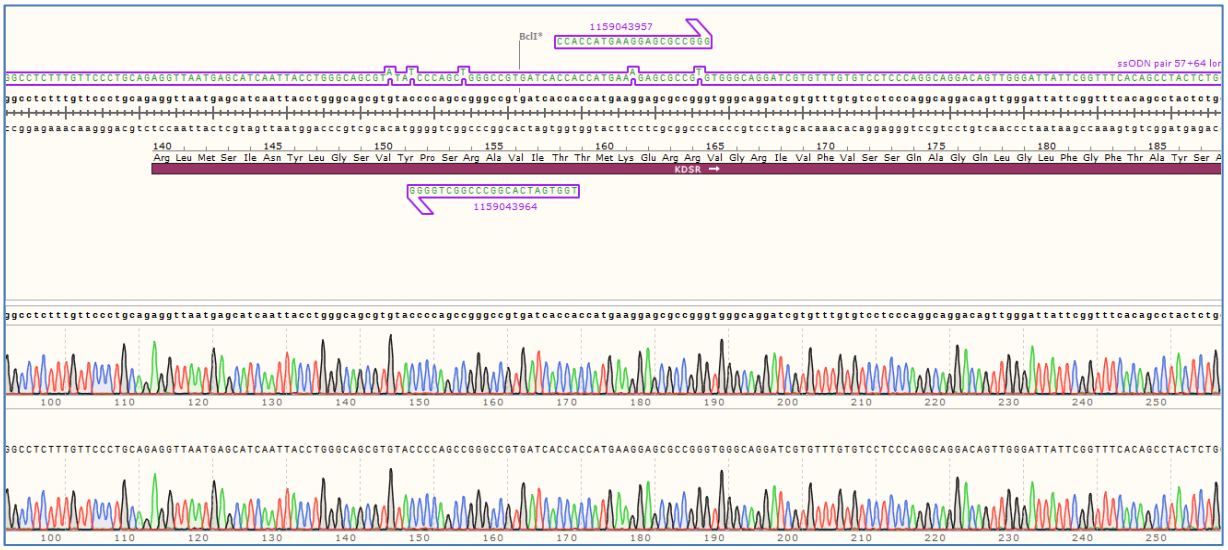
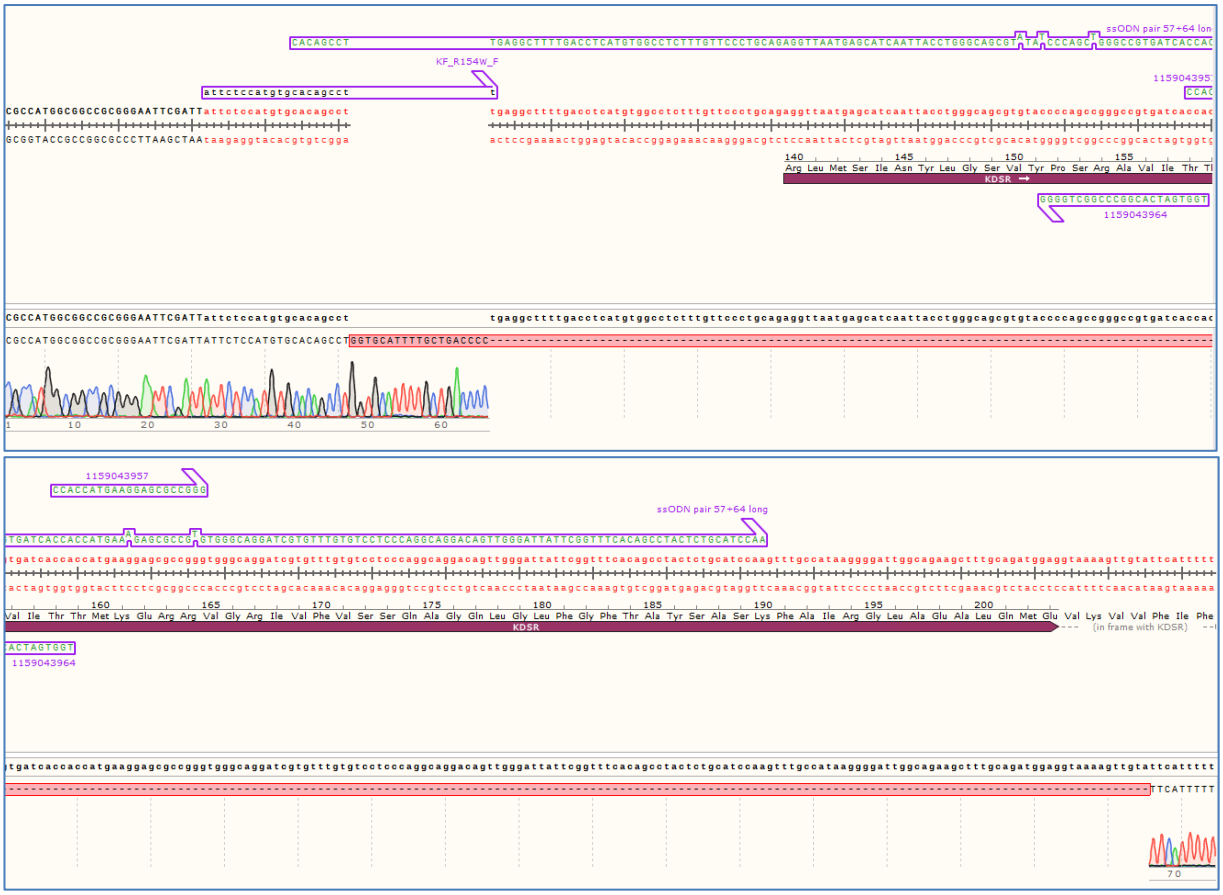


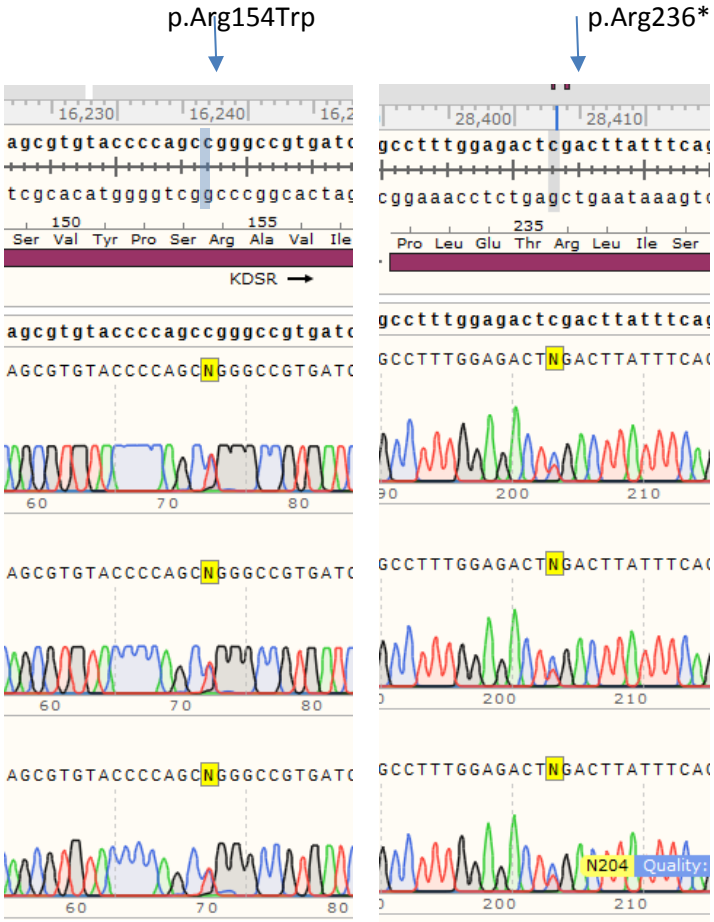
Figure 86 Condition 2, iPSC clone 9 yielded a heterozygous in-frame deletion. Upper panel shows amplicons from pGEM-T Easy colony 3 and 7 which align with the reference sequence. Middle and lower panels show pGEM-T Easy colony 4 with an in-frame insertion (19bp) deletion (244bp):




ssODN pair 57+64 lon
KF_R154W_F
KDSR
1159043964
ssODN pair 57+64 long
1159043957
KDSR
(in frame with KDSR)
1159043964

8.2.7 Sanger sequencing confirmed variants p.Arg154Trp and p.Arg236* in IPD123 iPSC clones

Sanger sequencing confirmed *KDSR* variants (18:61006104 G>A and 18:61018270 G>A) in patient-derived iPSC clones 1,2 and 4. The panels below show forward strand reference sequence at the top and sequencing amplicons from patient-derived iPSC below. Tracings show heterozygosity for reference C (blue) and variant T (red) nucleotides. *KDSR* is transcribed from the reverse strand.



