Molecular Genetic Investigation into Inherited Thrombocytopenia

by

Ben David Johnson

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

Birmingham Platelet Group

Institute of Cardiovascular Sciences

College of Medical and Dental Sciences

University of Birmingham

October 2016

Abstract

Inherited thrombocytopenias are a heterogeneous group of disorders characterised by abnormally low platelet counts, often with secondary qualitative defects in platelet function, which can be associated with abnormal bleeding. Next generation sequencing has only previously been employed in small-scale studies and for the confirmation of suspected variants. This study presents the first large-scale approach to fully categorise inherited thrombocytopenia. Ninety-five patients were recruited to the UK-Genotyping and Phenotyping (GAPP) study with inherited thrombocytopenia of unknown genetic aetiology. An average platelet count of $88x10^9$/L was observed across all individuals. Platelet function testing revealed a secondary phenotypic defect in addition to the reduction in platelet count in 71% (61/86). Of the 95 patients, 69 patients, encompassing 47 index cases, were analysed by whole exome sequencing. A variant with a positive prediction of pathogenicity was identified in 40% (19/47) of patients and overall plausible candidate variants of disease were identified in 69% of index cases. Subsequently an inherited thrombocytopenia specific gene panel was developed to serve as a pre-screen for patients prior to whole exome sequencing. In total, candidate variants in genes previously known to be implicated with disease were identified in 77% (20/26) of patients analysed. In conclusion, the application of a combined phenotyping and genotyping approach to patients with inherited thrombocytopenia is an effective and efficient means of complete diagnosis. It also has the added benefit of identifying novel candidate variants in genes not previously known to cause disease that may further our understanding of the processes surrounding haemostasis through subsequent functional studies.

Acknowledgements

Firstly, I would like to thank all collaborators (in particular members of the West Midlands Regional Genetics Service), consultants and referring physicians, providers of equipment and resources, and staff that have helped me over the past three years.

To everyone in the Watson (Platelet) group, thank you for all of your continued help and expertise throughout. Whether you have directly helped me or whether you have just been around to share a cup of tea and distract me from my work with office projectiles, lunges or pictures of cats, it has definitely helped.

Thank you to everyone in particular within the Morgan Group. Abs, Annabel (Annie Mac), Jane, Sarah and Sian. You have all been amazing and I really appreciate everything you have done for me during my thesis.

A special thank you to Neil. I'm sorry for being a pain over the past three years, for suggesting strange ideas that may not have fully worked and for repeating sequencing on patients that was completely unnecessary (I'll pay you back that £800 at some point). Thank you so much though, you have been a great supervisor and a good friend, and although I have moved to the dark side of industry I hope it can stay that way.

Lastly to my friends, family and my girlfriend Grace. You have kept me sane, supplied me with sugar and warm beverages, and have ultimately made sure that I have submitted this one piece of work that I have been slaving over for three years.

Thank you.

Contents

List of Figures

List of Tables

List of Abbreviations

Standard International units of measurement and standard chemical formulae are used. In addition, the following abbreviations are referred to in this thesis:

A	Adenine
AA	Arachidonic acid
ABD	Actin binding domain
ACD	Acid-citrate-dextrose
aCGH	Array comparative genomic hybridization
ACMG	American College of Medical Genetics and Genomics
ADP	Adenosine diphosphate
AML	Acute myelogenous leukaemia
ANA	Anti-nuclear antibodies
ATP	Adenosine triphosphate
BFU-MK	Burst-forming-unit megakaryocyte
BMEC	Bone marrow endothelial cell
BNPH	Bilateral periventricular nodular heterotopia
Bp	Base pair(s)
BPD	Bleeding and platelet disorders
BR	Broad range
BSA	Bovine serum albumin
BSS	Bernard Soulier syndrome
C	Cytosine
CA	5’ CA 3’ dinucleotide (repeat)
CAMT	Congenital amegakaryocytic thrombocytopenia
cDNA	Complementary DNA
CFU-MK	Colony-forming-unit megakaryocyte
CMP	Common myeloid progenitor
CNV	Copy number variant
DAD	Diaphanous auto-regulatory domain
DID	Diaphanous inhibitory domain
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
dNTP’s	Deoxyribonucleoside 5’-triphosphates
EDTA	Ethylenediaminetetraacetic acid
ETS	E26-transformation-specific
ExAC	Exome Aggregation Consortium
FPD/AML	Familial platelet disorder with a predisposition to acute myelogenous leukaemia
G	Guanine
GDP	Guanosine diphosphate
GMP	Granulocyte/macrophage lineage-restricted progenitor

GP	Glycoprotein
GT	Glanzmann thrombasthenia
GTP	Guanosine triphosphate
HLA	Human leukocyte antigen
HPO	Human phenotype ontology
HS	High sensitivity
HSC	Haematopoietic stem cell
IMS	Invaginated membrane system
IPF	Immature platelet fraction
IT	Inherited thrombocytopenia
ITP	Immune thrombocytopenic purpura
KO	Knock-out
LB	Lysogeny broth
LTA	Light transmission aggregometry
MAF	Minor allele frequency
MEP	Myelo-erythroid progenitor
MFI	Mean fluorescent intensity
MPV	Mean platelet volume
mRNA	Messenger RNA
PAM	Protospacer adjacent motif
PBS	Phosphate buffer saline
PCR	Polymerase chain reaction
PFA	Platelet function analyser
PPP	Platelet poor plasma
PRP	Platelet rich plasma
PS	Phosphatidylserine
PVDF	Polyvinylidene difluoride
QC	Quality control
RBC	Red blood cell
RNA	ribonucleic acid
Rpm	Revolutions per minute
RT	Reverse transcriptase
SAM	Segmental aterial mediolysis
SB	Sample buffer
SD	Standard deviation
SDS	Sodium dodecyl sulphate
SNP	Single nucleotide polymorphism
SNV	Single nucleotid variant
T	Thymine
TAE	Tris base, acetic acid, EDTA
TAR	Thrombocytopenia with absent radii
THC2	Thrombocytopenia 2
TPO	Thrombopoietin
Tris	Tri(hydroxymethyl)aminomethane

TTP	Thrombotic thrombocytopenia purpura
TXA2	Thromboxane A2
UK-GAPP	UK-Genotyping and phenotyping of platelets
UTR	Untranslated region
UV	Ultraviolet
VWD	von Willebrand disease
VWF	von-Willebrand factor
WAS	Wiskott-Aldrich syndrome
WGS	Whole genome sequencing
WT	Wild type
XLT	X-linked thrombocytopenia
XLTT	X-linked thrombocytopenia with thalassemia

1. Introduction

Inherited thrombocytopenias (ITs) are a heterogeneous group of disorders primarily characterised by a sustained reduction in platelet count. Clinically they manifest as a bleeding diathesis with varying severity, often dependant on the gene affected by genetic mutation and the presence of a secondary qualitative defect in platelet function. The pathophysiology behind a reduction in platelet count is often situational and to understand the variety within the group of disorders encompassed under IT it is often best to address the role of the individual genes and how their function can be disrupted.