8.2.8 Differentially expressed gene list between iPSC derived from IPD123 and stock iPSC lines QOLG, PODX and BIMA

Ensembl ID	HGNC symbol	Posterior probability	η	log2 fpkm expression estimates by condition and experiment					
				Kev1	Kev2	Kev3	Kresc1	Kresc2	Kresc3
ENSG00000152977	ZIC1	0.968759	-4.55594	0.28	0.27	0.19	4.66	3.69	5.93
ENSG00000174963	ZIC4	0.954105	-3.09541	0.12	0.10	0.05	1.66	1.50	1.90
ENSG00000196092	PAX5	0.949472	-3.24893	0.36	0.34	0.15	3.08	3.59	2.72
ENSG00000169071	ROR2	0.948823	-1.8113	1.30	1.66	1.51	3.70	3.79	3.57
ENSG00000158747	NBL1	0.919898	-1.2706	1.72	1.98	1.86	3.36	3.41	3.43
ENSG00000134201	GSTM5	0.919632	2.93076	2.00	1.78	2.14	0.01	0.04	0.27
ENSG00000157399	ARSE	0.891681	-2.44276	0.66	0.53	0.41	3.37	3.24	2.31
ENSG00000165623	UCMA	0.887947	2.41331	2.33	2.37	2.02	0.23	0.00	0.41
ENSG00000166863	TAC3	0.885075	2.37105	4.28	4.11	4.74	0.77	1.36	1.71
ENSG00000196834	POTEI	0.878937	2.45982	1.27	1.08	1.11	0.04	0.20	0.11
ENSG00000196159	FAT4	0.86657	-2.16544	0.48	0.41	0.59	2.44	3.06	1.93
ENSG00000139211	AMIGO2	0.858353	-2.03206	1.01	1.16	0.80	2.92	3.96	3.22
ENSG00000160321	ZNF208	0.846744	3.69305	1.62	1.19	1.25	0.02	0.02	0.01
ENSG00000156253	RWDD2B	0.826031	-1.72316	0.71	1.08	0.66	2.33	2.75	2.44
ENSG00000168505	GBX2	0.824519	-3.67712	0.98	0.45	0.51	4.64	7.74	4.42
ENSG00000127831	VIL1	0.823897	2.4748	2.66	3.54	3.30	0.36	0.30	1.01
ENSG00000105205	CLC	0.823194	2.1286	1.56	1.48	1.88	0.19	0.08	0.36
ENSG00000114757	PEX5L	0.800912	1.67518	2.49	2.41	2.51	0.77	1.12	0.56
ENSG00000198883	PNMA5	0.787376	2.36252	0.74	1.02	0.54	0.04	0.00	0.09
ENSG00000188219	POTEE	0.786358	3.02111	0.82	1.00	1.20	0.00	0.00	0.04
ENSG00000181333	HEPHL1	0.781828	2.03629	0.23	0.28	0.39	0.01	0.03	0.05
ENSG00000086967	MYBPC2	0.777005	2.20477	2.06	1.87	2.35	0.36	0.10	0.56
ENSG00000081189	MEF2C	0.775965	-2.22206	0.32	0.46	0.12	1.58	2.60	2.07
ENSG00000136944	LMX1B	0.773599	-2.45785	0.26	0.19	0.27	3.33	1.51	1.64
ENSG00000122756	CNTFR	0.754132	-2.39302	2.12	1.10	1.82	4.39	6.38	5.05
ENSG00000019582	CD74	0.753617	1.60361	5.61	5.46	5.52	3.10	2.50	3.76
ENSG00000242866	STRC	0.74487	1.88452	0.93	0.71	0.66	0.11	0.08	0.12
ENSG00000105974	CAV1	0.742091	-1.76216	1.65	2.06	1.44	4.87	3.80	3.84
ENSG00000183486	MX2	0.739462	1.76063	2.91	2.59	2.48	0.07	1.04	0.89
ENSG00000156265	MAP3K7CL	0.725487	-2.77807	0.10	0.09	0.04	0.92	2.55	1.28
ENSG00000198133	TMEM229B	0.714416	1.66951	1.01	0.99	0.81	0.28	0.13	0.14
ENSG00000172803	SNX32	0.713481	1.6191	1.55	1.23	1.23	0.07	0.38	0.37
ENSG00000125148	MT2A	0.708162	-2.0022	1.82	2.07	1.92	6.28	3.87	4.85

ENSG00000143499	SMYD2	0.708036	-2.22836	0.80	1.08	0.57	3.28	4.84	2.68
ENSG00000015413	DPEP1	0.707424	2.17744	2.25	2.36	1.65	0.08	0.12	0.65
ENSG00000164651	SP8	0.706126	-1.69986	2.30	1.97	2.28	5.34	3.78	4.97
ENSG00000171951	SCG2	0.697998	-1.52749	0.45	1.01	0.67	2.00	2.28	2.07
ENSG00000174469	CNTNAP2	0.697586	-0.815626	4.13	4.25	4.28	5.39	5.23	5.51
ENSG00000085491	SLC25A24	0.696063	-1.09623	2.53	2.95	3.07	4.37	4.48	4.30
ENSG00000135638	EMX1	0.692621	2.56652	3.54	2.32	1.75	0.58	0.14	0.16
ENSG00000006128	TAC1	0.691074	1.86551	2.60	1.75	1.80	0.22	0.70	0.39
ENSG00000198865	CCDC152	0.689667	1.67397	2.65	3.06	3.63	1.15	1.37	0.79
ENSG00000204252	HLA-DOA	0.686636	1.48394	4.23	4.11	4.19	2.26	1.49	2.56
ENSG00000151224	MAT1A	0.67533	2.11902	1.94	1.90	2.38	0.23	0.05	0.77
ENSG00000204257	HLA-DMA	0.674268	1.13879	2.78	2.54	2.85	1.18	1.44	1.58
ENSG00000188816	HMX2	0.670453	2.0016	2.66	2.42	2.26	0.27	0.23	1.12
ENSG00000085276	MECOM	0.667155	-2.34417	0.22	0.34	0.15	3.32	1.62	1.41
ENSG00000186417	GLDN	0.665901	1.64814	1.01	0.63	1.17	0.19	0.23	0.19
ENSG00000168675	LDLRAD4	0.660792	-1.77015	0.69	0.51	0.22	1.62	2.33	1.86
ENSG00000213424	KRT222	0.659207	1.68549	0.97	0.88	0.93	0.05	0.09	0.26
ENSG00000105650	PDE4C	0.656674	1.66267	1.07	1.63	1.80	0.41	0.19	0.39
ENSG00000119698	PPP4R4	0.655442	0.962424	3.31	3.38	3.30	1.89	1.95	2.34
ENSG00000172020	GAP43	0.653003	-1.11345	1.97	2.14	2.35	3.66	3.87	3.37
ENSG00000128039	SRD5A3	0.650613	0.798577	4.62	4.41	4.63	3.35	3.44	3.56
ENSG00000074803	SLC12A1	0.646607	-2.24813	0.03	0.06	0.02	0.68	0.55	0.46
ENSG00000177519	RPRM	0.64186	-1.25628	2.04	1.84	2.43	4.13	3.55	3.72
ENSG00000158859	ADAMTS4	0.64121	1.10201	1.97	1.92	2.23	1.08	0.86	1.01
ENSG00000135074	ADAM19	0.639874	-0.948498	3.41	3.75	3.44	4.95	5.01	4.69
ENSG00000167676	PLIN4	0.636168	1.53023	0.51	0.25	0.40	0.07	0.05	0.09
ENSG00000198113	TOR4A	0.633244	1.21616	2.78	2.61	2.42	1.14	1.05	1.54
ENSG00000102195	GPR50	0.633076	-1.48034	0.75	0.71	0.58	2.68	1.76	1.82
ENSG00000129451	KLK10	0.629354	1.71103	3.14	2.81	2.55	0.93	0.21	1.33
ENSG00000164683	HEY1	0.62743	-1.42901	0.66	0.85	0.67	1.90	2.63	1.78
ENSG00000185269	NOTUM	0.623122	-2.27796	0.03	0.04	0.00	0.34	0.45	0.67
ENSG00000204287	HLA-DRA	0.61803	1.46708	0.79	1.18	1.01	0.26	0.13	0.31
ENSG00000142619	PADI3	0.617291	1.66092	1.51	0.74	0.89	0.17	0.17	0.27
ENSG00000105499	PLA2G4C	0.616118	1.65392	3.81	2.82	3.26	0.84	0.76	1.71
Ensembl ID	HGNC								
ENSG00000183379	SYNDIG1L	0.61592	2.01617	1.86	1.93	1.58	0.03	0.14	0.74
ENSG00000173404	INSM1	0.614981	-1.32405	0.82	0.64	0.79	2.13	2.34	1.55
ENSG00000107551	RASSF4	0.611872	-0.669856	2.35	2.36	2.31	3.15	3.14	3.22
ENSG00000179772	FOXS1	0.609857	-2.24242	0.45	0.46	0.00	0.00	3.08	2.46

ENSG00000123453	SARDH	0.608876	1.29998	3.22	3.10	2.98	1.04	1.63	1.83
ENSG00000107317	PTGDS	0.608596	1.92134	2.02	2.98	3.11	0.58	0.11	0.80
ENSG00000196604	POTEF	0.606906	2.10115	2.34	2.19	1.65	0.00	0.72	0.16
ENSG00000095932	SMIM24	0.60662	1.99919	4.27	3.51	4.27	1.06	0.10	2.17
ENSG00000177700	POLR2L	0.604152	-0.674115	5.77	5.89	5.84	6.81	6.70	6.92
ENSG00000156097	GPR61	0.598862	1.7059	0.82	0.72	0.63	0.21	0.16	0.08
ENSG00000183067	IGSF5	0.595471	1.14299	1.93	1.56	1.76	0.79	0.56	0.88
ENSG00000184384	MAML2	0.59369	-1.41029	1.15	1.53	1.09	2.66	3.62	2.66
ENSG00000169515	CCDC8	0.593242	-0.960875	2.84	2.93	2.92	4.59	4.00	4.13
ENSG00000169035	KLK7	0.592842	1.77146	1.39	0.87	1.74	0.21	0.02	0.34
ENSG00000188257	PLA2G2A	0.586677	2.40617	3.94	1.70	3.72	0.43	0.02	0.67
ENSG00000105668	UPK1A	0.581969	1.67675	0.52	0.74	0.40	0.04	0.09	0.10
ENSG00000176014	TUBB6	0.580822	-1.34725	5.21	5.40	4.52	7.80	6.98	6.87
ENSG00000114115	RBP1	0.580457	-0.940367	4.30	4.07	4.41	5.65	5.41	5.90
ENSG00000204071	TCEAL6	0.575414	2.1074	1.43	1.24	0.49	0.04	0.04	0.15
ENSG00000180730	SHISA2	0.57171	1.71881	5.02	3.55	3.48	0.99	1.81	1.76
ENSG00000127418	FGFRL1	0.571134	-1.23353	2.87	2.61	2.52	3.95	5.04	4.40
ENSG00000144057	ST6GAL2	0.570403	-0.936134	2.20	2.46	2.05	3.37	3.36	3.62
ENSG00000213240	NOTCH2NL	0.569434	-1.07682	2.49	2.15	1.81	3.61	3.71	3.43
ENSG00000187569	DPPA3	0.565698	-2.05778	0.18	0.51	0.53	2.82	0.00	2.04
ENSG00000119147	C2orf40	0.563924	1.92182	1.68	2.75	1.94	0.13	0.72	0.02
ENSG00000124098	FAM210B	0.561871	-1.21259	2.19	2.41	2.52	3.77	4.73	3.82
ENSG00000213999	MEF2B	0.55873	2.19095	3.98	1.43	2.93	0.30	0.24	0.67
ENSG00000020181	ADGRA2	0.558118	-1.18713	0.93	1.08	0.91	1.77	2.30	2.61
ENSG00000049130	KITLG	0.55745	-1.24717	0.99	1.02	1.19	2.23	2.99	2.14
ENSG00000187922	LCN10	0.556682	2.22256	0.73	0.66	1.54	0.06	0.06	0.09
ENSG00000103184	SEC14L5	0.5547	1.63169	1.71	1.75	1.57	0.52	0.10	0.69
ENSG00000169442	CD52	0.554454	1.89902	2.82	1.67	2.76	0.50	0.62	0.00
ENSG00000003436	TFPI	0.552081	-1.84504	1.50	1.27	1.10	2.46	5.21	4.33
ENSG00000071575	TRIB2	0.55053	-0.713982	4.51	4.34	4.31	5.37	5.58	5.30
ENSG00000204385	SLC44A4	0.550172	1.3495	1.26	1.63	1.66	0.62	0.21	0.58
ENSG00000162460	TMEM82	0.549716	1.97035	1.12	0.33	0.83	0.07	0.06	0.08
ENSG00000130700	GATA5	0.546453	-2.06251	0.03	0.10	0.14	0.70	0.00	1.24
ENSG00000111452	ADGRD1	0.541672	1.44993	0.93	1.42	1.64	0.24	0.41	0.37
ENSG00000108379	WNT3	0.54069	-1.99918	0.25	0.58	0.40	3.29	2.98	1.10
ENSG00000169403	PTAFR	0.538731	1.46239	0.89	1.12	0.85	0.32	0.04	0.33
ENSG00000167754	KLK5	0.538599	1.81386	0.93	1.39	0.52	0.22	0.10	0.06
ENSG00000137486	ARRB1	0.537672	-1.5454	2.41	2.06	1.92	5.62	3.54	4.39
ENSG00000140836	ZFHX3	0.537171	-1.92578	0.47	0.48	0.56	2.38	4.39	1.37

ENSG00000108001	EBF3	0.536255	-2.51695	0.07	0.07	0.00	1.02	2.44	0.49
ENSG00000105971	CAV2	0.535599	-1.23445	0.92	0.81	0.74	2.45	1.69	1.95
ENSG00000158874	APOA2	0.534491	1.97021	1.14	1.54	1.10	0.14	0.13	0.10
ENSG00000111816	FRK	0.534386	1.57761	0.66	0.28	0.50	0.15	0.06	0.06
ENSG00000023171	GRAMD1B	0.531195	0.651706	2.95	2.96	3.13	2.12	2.22	2.27
ENSG00000088756	ARHGAP28	0.531185	-1.19223	1.73	2.65	2.61	4.13	3.89	3.94
ENSG00000102359	SRPX2	0.53083	-1.60462	0.07	0.00	0.07	0.33	0.47	0.41
ENSG00000185101	ANO9	0.530097	1.58154	2.32	1.82	2.15	0.47	0.09	1.09
ENSG00000076706	MCAM	0.527432	-1.26542	3.27	2.56	2.58	4.32	5.26	4.57
ENSG00000106327	TFR2	0.526623	1.13398	1.47	1.79	1.92	0.67	0.91	0.69
ENSG00000214491	SEC14L6	0.525847	-1.12282	0.34	0.30	0.30	0.75	0.99	0.93
ENSG00000138653	NDST4	0.525055	-3.62005	0.01	0.00	0.00	0.00	1.72	0.95
ENSG00000130829	DUSP9	0.519855	-1.05007	0.62	0.44	0.47	1.36	1.03	1.29
ENSG00000019505	SYT13	0.519209	1.05294	3.13	2.83	3.04	1.52	1.45	2.06
ENSG00000164007	CLDN19	0.517812	1.47717	3.00	2.54	2.76	0.85	0.56	1.72
ENSG00000258947	TUBB3	0.517529	-1.11705	5.30	6.50	5.75	7.60	7.60	7.66
ENSG00000133636	NTS	0.515999	1.88039	6.44	6.26	6.47	4.83	1.25	3.28
ENSG00000205832	C16orf96	0.514665	1.00842	1.27	1.01	1.20	0.49	0.40	0.61
ENSG00000148677	ANKRD1	0.514539	-1.99817	3.24	2.76	1.71	0.00	6.33	5.72
ENSG00000168062	BATF2	0.514261	1.33621	1.48	1.48	1.55	0.28	0.36	0.85
ENSG00000162522	KIAA1522	0.509632	0.457385	5.39	5.53	5.43	4.77	4.80	4.83
ENSG00000169184	MN1	0.508378	-1.42027	2.25	1.98	2.31	3.69	5.68	3.86
ENSG00000180336	C17orf104	0.507551	1.75376	2.87	1.49	2.74	0.32	0.39	0.92
ENSG00000205795	CYS1	0.505443	-1.91903	0.02	0.09	0.00	0.33	0.70	0.64
ENSG00000075223	SEMA3C	0.503807	-1.94421	1.82	0.63	0.96	2.92	5.00	3.63
ENSG00000198729	PPP1R14C	0.50343	-1.93765	0.13	0.06	0.00	0.55	1.36	0.80
ENSG00000170373	CST1	0.501568	2.41753	5.39	3.96	3.60	0.08	0.39	2.21
ENSG00000113448	PDE4D	0.501403	-1.20613	1.05	1.09	0.65	2.11	2.48	1.89