1.1. Megakaryopoiesis and thrombopoiesis

Platelets are the cellular product of thrombopoiesis and fragmentation of their precursor cell, the megakaryocyte. Like all blood cells, megakaryocytes are derived from haematopoietic stem cells (HSCs) and are formed via progressively differentiated progenitor cells. This follows a process known as megakaryopoiesis, but can be described as a section of the broader process of haematopoiesis (Summarised in Figure 1.1.1).

Haematopoiesis often occurs in unique microenvironment niches within the bone marrow that facilitate the self-renewal of HSCs as well as the initial differentiation into progenitor cells. There are at least two described niches within the bone marrow; the osteoblastic niche and the vascular niche (Taichman 2005). The osteoblastic niche refers to a subsection of the bone marrow next to the endosteal bone surface which is lined by osteoblasts. It is thought that this close connection to osteoblasts leads to

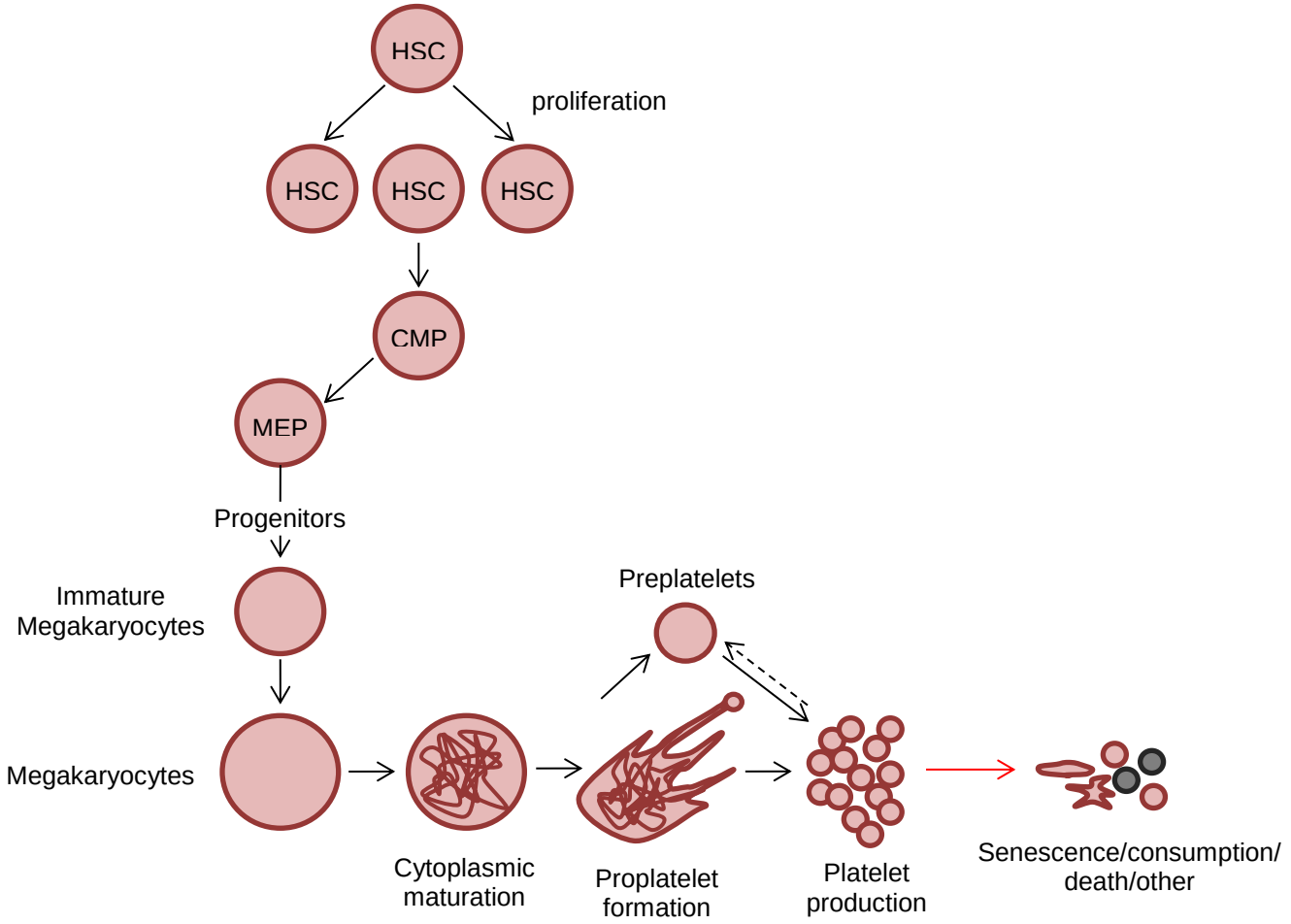


Figure 1.1.1 Schematic of the main steps of megakaryopoiesis. HSC – Haematopoietic Stem Cell, CMP – Common Myeloid Progenitor, MEP – Myelo-Erythroid progenitor

reciprocal connections between the two cell types creating an environment to maintain a quiescent stem cell population (Arai, Hirao et al. 2004; Passegue, Wagers et al. 2005). In contrast, the vascular niche consists of HSCs in close proximity to endothelial cells from fenestrated capillaries and is thought to promote proliferation and differentiation (Kopp, Avecilla et al. 2005). It is this mobilisation between niches, then, that is the driving force behind the differentiation of HSCs to haematopoietic progenitors. This is indeed true for the production of megakaryocytes, as the chemokine-mediated interaction between megakaryocyte progenitors and sinusoidal bone marrow endothelial cells (BMECs) is required for eventual thrombopoiesis (Avecilla, Hattori et al. 2004).