8.2.9 Differential expression of genes between K^{resc} and K^{ev} iMK

110 apoptosis-related genes were differentially expressed following wild-type expression rescue, however the directionality of effect was consistent with the observed phenotype in only 49%.

hgnc_symbol	posterior_ probability	η	Effect of rescue on expression	Role in apoptosis	Expected direction?
BIRC5	0.859458	1.17746	lower	anti	no
BIRC3	0.91735	-1.55897	higher	anti	yes
CYCS	0.76873	0.737613	lower	pro	yes
SMAD1	0.682123	0.688356	lower	pro	yes
CASP4	0.826503	-1.37485	higher	pro	no
TP53I3	0.721239	0.69799	lower	pro	yes
BCL6	0.727574	-1.20913	higher	anti	yes
TNF	0.984907	-2.21384	higher	pro	no
FASLG	0.725215	-1.56708	higher	pro	no
FAS	0.513357	1.08134	lower	pro	yes
TNFRSF14	0.957198	-1.37778	higher	pro	no
TNFSF9	0.627232	-0.923811	higher	pro	no
TNFRSF4	0.527496	-1.45364	higher	pro	no
TNFRSF10B	0.885798	1.52568	lower	pro	yes
NFKB2	0.880487	-1.00913	higher	pro	no
XAF1	0.968219	-2.88706	higher	anti	yes
MDM2	0.849905	0.915674	lower	anti	yes
CARD6	0.970352	-1.93038	higher	pro	yes
BTK	0.588989	-0.431367	higher	pro	yes
BOK	0.516819	-1.28078	higher	anti	no
BMF	0.833764	-1.02481	higher	pro	yes
AATK	0.91209	-2.21877	higher	pro	yes
ADCK3	0.636531	-0.727405	higher	pro	no
APLP1	0.629364	-1.56931	higher	pro	yes
AR	0.554308	-1.55824	higher	pro	yes
ATXN1	0.656994	-1.07358	higher	pro	yes
ANKRD1	0.949116	-2.73083	higher	pro	yes
AMIGO2	0.743479	-0.85869	higher	anti	no
AGT	0.568872	-1.27775	higher	pro	yes
ADORA2A	0.87418	-0.879692	higher	anti	no
ADAR	0.577008	-0.514785	higher	anti	no
CBL	0.728578	-0.50414	higher	anti	no
CDKN2A	0.695666	-1.10617	higher	pro	yes
SMAD6	0.893128	-1.47598	higher	anti	no
CNTFR	0.912848	-2.94323	higher	anti	no
CSF2	0.712479	-2.22557	higher	anti	no

CTGF	0.729686	-0.782798	higher	anti	no
CXCL12	0.959031	-2.15282	higher	anti	no
EGFR	0.77705	-1.97245	higher	pro	yes
EGLN3	0.805879	-0.642208	higher	pro	yes
EP300	0.781541	-0.633441	higher	anti	no
ETS1	0.909963	-1.06345	higher	anti	no
FGF2	0.69738	-1.25767	higher	anti	no
FNDC1	0.97424	-3.15428	higher	pro	yes
FOXC2	0.694217	-2.04861	higher	anti	no
FOXO3	0.89359	-0.629602	higher	pro	yes
FOXS1	0.862262	-1.64269	higher	anti	no
GIMAP5	0.842655	-1.26556	higher	anti	no
GLP1R	0.982165	-3.01014	higher	anti	no
HBB	0.879186	-1.42935	higher	anti	no
HPN	0.945018	-2.20702	higher	pro	yes
HTATIP2	0.855717	-0.973439	higher	pro	yes
ID3	0.924506	-1.36944	higher	anti	no
IER3	0.770921	-0.986243	higher	anti	no
IFI16	0.990042	-1.7426	higher	pro	yes
IFIH1	0.919085	-1.4627	higher	pro	yes
IFIT2	0.976982	-2.01658	higher	anti	no
IFIT3	0.989425	-1.85698	higher	anti	no
IFNB1	0.518676	-1.76216	higher	pro	yes
IGF1R	0.583195	-0.686801	higher	pro	yes
IKZF3	0.653532	-1.67952	higher	anti	no
INHA	0.831017	-1.03435	higher	anti	no
INHBA	0.688058	-1.33161	higher	anti	no
IRF1	0.877744	-1.18231	higher	pro	yes
JAG2	0.884249	-1.70093	higher	anti	no
JAK2	0.586158	-0.709145	higher	anti	no
JAK3	0.731286	-0.597716	higher	anti	no
JMY	0.857766	-0.887535	higher	pro	yes
JUN	0.943482	-0.907574	higher	pro	yes
KDR	0.968964	-1.3194	higher	anti	no
KLF10	0.829709	-0.940309	higher	pro	yes
KLF4	0.885083	-1.34524	higher	pro	yes
KLHL20	0.730172	-1.04147	higher	anti	no
MAP3K9	0.598893	-1.29354	higher	pro	yes
MECP2	0.592926	-0.596183	higher	pro	yes
MMP9	0.639567	-1.28558	higher	pro	yes
MSX1	0.740651	-1.76521	higher	pro	yes

MX1	0.965893	-3.16558	higher	pro	yes
NET1	0.541668	-0.610885	higher	pro	yes
NOTCH1	0.676096	-0.66106	higher	pro	yes
NR4A1	0.794843	-1.2356	higher	pro	yes
NUAK2	0.537843	-1.64694	higher	anti	no
P2RX1	0.945632	-0.621577	higher	pro	yes
PHLDA1	0.943176	-1.08534	higher	anti	no
PIM3	0.927169	-0.962459	higher	anti	no
PINK1	0.622044	-0.918473	higher	anti	no
PLK2	0.608529	-0.777948	higher	pro	yes
PRKCH	0.846332	-1.08344	higher	anti	no
PRNP	0.755303	-0.615728	higher	anti	no
PSEN2	0.590681	-0.712416	higher	anti	no
PSMB8	0.689222	-0.76341	higher	pro	yes
PSMB9	0.875515	-2.45658	higher	pro	yes
RB1CC1	0.528246	-0.662204	higher	pro	yes
RHOB	0.94674	-1.05753	higher	pro	yes
SERPINB9	0.699078	-0.330318	higher	anti	no
SGK1	0.629532	-0.888529	higher	anti	no
SHISA5	0.936581	-1.00269	higher	pro	yes
SKIL	0.611001	-0.537774	higher	anti	no
SNAI1	0.735469	-0.998406	higher	anti	no
SOCS3	0.918728	-1.36399	higher	anti	no
SOD2	0.513342	-0.617557	higher	anti	no
STAT1	0.937098	-1.08912	higher	pro	yes
STK4	0.528617	-0.488622	higher	pro	yes
TGFB1	0.71419	-0.43867	higher	pro	yes
THRA	0.915792	-1.03324	higher	anti	no
TSC22D3	0.848495	-1.2281	higher	anti	no
UNC5B	0.942161	-0.966327	higher	pro	yes
VDR	0.775913	-1.18259	higher	pro	yes
VEGFA	0.64843	-0.794009	higher	anti	no
WT1	0.928996	-2.4442	higher	anti	no

8.2.10 262 genes that contribute to enrichment of GO term GO:0007049 *cell cycle* in K^{resc} and K^{ev} iMK

Gene	Gene Name
AKT1	v-akt murine thymoma viral oncogene homolog 1 [Source:HGNC Symbol;Acc:391]
ANLN	anillin, actin binding protein [Source:HGNC Symbol;Acc:14082]
APITD1	apoptosis-inducing, TAF9-like domain 1 [Source:HGNC Symbol;Acc:23163]
ASPM	asp (abnormal spindle) homolog, microcephaly associated (Drosophila) [Source:HGNC Symbol;Acc:19048]
AURKA	aurora kinase A [Source:HGNC Symbol;Acc:11393]
AURKB	aurora kinase B [Source:HGNC Symbol;Acc:11390]
BCAT1	branched chain amino-acid transaminase 1, cytosolic [Source:HGNC Symbol;Acc:976]
BIRC5	baculoviral IAP repeat containing 5 [Source:HGNC Symbol;Acc:593]
BLM	Bloom syndrome, RecQ helicase-like [Source:HGNC Symbol;Acc:1058]
BRCA2	breast cancer 2, early onset [Source:HGNC Symbol;Acc:1101]
BUB1	budding uninhibited by benzimidazoles 1 homolog (yeast) [Source:HGNC Symbol;Acc:1148]
BUB1B	budding uninhibited by benzimidazoles 1 homolog beta (yeast) [Source:HGNC Symbol;Acc:1149]
BUB3	budding uninhibited by benzimidazoles 3 homolog (yeast) [Source:HGNC Symbol;Acc:1151]
CASC5	cancer susceptibility candidate 5 [Source:HGNC Symbol;Acc:24054]
CCNA2	cyclin A2 [Source:HGNC Symbol;Acc:1578]
CCNB1	cyclin B1 [Source:HGNC Symbol;Acc:1579]
CCNB2	cyclin B2 [Source:HGNC Symbol;Acc:1580]
CCNE1	cyclin E1 [Source:HGNC Symbol;Acc:1589]
CCNE2	cyclin E2 [Source:HGNC Symbol;Acc:1590]
CCNF	cyclin F [Source:HGNC Symbol;Acc:1591]
CDC123	cell division cycle 123 homolog (S. cerevisiae) [Source:HGNC Symbol;Acc:16827]
CDC20	cell division cycle 20 homolog (S. cerevisiae) [Source:HGNC Symbol;Acc:1723]
CDC25A	cell division cycle 25 homolog A (S. pombe) [Source:HGNC Symbol;Acc:1725]
CDC45	cell division cycle 45 homolog (S. cerevisiae) [Source:HGNC Symbol;Acc:1739]
CDC6	cell division cycle 6 homolog (S. cerevisiae) [Source:HGNC Symbol;Acc:1744]
CDC7	cell division cycle 7 homolog (S. cerevisiae) [Source:HGNC Symbol;Acc:1745]
CDCA3	cell division cycle associated 3 [Source:HGNC Symbol;Acc:14624]
CDCA5	cell division cycle associated 5 [Source:HGNC Symbol;Acc:14626]
CDCA8	cell division cycle associated 8 [Source:HGNC Symbol;Acc:14629]
CDK1	cyclin-dependent kinase 1 [Source:HGNC Symbol;Acc:1722]
CDKN3	cyclin-dependent kinase inhibitor 3 [Source:HGNC Symbol;Acc:1791]
CDT1	chromatin licensing and DNA replication factor 1 [Source:HGNC Symbol;Acc:24576]
CENPA	centromere protein A [Source:HGNC Symbol;Acc:1851]

CENPE	centromere protein E, 312kDa [Source:HGNC Symbol;Acc:1856]
CENPF	centromere protein F, 350/400kDa (mitosin) [Source:HGNC Symbol;Acc:1857]
CENPH	centromere protein H [Source:HGNC Symbol;Acc:17268]
CENPJ	centromere protein J [Source:HGNC Symbol;Acc:17272]
CENPK	centromere protein K [Source:HGNC Symbol;Acc:29479]
CENPM	centromere protein M [Source:HGNC Symbol;Acc:18352]
CENPN	centromere protein N [Source:HGNC Symbol;Acc:30873]
CENPO	centromere protein O [Source:HGNC Symbol;Acc:28152]
CENPP	centromere protein P [Source:HGNC Symbol;Acc:32933]
CENPQ	centromere protein Q [Source:HGNC Symbol;Acc:21347]
CENPV	centromere protein V [Source:HGNC Symbol;Acc:29920]
CENPW	centromere protein W [Source:HGNC Symbol;Acc:21488]
CEP55	centrosomal protein 55kDa [Source:HGNC Symbol;Acc:1161]
CHEK1	checkpoint kinase 1 [Source:HGNC Symbol;Acc:1925]
CHEK2	checkpoint kinase 2 [Source:HGNC Symbol;Acc:16627]
CKS1B	CDC28 protein kinase regulatory subunit 1B [Source:HGNC Symbol;Acc:19083]
CKS2	CDC28 protein kinase regulatory subunit 2 [Source:HGNC Symbol;Acc:2000]
CLIP1	CAP-GLY domain containing linker protein 1 [Source:HGNC Symbol;Acc:10461]
CUL1	cullin 1 [Source:HGNC Symbol;Acc:2551]
DBF4	DBF4 homolog (S. cerevisiae) [Source:HGNC Symbol;Acc:17364]
DDX11	DEAD/H (Asp-Glu-Ala-Asp/His) box helicase 11 [Source:HGNC Symbol;Acc:2736]
DHFR	dihydrofolate reductase [Source:HGNC Symbol;Acc:2861]
DLGAP5	discs, large (Drosophila) homolog-associated protein 5 [Source:HGNC Symbol;Acc:16864]
DSCC1	defective in sister chromatid cohesion 1 homolog (S. cerevisiae) [Source:HGNC Symbol;Acc:24453]
DSN1	DSN1, MIND kinetochore complex component, homolog (S. cerevisiae) [Source:HGNC Symbol;Acc:16165]
E2F1	E2F transcription factor 1 [Source:HGNC Symbol;Acc:3113]
E2F2	E2F transcription factor 2 [Source:HGNC Symbol;Acc:3114]
E2F5	E2F transcription factor 5, p130-binding [Source:HGNC Symbol;Acc:3119]
EIF4EBP1	eukaryotic translation initiation factor 4E binding protein 1 [Source:HGNC Symbol;Acc:3288]
ENSA	endosulfine alpha [Source:HGNC Symbol;Acc:3360]
ERCC6L	excision repair cross-complementing rodent repair deficiency, complementation group 6-like [Source:HGNC Symbol;Acc:20794]
ESPL1	extra spindle pole bodies homolog 1 (S. cerevisiae) [Source:HGNC Symbol;Acc:16856]
EXO1	exonuclease 1 [Source:HGNC Symbol;Acc:3511]
EZH2	enhancer of zeste homolog 2 (Drosophila) [Source:HGNC Symbol;Acc:3527]
FAM83D	family with sequence similarity 83, member D [Source:HGNC Symbol;Acc:16122]

FANCD2	Fanconi anemia, complementation group D2 [Source:HGNC Symbol;Acc:3585]
FBXO5	F-box protein 5 [Source:HGNC Symbol;Acc:13584]
FEN1	flap structure-specific endonuclease 1 [Source:HGNC Symbol;Acc:3650]
FOXM1	forkhead box M1 [Source:HGNC Symbol;Acc:3818]
GINS1	GINS complex subunit 1 (Psf1 homolog) [Source:HGNC Symbol;Acc:28980]
GINS2	GINS complex subunit 2 (Psf2 homolog) [Source:HGNC Symbol;Acc:24575]
GMNN	geminin, DNA replication inhibitor [Source:HGNC Symbol;Acc:17493]
GSG2	germ cell associated 2 (haspin) [Source:HGNC Symbol;Acc:19682]
GSPT1	G1 to S phase transition 1 [Source:HGNC Symbol;Acc:4621]
GTSE1	G-2 and S-phase expressed 1 [Source:HGNC Symbol;Acc:13698]
H2AFX	H2A histone family, member X [Source:HGNC Symbol;Acc:4739]
HAUS1	HAUS augmin-like complex, subunit 1 [Source:HGNC Symbol;Acc:25174]
HAUS4	HAUS augmin-like complex, subunit 4 [Source:HGNC Symbol;Acc:20163]
HAUS6	HAUS augmin-like complex, subunit 6 [Source:HGNC Symbol;Acc:25948]
HAUS8	HAUS augmin-like complex, subunit 8 [Source:HGNC Symbol;Acc:30532]
HELLS	helicase, lymphoid-specific [Source:HGNC Symbol;Acc:4861]
HSP90AA1	heat shock protein 90kDa alpha (cytosolic), class A member 1 [Source:HGNC Symbol;Acc:5253]
ILF3	interleukin enhancer binding factor 3, 90kDa [Source:HGNC Symbol;Acc:6038]
INCENP	inner centromere protein antigens 135/155kDa [Source:HGNC Symbol;Acc:6058]
ITGB3BP	integrin beta 3 binding protein (beta3-endonexin) [Source:HGNC Symbol;Acc:6157]
KHDRBS1	KH domain containing, RNA binding, signal transduction associated 1 [Source:HGNC Symbol;Acc:18116]
KIF11	kinesin family member 11 [Source:HGNC Symbol;Acc:6388]
KIF15	kinesin family member 15 [Source:HGNC Symbol;Acc:17273]
KIF20A	kinesin family member 20A [Source:HGNC Symbol;Acc:9787]
KIF20B	kinesin family member 20B [Source:HGNC Symbol;Acc:7212]
KIF22	kinesin family member 22 [Source:HGNC Symbol;Acc:6391]
KIF23	kinesin family member 23 [Source:HGNC Symbol;Acc:6392]
KIF2C	kinesin family member 2C [Source:HGNC Symbol;Acc:6393]
KIFC1	kinesin family member C1 [Source:HGNC Symbol;Acc:6389]
KPNA2	karyopherin alpha 2 (RAG cohort 1, importin alpha 1) [Source:HGNC Symbol;Acc:6395]
LIG1	ligase I, DNA, ATP-dependent [Source:HGNC Symbol;Acc:6598]
LRRCC1	leucine rich repeat and coiled-coil domain containing 1 [Source:HGNC Symbol;Acc:29373]
MAD2L1	MAD2 mitotic arrest deficient-like 1 (yeast) [Source:HGNC Symbol;Acc:6763]
MAD2L2	MAD2 mitotic arrest deficient-like 2 (yeast) [Source:HGNC Symbol;Acc:6764]
MCM10	minichromosome maintenance complex component 10 [Source:HGNC Symbol;Acc:18043]