The first step for differentiation into megakaryocytes from HSCs is a commitment to the myelo-erythroid lineage and subsequent divergence from the myelo-lymphoid lineage. This first branch point in the two lineages gives rise to the common myeloid progenitor (CMP), a progenitor with the propensity to differentiate to megakaryocyte/erythrocyte or granulocyte/macrophage progenitors (Akashi, Traver et al. 2000). CMPs show a broad expression profile that shares similarities with both long and short-term HSCs with a relatively high expression of T-cell Acute Lymphocytic Leukaemia 1 (TAL1) but an increasing expression of the transcription factors Nuclear Erythroid, Factor 2 (NF-E2) and GATA binding protein 1 (GATA1)(Akashi, Traver et al. 2000). CMPs can subsequently differentiate into one of the two progenitors; the megakaryocyte/erythrocyte lineage-restricted progenitors (MEPs) or granulocyte/macrophage lineage-restricted progenitors (GMPs). MEPs incorporate a number of progressively differentiated progenitor cells that retain bi-potency and an altered expression profile with elevated GATA binding protein 2

(GATA2) as well as NF-E2 and GATA1 levels when compared with CMPs (Debili, Coulombel et al. 1996; Akashi, Traver et al. 2000).

From this stage the commitment to the formation of megakaryocytes as opposed to erythrocytes is largely dependent on the expression of chemokines and transcription factors. One of the most important chemokines for the formation of megakaryocytes is thrombopoietin (TPO). Synthesised predominately in the liver, TPO is the megakaryocyte specific chemokine that primarily drives megakaryopoiesis and thrombopoiesis (Jelkmann 2001). TPO exerts its synergistic effect on megakaryocyte production through the homodimerisation of its receptor; the cellular homologue of the myeloproliferative leukaemia virus oncogene (c-Mpl), which is one of the definitive cell surface markers of early differentiation along the myelo-erythroid lineage (Kaushansky, Lok et al. 1994; Akashi, Traver et al. 2000). It works in a dose-dependent manner, allowing for progression of HSCs to CMP and MEP progenitors and subsequently burst-forming unit-megakaryocytes (BFU-MKs) and colony-forming unit-megakaryocytes (CFU-MKs), the primitive and mature lineage specific progenitors, respectively (Harker, Marzec et al. 1996; Kaushansky 2008). However, TPO exerts an effect throughout haematopoiesis suggesting that downstream TPO-signalling coupled with non-TPO signalling is important for the production of mature megakaryocytes (Solar, Kerr et al. 1998; Fox, Priestley et al. 2002). Indeed, alternative signalling pathways such as the SDF1/CXCR4 axis, Src family kinases, integrin signalling and platelet factor 4 (PF4) related signalling have also been shown to play a role in *in vitro* and *in vivo* megakaryopoiesis and thrombopoiesis (Yu and Cantor 2012).

Additionally, transcription factors also play a critical role in allowing lineage specific progression. The transcription factor Runt-related transcription factor 1 (RUNX1) acts a master regulator of haematopoiesis. It plays a crucial role in the maintenance, proliferation and differentiation of HSCs due its localisation, transactivation and ability to act as a micro-RNA hub throughout haematopoiesis (North, Stacy et al. 2004; Dowdy, Xie et al. 2010; Rossetti and Sacchi 2013). One of the main downstream transcriptional targets of RUNX1 is the transcription factor PU.1 (Imperato, Cauchy et al. 2015). This is expressed in the myelo-lymphoid lineage in a mutually exclusive pattern to GATA1 which is increasingly expressed during megakaryopoiesis in the myelo-erythroid lineage (Arinobu, Mizuno et al. 2007). This variance in differentiation patterns is maintained within megakaryocyte fated progenitors due to functional cross-antagonism of the transcription factors Friend Leukaemia Integration 1 transcription factor (FLI1) and Kruppel-Like Factor 1 (Erythroid) (KLF1) (Starck, Cohet et al. 2003). Additional transcription factors such as; ETV6, FOG1 and SRF with its co-factor MKL1, in tandem with the transcriptional repressor GFI1B, ensure maturation of the megakaryocyte by binding promoter regions in megakaryocyte expressed genes (Tijssen and Ghevaert 2013; Foudi, Kramer et al. 2014).

In the production of platelets megakaryocyte precursors have to undergo two further steps; endomitosis and proplatelet formation. Endomitosis is the process in which committed megakaryoblasts develop progressively larger and become polyploid promegakaryocytes and then megakaryocytes with the propensity to form platelets (Bluteau, Lordier et al. 2009). Cells follow a normal cell cycle progressing through G1, S and G2 phases followed by an altered mitotic phase. It is within this adapted mitosis that megakaryocytes form a mid-zone in anaphase but have a deficiency in

cytokinesis and cleavage furrow formation (Geddis and Kaushansky 2006; Lordier, Jalil et al. 2008). This process is underpinned by the RUNX1-mediated silencing of non-muscle myosin IIB heavy chain (MYH10) at the contractile ring (Lordier, Bluteau et al. 2012). The net result of which is that megakaryocytes are allowed to form a multilobulated polyploid nucleus that follows geometric distribution from 2N to 64N. Such an increase in genetic content is not a necessity to shed platelets but it has been tied to an associated functional gene amplification (Raslova, Roy et al. 2003).

The theory behind an increase in functional gene amplification is that it allows for compensation of the next step towards thrombopoiesis; proplatelet formation. Following nuclear polyploidisation, megakaryocytes undergo cytoplasmic maturation involving the formation of an invaginated membrane system (IMS). The IMS is an extensive complex of cisternae and tubules continuous with the plasma membrane (Radley and Haller 1982). As a hallmark of mature megakaryocytes, it acts a reservoir of membrane for the formation of proplatelets (Schulze, Korpal et al. 2006).

As the IMS is formed it is loaded with the phospholipid, phosphatidylinositol 4,5-bisphosphate (PI4, P-5 [2]), which associates with filamentous actin promoting polymerisation through Rho-like GTPases and Wiskott-Aldrich syndrome (WASp) family proteins (Schulze, Korpal et al. 2006). Cytoskeletal rearrangements like this power proplatelet formation allowing for elongation of proplatelet extensions into sinusoidal blood vessels (Thon, Montalvo et al. 2010). Microtubule β1-tubulin polymerisation supports the growing mass of proplatelet elongations but sliding of overlapping microtubules utilising the motor protein dynein is essential for full formation (Patel, Richardson et al. 2005). Filamentous-actin (F-actin) is also present throughout proplatelets and forms assembly points allowing proplatelet branching

and bending (Italiano, Lecine et al. 1999). Not only does the cytoskeleton create the necessary structure to support proplatelet elongation, but it also acts as a highway for the transportation of granules and organelles into the budding proplatelet tips. As megakaryocyte precursors enter endomitosis and begin to form an IMS they also begin to increase production of platelet specific granules; namely alpha (α) and dense (δ) granules (Behnke 1968). These granules are derived from budding vesicles from the trans-Golgi network and are transported bi-directionally until they reach proplatelet tips (Behnke 1968; Richardson, Shivdasani et al. 2005). Their transport is facilitated by microtubules and follows the same directional pattern as plus end-directed microtubule motor kinesins (Richardson, Shivdasani et al. 2005).