MCM2	minichromosome maintenance complex component 2 [Source:HGNC Symbol;Acc:6944]
MCM3	minichromosome maintenance complex component 3 [Source:HGNC Symbol;Acc:6945]
MCM4	minichromosome maintenance complex component 4 [Source:HGNC Symbol;Acc:6947]
MCM5	minichromosome maintenance complex component 5 [Source:HGNC Symbol;Acc:6948]
MCM6	minichromosome maintenance complex component 6 [Source:HGNC Symbol;Acc:6949]
MCM7	minichromosome maintenance complex component 7 [Source:HGNC Symbol;Acc:6950]
MCM8	minichromosome maintenance complex component 8 [Source:HGNC Symbol;Acc:16147]
MCMBP	minichromosome maintenance complex binding protein [Source:HGNC Symbol;Acc:25782]
MDM2	Mdm2, p53 E3 ubiquitin protein ligase homolog (mouse) [Source:HGNC Symbol;Acc:6973]
MELK	maternal embryonic leucine zipper kinase [Source:HGNC Symbol;Acc:16870]
MKI67	antigen identified by monoclonal antibody Ki-67 [Source:HGNC Symbol;Acc:7107]
MND1	meiotic nuclear divisions 1 homolog (S. cerevisiae) [Source:HGNC Symbol;Acc:24839]
MNS1	meiosis-specific nuclear structural 1 [Source:HGNC Symbol;Acc:29636]
MSH2	mutS homolog 2, colon cancer, nonpolyposis type 1 (E. coli) [Source:HGNC Symbol;Acc:7325]
MSH3	mutS homolog 3 (E. coli) [Source:HGNC Symbol;Acc:7326]
MSH6	mutS homolog 6 (E. coli) [Source:HGNC Symbol;Acc:7329]
MYBL2	v-myb myeloblastosis viral oncogene homolog (avian)-like 2 [Source:HGNC Symbol;Acc:7548]
NCAPG	non-SMC condensin I complex, subunit G [Source:HGNC Symbol;Acc:24304]
NCAPG2	non-SMC condensin II complex, subunit G2 [Source:HGNC Symbol;Acc:21904]
NCAPH	non-SMC condensin I complex, subunit H [Source:HGNC Symbol;Acc:1112]
NDC80	NDC80 kinetochore complex component homolog (S. cerevisiae) [Source:HGNC Symbol;Acc:16909]
NES	nestin [Source:HGNC Symbol;Acc:7756]
NOLC1	nucleolar and coiled-body phosphoprotein 1 [Source:HGNC Symbol;Acc:15608]
NUF2	NUF2, NDC80 kinetochore complex component, homolog (S. cerevisiae) [Source:HGNC Symbol;Acc:14621]
NUP85	nucleoporin 85kDa [Source:HGNC Symbol;Acc:8734]
NUSAP1	nucleolar and spindle associated protein 1 [Source:HGNC Symbol;Acc:18538]
ODF2	outer dense fiber of sperm tails 2 [Source:HGNC Symbol;Acc:8114]
OIP5	Opa interacting protein 5 [Source:HGNC Symbol;Acc:20300]
ORC1	origin recognition complex, subunit 1 [Source:HGNC Symbol;Acc:8487]
ORC5	origin recognition complex, subunit 5 [Source:HGNC Symbol;Acc:8491]
ORC6	origin recognition complex, subunit 6 [Source:HGNC Symbol;Acc:17151]
PBK	PDZ binding kinase [Source:HGNC Symbol;Acc:18282]
PCNA	proliferating cell nuclear antigen [Source:HGNC Symbol;Acc:8729]
PKMYT1	protein kinase, membrane associated tyrosine/threonine 1 [Source:HGNC Symbol;Acc:29650]
PLK1	polo-like kinase 1 [Source:HGNC Symbol;Acc:9077]
PLK4	polo-like kinase 4 [Source:HGNC Symbol;Acc:11397]

PMF1	polyamine-modulated factor 1 [Source:HGNC Symbol;Acc:9112]
POLA1	polymerase (DNA directed), alpha 1, catalytic subunit [Source:HGNC Symbol;Acc:9173]
POLA2	polymerase (DNA directed), alpha 2 (70kD subunit) [Source:HGNC Symbol;Acc:30073]
POLD1	polymerase (DNA directed), delta 1, catalytic subunit 125kDa [Source:HGNC Symbol;Acc:9175]
POLD2	polymerase (DNA directed), delta 2, regulatory subunit 50kDa [Source:HGNC Symbol;Acc:9176]
POLD3	polymerase (DNA-directed), delta 3, accessory subunit [Source:HGNC Symbol;Acc:20932]
POLE	polymerase (DNA directed), epsilon [Source:HGNC Symbol;Acc:9177]
POLE2	polymerase (DNA directed), epsilon 2 (p59 subunit) [Source:HGNC Symbol;Acc:9178]
PPP1CC	protein phosphatase 1, catalytic subunit, gamma isozyme [Source:HGNC Symbol;Acc:9283]
PPP5C	protein phosphatase 5, catalytic subunit [Source:HGNC Symbol;Acc:9322]
PRC1	protein regulator of cytokinesis 1 [Source:HGNC Symbol;Acc:9341]
PRIM1	primase, DNA, polypeptide 1 (49kDa) [Source:HGNC Symbol;Acc:9369]
PSMA1	proteasome (prosome, macropain) subunit, alpha type, 1 [Source:HGNC Symbol;Acc:9530]
PSMA2	proteasome (prosome, macropain) subunit, alpha type, 2 [Source:HGNC Symbol;Acc:9531]
PSMA3	proteasome (prosome, macropain) subunit, alpha type, 3 [Source:HGNC Symbol;Acc:9532]
PSMA4	proteasome (prosome, macropain) subunit, alpha type, 4 [Source:HGNC Symbol;Acc:9533]
PSMB5	proteasome (prosome, macropain) subunit, beta type, 5 [Source:HGNC Symbol;Acc:9542]
PSMC3	proteasome (prosome, macropain) 26S subunit, ATPase, 3 [Source:HGNC Symbol;Acc:9549]
PSMC5	proteasome (prosome, macropain) 26S subunit, ATPase, 5 [Source:HGNC Symbol;Acc:9552]
PSMD14	proteasome (prosome, macropain) 26S subunit, non-ATPase, 14 [Source:HGNC Symbol;Acc:16889]
PSMD3	proteasome (prosome, macropain) 26S subunit, non-ATPase, 3 [Source:HGNC Symbol;Acc:9560]
PSMD6	proteasome (prosome, macropain) 26S subunit, non-ATPase, 6 [Source:HGNC Symbol;Acc:9564]
PSME3	proteasome (prosome, macropain) activator subunit 3 (PA28 gamma; Ki) [Source:HGNC Symbol;Acc:9570]
PTTG1	pituitary tumor-transforming 1 [Source:HGNC Symbol;Acc:9690]
RACGAP1	Rac GTPase activating protein 1 [Source:HGNC Symbol;Acc:9804]
RAD51	RAD51 homolog (S. cerevisiae) [Source:HGNC Symbol;Acc:9817]
RAD51C	RAD51 homolog C (S. cerevisiae) [Source:HGNC Symbol;Acc:9820]
RAD54B	RAD54 homolog B (S. cerevisiae) [Source:HGNC Symbol;Acc:17228]
RAD54L	RAD54-like (S. cerevisiae) [Source:HGNC Symbol;Acc:9826]
RAN	RAN, member RAS oncogene family [Source:HGNC Symbol;Acc:9846]
RANBP1	RAN binding protein 1 [Source:HGNC Symbol;Acc:9847]
RANGAP1	Ran GTPase activating protein 1 [Source:HGNC Symbol;Acc:9854]
RBBP8	retinoblastoma binding protein 8 [Source:HGNC Symbol;Acc:9891]
RCC1	regulator of chromosome condensation 1 [Source:HGNC Symbol;Acc:1913]
RCC2	regulator of chromosome condensation 2 [Source:HGNC Symbol;Acc:30297]
RFC1	replication factor C (activator 1) 1, 145kDa [Source:HGNC Symbol;Acc:9969]

RFC2	replication factor C (activator 1) 2, 40kDa [Source:HGNC Symbol;Acc:9970]
RFC3	replication factor C (activator 1) 3, 38kDa [Source:HGNC Symbol;Acc:9971]
RFC4	replication factor C (activator 1) 4, 37kDa [Source:HGNC Symbol;Acc:9972]
RFC5	replication factor C (activator 1) 5, 36.5kDa [Source:HGNC Symbol;Acc:9973]
RGCC	regulator of cell cycle [Source:HGNC Symbol;Acc:20369]
RPA2	replication protein A2, 32kDa [Source:HGNC Symbol;Acc:10290]
RPA3	replication protein A3, 14kDa [Source:HGNC Symbol;Acc:10291]
RPS27A	ribosomal protein S27a [Source:HGNC Symbol;Acc:10417]
RRM2	ribonucleotide reductase M2 [Source:HGNC Symbol;Acc:10452]
RRS1	RRS1 ribosome biogenesis regulator homolog (S. cerevisiae) [Source:HGNC Symbol;Acc:17083]
RUVBL1	RuvB-like 1 (E. coli) [Source:HGNC Symbol;Acc:10474]
SAC3D1	SAC3 domain containing 1 [Source:HGNC Symbol;Acc:30179]
SEH1L	SEH1-like (S. cerevisiae) [Source:HGNC Symbol;Acc:30379]
SGOL1	shugoshin-like 1 (S. pombe) [Source:HGNC Symbol;Acc:25088]
SKA1	spindle and kinetochore associated complex subunit 1 [Source:HGNC Symbol;Acc:28109]
SKA3	spindle and kinetochore associated complex subunit 3 [Source:HGNC Symbol;Acc:20262]
SKP2	S-phase kinase-associated protein 2, E3 ubiquitin protein ligase [Source:HGNC Symbol;Acc:10901]
SLBP	stem-loop binding protein [Source:HGNC Symbol;Acc:10904]
SMC2	structural maintenance of chromosomes 2 [Source:HGNC Symbol;Acc:14011]
SPAG5	sperm associated antigen 5 [Source:HGNC Symbol;Acc:13452]
SPC24	SPC24, NDC80 kinetochore complex component, homolog (S. cerevisiae) [Source:HGNC Symbol;Acc:26913]
SPC25	SPC25, NDC80 kinetochore complex component, homolog (S. cerevisiae) [Source:HGNC Symbol;Acc:24031]
STMN1	stathmin 1 [Source:HGNC Symbol;Acc:6510]
STRA13	stimulated by retinoic acid 13 homolog (mouse) [Source:HGNC Symbol;Acc:11422]
SUV39H2	suppressor of variegation 3-9 homolog 2 (Drosophila) [Source:HGNC Symbol;Acc:17287]
TBRG4	transforming growth factor beta regulator 4 [Source:HGNC Symbol;Acc:17443]
TCF3	transcription factor 3 (E2A immunoglobulin enhancer binding factors E12/E47) [Source:HGNC Symbol;Acc:11633]
TFDP1	transcription factor Dp-1 [Source:HGNC Symbol;Acc:11749]
TIMELESS	timeless homolog (Drosophila) [Source:HGNC Symbol;Acc:11813]
TIPIN	TIMELESS interacting protein [Source:HGNC Symbol;Acc:30750]
TOP2A	topoisomerase (DNA) II alpha 170kDa [Source:HGNC Symbol;Acc:11989]
TPX2	TPX2, microtubule-associated, homolog (Xenopus laevis) [Source:HGNC Symbol;Acc:1249]
TRIP13	thyroid hormone receptor interactor 13 [Source:HGNC Symbol;Acc:12307]
TTK	TTK protein kinase [Source:HGNC Symbol;Acc:12401]
TUBB	tubulin, beta class I [Source:HGNC Symbol;Acc:20778]

TUBB4B	tubulin, beta 4B class IVb [Source:HGNC Symbol;Acc:20771]
TUBG1	tubulin, gamma 1 [Source:HGNC Symbol;Acc:12417]
TUBGCP4	tubulin, gamma complex associated protein 4 [Source:HGNC Symbol;Acc:16691]
TYMS	thymidylate synthetase [Source:HGNC Symbol;Acc:12441]
UBA52	ubiquitin A-52 residue ribosomal protein fusion product 1 [Source:HGNC Symbol;Acc:12458]
UBE2C	ubiquitin-conjugating enzyme E2C [Source:HGNC Symbol;Acc:15937]
UBE2I	ubiquitin-conjugating enzyme E2I [Source:HGNC Symbol;Acc:12485]
UBE2S	ubiquitin-conjugating enzyme E2S [Source:HGNC Symbol;Acc:17895]
USP16	ubiquitin specific peptidase 16 [Source:HGNC Symbol;Acc:12614]
VRK1	vaccinia related kinase 1 [Source:HGNC Symbol;Acc:12718]
WEE1	WEE1 homolog (S. pombe) [Source:HGNC Symbol;Acc:12761]
XRCC2	X-ray repair complementing defective repair in Chinese hamster cells 2 [Source:HGNC Symbol;Acc:12829]
YEATS4	YEATS domain containing 4 [Source:HGNC Symbol;Acc:24859]
ZWILCH	Zwilch, kinetochore associated, homolog (Drosophila) [Source:HGNC Symbol;Acc:25468]
ZWINT	ZW10 interactor [Source:HGNC Symbol;Acc:13195]

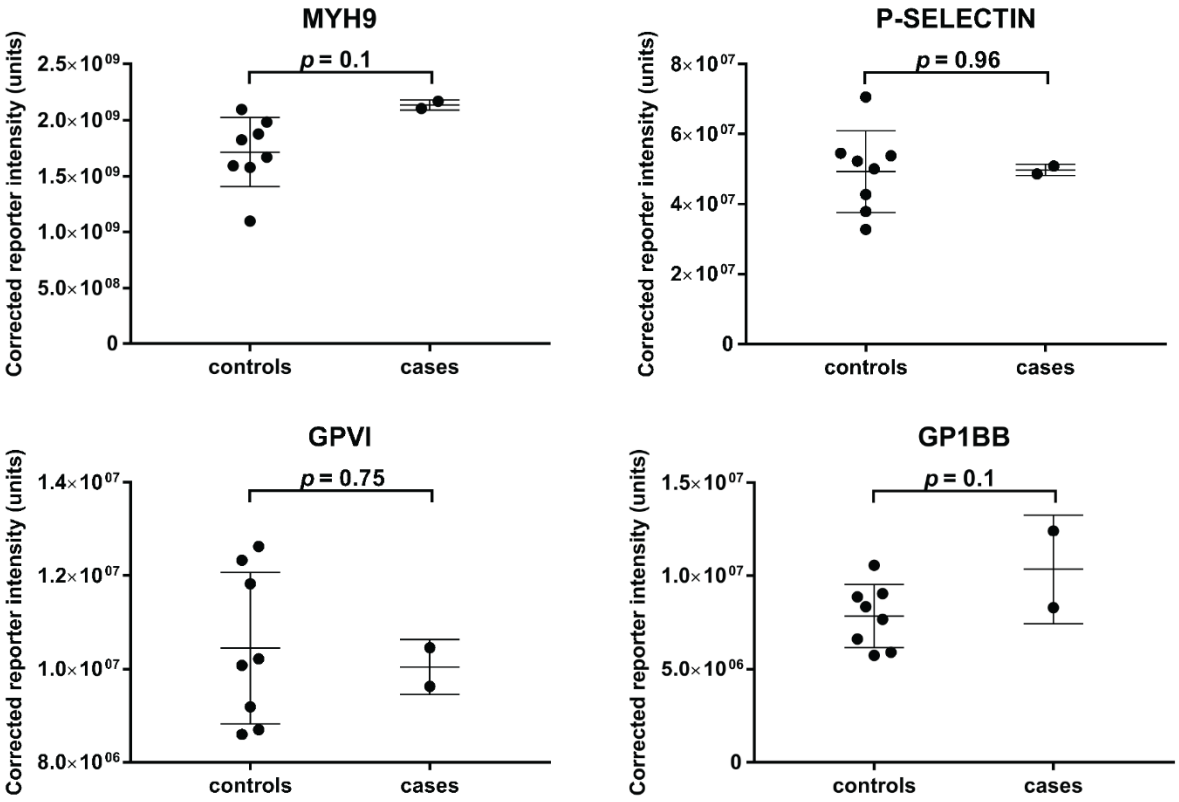
8.2.11 60 genes that contribute to enrichment of GO term GO:0045047 *protein targeting to ER* in K^{resc} and K^{ev} iMK