Once formed, proplatelets have a unique architecture allowing for platelet release into sinusoidal blood vessels. As mentioned previously, maturing megakaryocytes relocate towards the vascular niche where an altered extracellular matrix is found, supporting their commitment to differentiate. Localisation towards the vascular network also exposes megakaryocytes to increasing concentrations of blood components. One particular component, sphingosine 1 phosphate (S1P), guides proplatelet extensions into the lumen of bone marrow sinusoids through an increasing gradient (Zhang, Orban et al. 2012). Once extended into the vascular network, downstream S1P signalling, through the activation of megakaryocyte S1P receptors (S1pr1), leads to platelet release into the circulating blood (Zhang, Orban et al. 2012). Before entering the vascular network though, proplatelet extensions still have to break through into the lumen and this is a process thought to be mediated by podosomes (Schachtner, Calaminus et al. 2013). Podosome formation and stability is dependent on actin polymerisation via the Arp2/3 complex and WASp to create an

actin core (Schachtner, Calaminus et al. 2013). Upon podosome mediated degradation, proplatelet intrusions can reach into the vascular lumen and blood shear forces aid in release of proplatelet fragments from the megakaryocyte body into circulation (Junt, Schulze et al. 2007).

Megakaryocytes do not release uniform platelets into circulation, however; they are believed to be released as a heterogeneous mix of fragments. Contained within this mix of fragments are discoid particles that are 2-10μm in diameter known as preplatelets (Thon, Devine et al. 2012). Preplatelets represent a precursor to mature circulating platelets and can be reversibly converted into barbell-shaped proplatelet structures with two platelet-sized bulbs on each end (Schwertz, Koster et al. 2010; Thon, Devine et al. 2012). This process is governed by the biophysical properties of microtubules and the subsequent abscission of barbell-shaped proplatelets to mature platelets is also mediated by polar located microtubule coils (Thon, Devine et al. 2012; Thon, Macleod et al. 2012). Tight cytoskeletal regulation of platelet maturation in circulation ensures that platelets are formed to a uniform size, however, there is also a strong genetic impact on overall platelet size (Gieger, Radhakrishnan et al. 2011).

Following platelet release, the remaining senescent megakaryocyte nucleus, devoid of the majority of its cytoplasmic components, is disposed of by apoptosis and phagocytosis (Gordge 2005).

1.2. Functional role and lifespan of platelets

Platelets, although anuclear cellular fragments, play a highly important role within haemostasis following vascular injury that was first described by Giulio Bizzozero (Brewer 2006).

Platelets are the smallest of human blood cells with a diameter of around 2µm. They were first described by Max Schultze in 1865 in a study focusing on white blood cells (Brewer 2006). Lacking in a nucleus, platelets have a highly specialised cytoplasmic content that consists of a complex actin and microtubule cytoskeleton which allows for rapid shape change upon activation (Boyles, Fox et al. 1985). The platelet cytoskeleton is a spectrin mesh reinforced by a coiled microtubule tendril that sits beneath the plasma membrane giving platelets their discoid resting shape (Hartwig and DeSisto 1991; Schwer, Lecine et al. 2001). The spectrin strands are interconnected by actin filaments which are cross linked by “scaffolding” molecules such as filamin A (FLNA), filamin B (FLNB) and α-actinin (Tablin, Reeber et al. 1988).

Also located within the unique cytoplasm is a large quantity of platelet specific granules. As mentioned before, platelets contain mostly α-and-dense granules; secretory exosomes that release varying content through membrane fusion (Reed, Fitzgerald et al. 2000). α-granules are the most abundant with a highly organised interior structure. They contain a variety of adhesion molecules, chemokines, adhesive proteins, growth factors and other proteins (Harrison and Cramer 1993). Dense granules, named due to the presence of a highly dense core when viewed by electron microscopy, contain more small compound molecules such as metallic ions and adenosine triphosphate (ATP), guanosine triphosphate (GTP), adenosine

diphosphate (ADP) and guanosine diphosphate (GDP) nucleotides (Youssefian, Masse et al. 1997). Both granules are secreted upon activation of platelet bound membrane receptors through G-protein coupled signalling, the net effect of which is the amplification of the platelet activation pathway.

In normal conditions, platelets follow a sentinel role, unresponsive as they circulate throughout the vascular network (Ruggeri and Mendolicchio 2007). Upon vessel injury, when the endothelial cell layer is damaged exposing the extracellular matrix to circulation, platelets quickly become active. Activation of platelets generally follows three main steps; adhesion, secretion and aggregation.

Platelet adhesion proceeds in slightly different conditions dependant on the shear forces present at the site of vascular injury. In high shear conditions, exposure of collagen from the subendothelial layer allows association with circulating von Willebrand factor (VWF). VWF multimers allow for platelet capture by cross-linking platelet GP1B/V/IX receptor complexes (Clemetson and Clemetson 1995). Tethered platelets are then allowed to roll across the endothelium through endothelial P-selectin binding to its ligand PSGL1 in addition to the GP1b/V/IX complex (Romo, Dong et al. 1999; Frenette, Denis et al. 2000). In low shear conditions, platelets adhere by binding subendothelial collagen directly, cross-linking it using the GP1a/IIa integrin and GPVI which leads to downstream collagen signalling and activation (Nieswandt and Watson 2003).

Once adhered, activated platelets release several granular components which modulate the function of interacting platelets and blood cells. ADP, released from dense granules is predicted to be the prominent amplifier of activation (Gachet 2001).

Secreted ADP binds to its cellular receptors $P2Y_1$ and $P2Y_{12}$ on the platelet surface, mediating platelet shape change and further positive feedback loop associated secretion, respectively (Fabre, Nguyen et al. 1999; Remijn, Wu et al. 2002). Serotonin (5-HT), also located within dense granules, is a known vasoconstrictor and plays a procoagulant role by retaining proteins such as fibrinogen and thrombospondin on the platelet surface (Dale, Friese et al. 2002). α-granules, containing larger adhesive, coagulation and protease inhibitor proteins, are also secreted upon activation. The released alpha granule content aids in platelet activation and the clotting cascade, as well as exposing P-selectin (CD62P) on the platelet surface due to fusion of the granules with the plasma membrane. Exposing P-selectin on the cell surface has an inflammatory response mediating the binding of platelets to other blood cells including neutrophils, monocytes and leukocytes at the site of vascular injury (Furie, Furie et al. 2001; Singbartl, Forlow et al. 2001).

After granule secretion, the next step is the aggregation of platelets to one another to form a haemostatic plug. This process is underpinned by the platelet receptor GPIIb/IIIa ($\alpha_{IIb}\beta_3$-integrin), linking activated platelets through fibrinogen bridges in a two-step process. α-granule exocytosis increases the number of GPIIb/IIIa molecules on the platelet cell surface allowing for binding of soluble plasma fibrinogen, while immobilised fibrinogen on preactivated platelets at the site of vessel injury serves as a substrate for circulating resting platelets. Under high shear forces, platelet aggregation can also occur through the bridging of VWF to the GP1b/V/IX complex which leads to inside-out GPIIb/IIIa signalling and the formation of stable fibrinogen complexes (Ruggeri 1993).

Once platelets begin to aggregate at the site of vascular injury, it can lead to receptor-mediated generation of thrombin allowing the clotting cascade to form a stable platelet-fibrin plug. Upon platelet activation, the constituents of the plasma membrane are altered so that molecules, such as phosphatidylserine (PS), are transported in a flip-flop movement from the inner-to-outer leaflet (Dormann, Kardoeus et al. 1998). Platelets utilise this process to act as a catalyst; amplifying the small amount of thrombin exposed by the endothelia through their own activation and providing a pro-coagulant surface for members of the clotting cascade to bind and form thrombin (Sims, Wiedmer et al. 1989). Also within this process, microvesicles are released due to Ca^{2+}-mediated activation of calpain aiding in coagulation (Wiedmer, Shattil et al. 1990).

An important part of platelet activation is signal transduction allowing for shape change to facilitate the complete formation of a platelet plug. Binding of platelet agonists to cell surface receptors coupled to the Gq-family of G-proteins such as; PAR1, PAR4, and TxA_2-receptor, cause a change in intracellular messenger molecules. The knock on effect is a positive feedback loop, amplifying platelet activation and accommodating the necessary shape change required through re-modelling of the platelet cytokskeleton to allow platelets to spread.

Platelets do not only play a role in the haemostatic regulation in response to vessel injury, they are also a key modulator in a number of processes. They are also active components of the immune system and response, implicated in cancer, a marker of cardiovascular disease and the key regulator behind thrombosis where they are aberrantly activated in the formation of a platelet plaque (Willoughby, Holmes et al. 2002; Smyth, McEver et al. 2009; Bambace and Holmes 2011).

Following activation, platelets are cleared from the circulation usually via the spleen or by phagocytosis. Platelets that have not undergone activation usually have a circulating lifespan of around 7-10 days before they are cleared in the same way. This lifespan is tightly regulated by an intrinsic balance between levels of pro-apoptotic proteins versus pro-survival proteins. Bcl-X_L (encoded by *BCL2L1*), from the Bcl-2 family of pro-survival proteins, is the main connection to platelet lifespan (Mason, Carpinelli et al. 2007). As platelets age, Bcl-X_L levels diminish as the inherited protein from megakaryopoiesis degrades (Bertino, Qi et al. 2003). When levels of Bcl-X_L dip below a threshold, they are insufficient to be able to suppress the pro-apoptotic proteins BAK and BAX leading to platelet death (Wagner, Claudio et al. 2000). It is this degradation over time that acts as a molecular clock for platelets, governing their lifespan; because of this thrombopoiesis has to be tightly regulated in order to maintain a healthy number of circulating platelets.

1.3. Platelet function testing

A platelet's function is often disrupted so that their role in haemostasis upon vascular injury is impaired. Several tests are currently available to test the function of platelets, many of which are currently used in clinical practice. Functional assays are continually progressing in order to achieve more accurate and efficient clinical and research results.

Platelet function testing began with the development of an evaluation of bleeding time by the Duke procedure in 1910 (Duke 1983). Platelet function testing now focuses on a number of key areas including; aggregation, platelet adhesion under

shear stress, analysing secretion and platelet cell surface markers upon activation, and metabolite release.

One of the first developed was the platelet aggregation test in platelet-rich plasma (PRP) known as light transmission aggregometry (LTA). Developed by Born in the 1960s, it functions to determine the effect of agonists on platelets by analysing the amount of light able to pass through a sample (Born 1962). Upon activation, platelets will aggregate in a GPIIb/IIIa manner, converting the opaque PRP clear and increasing the light transmission through the sample. Although one of the first developed assays, it is still considered a gold standard in clinical testing and forms the first step in many methodologies to determine platelet function (Kottke-Marchant and Corcoran 2002). Many agonists are now available to test for a wide array of potential defects and LTA can be coupled with lumiaggregometry to simultaneously determine the secretion of adenine nucleotides from platelet granules by binding the luciferin-luciferase reagent (Holmsen, Holmsen et al. 1966). Aggregometry assays have also progressed beyond LTA to increase efficiency by using a 96 well plate format in the Optimul assay or by assessing aggregation in whole blood through impedance measurements in the Multiplate assay (Toth, Calatzis et al. 2006; Chan, Armstrong et al. 2011).

Platelet adhesion assays under shear stress aim to determine the capabilities of platelets to adhere and aggregate to a coated surface under physiological conditions. Many groups employ their own micro-fluidic system to determine platelet adherence to a number of coated surfaces, predominately collagen and fibrinogen, however commercially available systems including the Platelet Function Analyser – PFA-100 and the IMPACT analyser are available (Kundu, Heilmann et al. 1995; Varon, Dardik

et al. 1997). Both systems work on whole blood flowed or span over pre-coated cartridges and aim to determine the efficacy of primary haemostasis.

To determine the function of platelet secretion and quantification of platelet cell surface markers, flow cytometry is often used. Flow cytometry rapidly measures specific characteristics of platelets and can be adapted to determine the platelet count in whole blood in a similar fashion to how impedance and optical counts are taken. It also has the added benefit of being able to determine a platelet's response to agonists, namely their granular secretion upon activation, the presence of pre-activation and the quantity of cell surface receptors which can elucidate many forms of platelet-related disorders.

In physiological conditions, when platelets are stimulated, thromboxane A2 (TXA2) is rapidly synthesised from the metabolism of arachidonic acid (AA) and released from platelets at the site of vascular injury (Kehrel and Brodde 2013). TXA2 is quickly transformed into the inactive metabolite TXB2 in the circulation, and it is this metabolite that can be measured given an indication on the activation state of platelets as well as their response in a number of diseases (Kehrel and Brodde 2013).

1.4. Inherited thrombocytopenia – Symptoms, treatment and misdiagnosis

Many diseases can affect the function, number and physiology of platelets. More often than not, alterations in platelets are as a secondary symptom to a more severe multisystem disorder or one affecting the immune system. In these cases, platelet numbers and function can crash dramatically but usually improve as the disease

state of the patient resolves or in response to transfusion. The clinical state of being thrombocytopenic is therefore a common occurrence in a hospital environment, however, inherited platelet based bleeding disorders are clinically rarer and considerably more variable in severity than acquired thrombocytopenia. Broadly speaking, platelet based bleeding disorders can be split into two main categories; platelet function disorders and disorders of platelet numbers.