Gene	Gene Name
RPL10	ribosomal protein L10 [Source:HGNC Symbol;Acc:10298]
RPL10A	ribosomal protein L10a [Source:HGNC Symbol;Acc:10299]
RPL12	ribosomal protein L12 [Source:HGNC Symbol;Acc:10302]
RPL13A	ribosomal protein L13a [Source:HGNC Symbol;Acc:10304]
RPL14	ribosomal protein L14 [Source:HGNC Symbol;Acc:10305]
RPL15	ribosomal protein L15 [Source:HGNC Symbol;Acc:10306]
RPL17	ribosomal protein L17 [Source:HGNC Symbol;Acc:10307]
RPL18	ribosomal protein L18 [Source:HGNC Symbol;Acc:10310]
RPL18A	ribosomal protein L18a [Source:HGNC Symbol;Acc:10311]
RPL19	ribosomal protein L19 [Source:HGNC Symbol;Acc:10312]
RPL21	ribosomal protein L21 [Source:HGNC Symbol;Acc:10313]
RPL22	ribosomal protein L22 [Source:HGNC Symbol;Acc:10315]
RPL23A	ribosomal protein L23a [Source:HGNC Symbol;Acc:10317]
RPL27	ribosomal protein L27 [Source:HGNC Symbol;Acc:10328]
RPL27A	ribosomal protein L27a [Source:HGNC Symbol;Acc:10329]
RPL29	ribosomal protein L29 [Source:HGNC Symbol;Acc:10331]
RPL3	ribosomal protein L3 [Source:HGNC Symbol;Acc:10332]
RPL32	ribosomal protein L32 [Source:HGNC Symbol;Acc:10336]
RPL35	ribosomal protein L35 [Source:HGNC Symbol;Acc:10344]
RPL35A	ribosomal protein L35a [Source:HGNC Symbol;Acc:10345]
RPL37A	ribosomal protein L37a [Source:HGNC Symbol;Acc:10348]
RPL4	ribosomal protein L4 [Source:HGNC Symbol;Acc:10353]
RPL5	ribosomal protein L5 [Source:HGNC Symbol;Acc:10360]
RPL6	ribosomal protein L6 [Source:HGNC Symbol;Acc:10362]
RPL7	ribosomal protein L7 [Source:HGNC Symbol;Acc:10363]
RPL7A	ribosomal protein L7a [Source:HGNC Symbol;Acc:10364]
RPL8	ribosomal protein L8 [Source:HGNC Symbol;Acc:10368]
RPLP0	ribosomal protein, large, P0 [Source:HGNC Symbol;Acc:10371]
RPLP2	ribosomal protein, large, P2 [Source:HGNC Symbol;Acc:10377]
RPN2	ribophorin II [Source:HGNC Symbol;Acc:10382]
RPS10	ribosomal protein S10 [Source:HGNC Symbol;Acc:10383]
RPS11	ribosomal protein S11 [Source:HGNC Symbol;Acc:10384]

RPS12	ribosomal protein S12 [Source:HGNC Symbol;Acc:10385]
RPS13	ribosomal protein S13 [Source:HGNC Symbol;Acc:10386]
RPS14	ribosomal protein S14 [Source:HGNC Symbol;Acc:10387]
RPS16	ribosomal protein S16 [Source:HGNC Symbol;Acc:10396]
RPS18	ribosomal protein S18 [Source:HGNC Symbol;Acc:10401]
RPS19	ribosomal protein S19 [Source:HGNC Symbol;Acc:10402]
RPS2	ribosomal protein S2 [Source:HGNC Symbol;Acc:10404]
RPS20	ribosomal protein S20 [Source:HGNC Symbol;Acc:10405]
RPS21	ribosomal protein S21 [Source:HGNC Symbol;Acc:10409]
RPS23	ribosomal protein S23 [Source:HGNC Symbol;Acc:10410]
RPS24	ribosomal protein S24 [Source:HGNC Symbol;Acc:10411]
RPS25	ribosomal protein S25 [Source:HGNC Symbol;Acc:10413]
RPS27A	ribosomal protein S27a [Source:HGNC Symbol;Acc:10417]
RPS3	ribosomal protein S3 [Source:HGNC Symbol;Acc:10420]
RPS3A	ribosomal protein S3A [Source:HGNC Symbol;Acc:10421]
RPS4X	ribosomal protein S4, X-linked [Source:HGNC Symbol;Acc:10424]
RPS4Y1	ribosomal protein S4, Y-linked 1 [Source:HGNC Symbol;Acc:10425]
RPS5	ribosomal protein S5 [Source:HGNC Symbol;Acc:10426]
RPS6	ribosomal protein S6 [Source:HGNC Symbol;Acc:10429]
RPS7	ribosomal protein S7 [Source:HGNC Symbol;Acc:10440]
RPS8	ribosomal protein S8 [Source:HGNC Symbol;Acc:10441]
RPS9	ribosomal protein S9 [Source:HGNC Symbol;Acc:10442]
RPSA	ribosomal protein SA [Source:HGNC Symbol;Acc:6502]
SEC11A	SEC11 homolog A (S. cerevisiae) [Source:HGNC Symbol;Acc:17718]
SEC61B	Sec61 beta subunit [Source:HGNC Symbol;Acc:16993]
SRP9	signal recognition particle 9kDa [Source:HGNC Symbol;Acc:11304]
SRPRB	signal recognition particle receptor, B subunit [Source:HGNC Symbol;Acc:24085]
UBA52	ubiquitin A-52 residue ribosomal protein fusion product 1 [Source:HGNC Symbol;Acc:12458]

8.2.12 Proteomics assessment of ABCC4 homozygous variant platelets

As detailed in section 5.7 assessment of platelet proteomics showed low levels of ABCC4 protein in platelets from affected cases. There was no difference in the quantification of MYH9, P-selectin, GPVI and GP1B in cases compared with a bank of controls.



8.2.13 Compounds assessed in ABCC4 pedigree by Metabolon global profiling

Compound	Pathway
cAMP	cAMP
bilirubin (E,E)*	Bile acids
bilirubin (E,Z or Z,E)*	
Bilirubin	
I-urobilinogen	
L-urobilin	
Chenodeoxycholate	Primary Bile Acid Metabolism
Cholate	
Glycochenodeoxycholate	
glycochenodeoxycholate glucuronide (1)	
glycochenodeoxycholate sulfate	
Glycocholate	
glycocholate glucuronide (1)	
tauro-beta-muricholate	
Taurochenodeoxycholate	
Taurocholate	
3b-hydroxy-5-cholenoic acid	Secondary Bile Acid Metabolism
7-ketodeoxycholate	
Deoxycholate	
glycocholenate sulfate*	
Glycodeoxycholate	
glycodeoxycholate sulfate	
Glycohyocholate	
Glycolithocholate	
glycolithocholate sulfate*	
Glycoursodeoxycholate	
Hyocholate	
taurocholenate sulfate	
Taurodeoxycholate	
taurolithocholate 3-sulfate	
Tauroursodeoxycholate	
Ursodeoxycholate	
Cortisol	Steroid
Cortisone	
dehydroisoandrosterone sulfate (DHEA-S)	
Urate	Xanthine Metabolism
7-methylurate	
5-oxoproline	Glutathione Metabolism
5-oxoproline	
cys-gly, oxidized	
cysteine-glutathione disulfide	
Cysteinylglycine	