IT is a bleeding diathesis that is categorised by a reduction in platelet count that manifests as a clinical coagulopathy. A normal circulating platelet count in a healthy individual in whole blood varies between 150-450x10^9/L. An individual's platelet count is a unique reflection on a number of genetic and environmental factors including age, ethnicity and gender (Whitfield and Martin 1985; Biino, Balduini et al. 2011; Biino, Gasparini et al. 2012). This platelet count can fluctuate within an individual but is largely confined to differ by less than ±50x10^9/L when measured. To steadily maintain a healthy platelet count within this range, a delicate balance between the production of platelets through megakaryopoiesis and thrombopoiesis and their consumption by activation or removal after senescence is needed. IT often arises when a step in this harmony is disrupted by genetic mutation.

Clinically, IT is defined by a reduction in platelet count to less than 150x10^9/L circulating platelets in whole blood. Its presentation often varies in severity due to the functional effect of genetic variation as well as the presence of a secondary qualitative defect in platelet function. Most patients suffer from mild symptoms related to excessive bleeding including; petechiae, epistaxis, easy cutaneous bruising, periodontal bleeding following brushing and menorrhagia in women that have reached menarche. Patients with a more severe form of IT may present with the

disease from birth in addition to an abnormality affecting the immune system or skeleton. In some cases, the presence of IT can go undiagnosed for many years due to the lack of haemostatic challenge; which is more common in men due to the haemostatic pressures that women face through the menstrual cycle. Although many of the major complications of IT are mild in presentation, they can potentially be life threatening upon certain challenges such as child birth, major surgery or trauma. For this reason, many patients will not receive active treatment but will instead be monitored regularly and the appropriate cause of action be taken if a risk of bleeding is known.

The synthetic antidiuretic hormone desmopressin (DDAVP) is the most commonly used therapy given prior to a planned procedure or following exposure to significant trauma (Cattaneo 2002). It is best administered intravenously by slow infusion over 30 minutes up to a maximum dose of 20µg. Mimicking endogenous vasopressin, desmopressin acts to reduce the excretion of water as urine but also stimulates the release of VWF from endothelial cells by binding the V2 receptor. An increase in VWF release slightly shortens bleeding time in patients with a normal dense granule function (Rao, Ghosh et al. 1995).

Fibrinolytic inhibitor drugs are used in a similar preventive way to desmopressin but show effectiveness mostly in minor operative dental and otolaryngology procedures (Hayward 2005). Fibrinolytic inhibitors include epsilon-aminocaproic acid and tranexamic acid and function to interfere with the formation of plasmin from its precursor plasminogen allowing prolongation of a formed clot.

In few cases, the excessive associated bleeding cannot be controlled prior to haemostatic risk, this is most often in severe cases of thrombocytopenia and a more active treatment plan is needed. The most commonly used active treatment is platelet transfusion and in ideal circumstances human leukocyte antigen (HLA)-matched single donor platelets are used, but leukocyte-depleted blood components can be used. Transfusions are used sparingly to avoid alloimmunisation which may lead to refractoriness to therapy (Fiore, Firah et al. 2012). Therapies such as recombinant factor VIIIa, are reserved for serious cases when where a lack of responsiveness to transfusion has occurred due to alloimmunisation (Kaleelrahman, Minford et al. 2004).

Recent advances in treatment has seen the use of TPO mimetics, haematopoietic stem cell transplantation and stem cell gene therapy, all being utilised in individual platelet based bleeding disorders (Locatelli, Rossi et al. 2003; Boztug, Schmidt et al. 2010; Pecci, Gresele et al. 2010).

Correct diagnosis of IT is crucial to be able to administer the correct treatment plan. Often cases of IT are misdiagnosed as immune thrombocytopenic purpura (ITP), an autoimmune disorder which leads to immune mediated thrombocytopenia (Bader-Meunier, Proulle et al. 2003). ITP is categorised by either a presence of platelet autoantibodies causing phagocytosis and subsequent removal of platelets or T lymphocyte-mediated platelet lysis (Cines, Bussel et al. 2004). Platelet counts in ITP can crash dramatically to less than $10x10^9$/L requiring a more intensive therapy to avoid significant bleeding. Initial treatment aims to create a stable platelet count and is usually achieved through the use of intravenous gammaglobulin, the corticosteroid methylprednisolone or a combination of both (Cines, Bussel et al. 2004). Both

treatments cause the spleen to avoid platelets tagged for removal, therefore maintaining the platelet count. They can be quite harsh and are associated with a number of side effects which may affect a patient's wellbeing. If a stable platelet count cannot be achieved with either treatment, a number of other treatments including anti-D and recombinant factor VIIIa can be used and in worst cases, a splenectomy is recommended. As most treatments target the over-excessive clearance of platelets, they often have little-to-no effect on patients suffering from IT, meaning that patients can go through several unnecessary therapies that will offer no benefit. As such a clear distinguished diagnosis between the two diseases should be achieved before therapy is started. This is often difficult to determine without pre-testing for a number of genetic mutations or by assessing how the patient responds to treatment, so a definitive family history and a series of platelet counts over a time period are crucial steps.

1.5. Genetic causes of inherited thrombocytopenia

In comparison to acquired ITP, ITs are relatively rare in their prevalence. Most recently they have been reported as being present in 2.7 per 10,000 live births, although their prevalence varies considerably between each individual disease. To date, 35 genes have been reported to cause 30 forms of IT (Table 1.5.1). Each form varies in its clinical presentation, severity, association with secondary symptoms and cause. Many groups have sought to separate the spectrum of disorders by a number of factors including; platelet count, mean platelet volume (MPV)/platelet size or presence of secondary symptoms. To better understand the functional effect of genetic variation it is best to consider each individual disease by how it causes a reduction in platelet count. As mentioned previously, many disorders are as a result

of a disruption to the balance of platelet production and platelet consumption/senescence so they will be explained in this way and shown below.

Area of mutational effect	Gene	Protein	Associated disease	Inheritance pattern
Megakaryopoiesis	*ETV6*	Transcription factor ETV6	Thrombocytopenia 5 (THC5)	Autosomal dominant
	FLI1	Friend leukemia integration 1 transcription factor	Paris Trousseau type thrombocytopenia/Jacobsen (11q23 del)	Autosomal dominant/recessive
	FYB	Adhesion and degranulation-promoting adaptor protein	Novel thrombocytopenia	Autosomal recessive
	GATA1	Erythroid transcription factor	GATA1 related disease (X-linked thrombocytopenia [XLT] and X-linked thrombocytopenia with β-thalassemia [XLTT])	X-linked recessive
	GFI1B	Zinc finger protein Gfi-1b	Grey Platelet Syndrome + novel thrombocytopenia	Autosomal dominant
	HOXA11	Homeobox protein Hox-A11	Amegakaryocytic thrombocytopenia with radio-ulnar synostosis	Autosomal dominant
	MECOM	MDS1 and EVI1 complex locus		
	MPL	Thrombopoietin receptor	Congenital amegakaryocytic thrombocytopenia	Autosomal recessive
	NBEAL2	Neurobeachin-like protein 2	Grey Platelet Syndrome	Autosomal recessive
	RBM8A	RNA-binding protein 8A	Thrombocytopenia with absent radii	Autosomal recessive
	RUNX1	Runt-related transcription factor	Familial platelet disorder and predisposition to AML	Autosomal dominant
	THPO	Thrombopoietin	Mild novel thrombocytopenia (heterozygous)	Autosomaldominant
Proplatelet formation/platelet production	*ACTN1*	Alpha-actinin-1	Bleeding disorder, platelet-type 15	Autosomal dominant
	ANKRD26	N/A	ANKRD26-related thrombocytopenia	Autosomal dominant
	CYCS	Cytochrome C	CYCS-related	Autosomal dominant

			thrombocytopenia/THC4	
	DIAPH1	Rho-effector diaphanous-related formin 1	Novel macrothrombocytopenia with hearing loss	Autosomal dominant
	FLNA	Filamin A	Isolated macrothrombocytopenia	X-linked recessive
	GP1BA	Platelet glycoprotein 1b alpha chain	Bernard Soulier Syndrome + (Platelet type von-Willebrand disease – *GP1BA* only)	Autosomal dominant/recessive
	GP1BB	Platelet glycoprotein 1b beta chain		
	GP9	Platelet glycoprotein IX		
	ITGA2B	Integrin alpha-IIb	Glanzmann thrombasthenia	Autosomal recessive
	ITGB3	Integrin beta-3		
	MKL1	MKL/myocardin-like protein 1	Thrombocytopenia with immunodeficiency	Autosomal recessive
	MYH9	Myosin 9	MYH9 related disease	Autosomal dominant
	PRKACG	cAMP-dependant protein kinase catalytic subunit gamma	Bleeding disorder, platelet-type 19	Autosomal recessive
	SRC	SRC proto-oncogene, non-receptor tyrosine kinase	Novel thrombocytopenia with bone pathologies	Autosomal dominant
	TRPM7	Transient receptor potential melastatin-like 7 channel	Novel macrothrombocytopenia with atrial fibrillation	Autosomal dominant
	TUBB1	Tubulin beta-1 chain	TUBB1-related macrothrombocytopenia	Autosomal dominant
	WAS	Wiskott-Aldrich syndrome protein	Wiskott-Aldrich syndrome, X-linked thrombocytopenia (XLT)	X-linked recessive
Platelet clearance/other	*ABCG5*	ABC transporter G family member 5	Thrombocytopenia associated with sitosterolaemia	Autosomal recessive
	ABCG8	ABC transporter G family member 8		
	ADAMTS13	A disintegrin and	Thrombotic thrombocytopenia	Autosomal recessive

		metalloproteinase with thrombospondin motifs 13	purpura, Upshaw-Schulman syndrome	
	GNE	Glucosamine (UDP-N-Acetyl)-2-Epimerase	GNE related myopathy with congenital thrombocytopenia	Autosomal recessive
	STIM1	Stromal interaction molecule 1	Stormorken Syndrome	Autosomal dominant
	VWF	von Willebrand factor	von Willebrand disease type 2B	Autosomal dominant/recessive

Table 1.5.1 Current forms of inherited thrombocytopenia and their causative genes and inheritance patterns

1.5.1. Disorders of megakaryopoiesis

The commitment to differentiate and progress along the myelo-erythroid lineage in the formation of megakaryocytes is a highly controlled process as mentioned previously. The main levels of regulation are transcriptional and signalling related, and as such the myriad of diseases associated with a disruption in HSCs proliferation, differentiation and megakaryocyte maturation are often as a result of a disruption in these processes.

1.5.1.1. *THPO* – Mild novel thrombocytopenia (heterozygous)

As the main regulator behind megakaryopoiesis and the differentiation of HSCs to megakaryocytes, the presence of TPO, encoded by the gene *THPO*, is essential for platelet production. Recently, the first reported variant causative of IT was identified in the *THPO* gene (Dasouki, Rafi et al. 2013). Present in a large Micronesian family, with a history of death due to idiopathic aplastic anaemia, a rare frequency variant, c.112C>A; p.Arg38Cys, was identified by whole exome sequencing. Identified as R17C in the paper, the variant has only been identified once previously in the Exome Aggregation Consortium (ExAC) database (rs760797899) in a heterozygous state. The variant was homozygous in the proband suffering from aplastic anaemia, was absent in a related unaffected individual and, interestingly, was present in a heterozygous state in both parents and a non-anaemic sibling who suffered from a mild asymptomatic thrombocytopenia. Overexpression of the mutant TPO in the UT7-TPO cell line caused a 2.5 fold decrease in proliferation at low dosage suggesting that mutant TPO is defective in megakaryopoiesis (Dasouki, Rafi et al. 2013).

1.5.1.2. *c-Mpl* – Congenital amegakaryocytic thrombocytopenia

c-Mpl, the receptor for the thrombopoietin ligand, is expressed on the cell surface of HSCs and megakaryocyte progenitors promoting their differentiation. Its association with IT was first suggested through a lack of *c-Mpl* mRNA in bone marrow mononuclear cells and an elevated serum level of TPO in a patient suffering from congenital amegakaryocyte thrombocytopenia (CAMT) (MIM #604498) (Muraoka, Ishii et al. 1997). In a subsequent study, the genomic sequence of *c-Mpl* was analysed in a 10-year old CAMT patient that revealed the presence of compound heterozygous inheritance of a nonsense and frameshift causing variant; c.556C>T;p.Q186* and c.1499delT (Ihara, Ishii et al. 1999). CAMT is a rare disorder expressed in infancy associated by isolated thrombocytopenia but without secondary associated defects. This progresses throughout childhood to a complete pancytopenia and bone marrow failure usually by four years of age but can occur at a much faster rate (Stoddart, Connor et al. 2013). To date, genetic variations have been identified throughout the coding region of the *c-Mpl* gene, however the majority are present in exons 1-3 (van den Oudenrijn, Bruin et al. 2000). Clinically, CAMT presents as a sustained reduction in platelet count to severe levels ($<10x10^9$/L in whole blood), however a subset of patients have counts that fluctuate above $50x10^9$/L, often due to missense mutations that retain partial receptor function which leads to a delayed onset of bone marrow failure (King, Germeshausen et al. 2005; Germeshausen, Ballmaier et al. 2006). CAMT remains epidemiologically rare and the only feasible treatment method is HSC transplantation from related carrier or unrelated matched donors (Frangoul, Keates-Baleeiro et al. 2010).

1.5.1.3. *RBM8A* – Thrombocytopenia with absent radii

Thrombocytopenia with absent radii (TAR) is a syndromic disorder originally thought to be characterised by the inheritance of a 200kb microdeletion of chromosome 1q21.1 (Klopocki, Schulze et al. 2007). As with CAMT, it presents from birth with a severe thrombocytopenia, however it is also associated with skeletal abnormalities where infants have bilateral absent radii but retention of the thumb (Hall, Levin et al. 1969). Unlike CAMT however, thrombocytopenia usually improves with age over the first two years negating the need for drastic transplantation treatment. Recently TAR has been shown to be caused by the co-inheritance of a null allele, primarily the microdeletion on chromosome 1q21.1, and one of two rare single nucleotide polymorphisms (SNPs) within the *RBM8A* gene (Albers, Paul et al. 2012). The *RBM8A* gene encodes the Y14 protein, a subunit of the exon-junction complex which forms the active nonsense-mediated mRNA decay complex. The SNPs present within patients are either a low frequency SNP in the 5' untranslated region (5'UTR) or a SNP in the first exon of *RBM8A.* Both variants are located in active regulatory domains of the *RBM8A* gene and reduce promoter activity in a cell specific pattern. This causes inefficiency in a dose dependant manner with the co-inherited null allele to bring levels of Y14 below a critical functional threshold (Jan 2012). The two regulatory SNPs also reduce binding of the transcriptional repressor EVI1 which may lead to aberrant downstream transcriptional regulation through defective c-Mpl signalling although a lack of JAK2 phosphorylation is noted in TAR patients (Ballmaier, Schulze et al. 1998; Albers, Paul et al. 2012).

1.5.1.4. *HOXA11* and *MECOM* – Amegakaryocytic thrombocytopenia with radio-ulnar syntosis

HOXA11 is a DNA binding protein characterised by the 183bp long highly conserved homeobox domain. Like many other homeobox domain proteins, HOXA11 is expressed ubiquitously in human cord blood and haematopoietic precursor cells (Horvat-Switzer and Thompson 2006). First described in 2000 by Thompson and Nguyen, a single nucleotide deletion in the third helix of the homeobox domain leading to a frameshift and premature stop codon was associated with amegakaryocytic thrombocytopenia with radio-ulnar syntosis (Thompson and Nguyen 2000). The disorder, similar to TAR, presents with bone marrow failure, an initial thrombocytopenia leading to pancytopenia, and skeletal abnormalities. Amegakaryocytic thrombocytopenia with radio-ulnar syntosis functionally distinguishes itself from previous disorders due to fusion of radius and ulnar as the hallmark phenotype and lack of progressive improvement in platelet count over time (Thompson, Woodruff et al. 2001). The connection between how variants in *HOXA11* cause amegakaryocytic thrombocytopenia with radio-ulnar syntosis is yet to be fully established. Unlike wild type (WT) HOXA11, truncated HOXA11 lacks DNA binding capability in the presence of the TALE transcription factor, Meis1b (Horvat-Switzer and Thompson 2006). However, both WT and mutant forms of HOXA11 impair megakaryocyte differentiation and CD61 surface expression *in vitro* (Horvat-Switzer and Thompson 2006).

Recently the first variants to cause amegakaryocytic thrombocytopenia with radio-ulnar syntosis in the gene *MECOM* have been described (Niihori, Ouchi-Uchiyama et al. 2015). Three missense mutations were identified by whole exome sequencing in

patients without *HOXA11* mutations. The mutations are clustered around the C-terminal 8th zinc finger motif of the protein EVI1, encoded by *MECOM.* EVI1 is a transcriptional repressor that has been previously implicated to suppress expression of *RBM8A* in patients with TAR (Albers, Paul et al. 2012). It binds DNA through two motifs, a GATA-like motif and an E26-transformation-specific (ETS)-like motif located towards the C-terminus. In patients with *MECOM* mutants, this ETS-domain DNA binding ability is reduced. EVI1 also has the potential to bind a number of haemato-specific transcription factors indicating that the cause of amegakaryocytic thrombocytopenia with radio-ulnar synostosis may be related to the functional properties of EVI1 (Laricchia-Robbio, Fazzina et al. 2006; Senyuk, Sinha et al. 2007; Laricchia-Robbio, Premanand et al. 2009; Niihori, Ouchi-Uchiyama et al. 2015).

1.5.1.5. *GATA1* – GATA1 related disease (XLT and XLTT)

As mentioned previously, megakaryopoiesis is under high transcriptional regulation by a number of transcription factors. GATA1 which is expressed in a myelo-erythroid restrictive pattern defines the progression from HSCs to CMPs and subsequently MEPs. Mutations in *GATA1* therefore reflect the shared progenitors causing disorders in both thrombopoiesis and erythropoiesis including; X-linked thrombocytopenia (XLT) with dyserythropoiesis, X-linked thrombocytopenia with thalassemia (XLTT) and rarely congenital erythropoietic porphyria (Ciovacco, Raskind et al. 2008). GATA1 is structurally defined by the presence of zinc-finger DNA binding domains at both the N and C terminus with the majority of the variants leading to disease occurring in the N-terminal zinc-finger domain (N-f) which facilitates GATA1's cofactor and DNA binding capabilities. Variants are further refined spatially; with missense variants disrupting GATA1's ability to bind it's co-

factor FOG1 solely causing XLT and variants disrupting GATA1's DNA binding ability leading to XLTT (Nichols, Crispino et al. 2000; Yu, Niakan et al. 2002). Both disorders have phenotypic similarities; with defective megakaryocyte maturation and mature megakaryocyte dimorphism which leads to abnormal platelet size and shape, however platelet count tends to be milder in XLTT. Although inheritance is X-linked, a mild phenotype can occur in female heterozygotes due to skewed X-inactivation (Nichols, Crispino et al. 2000).

1.5.1.6. *RUNX1* – Familial platelet disorder and predisposition to AML

Familial platelet disorder with predisposition to acute myelogenous leukaemia (FPD/AML) (MIM #601399) is an autosomal dominant disorder characterised by a quantitative and qualitative platelet defect and an increased risk of AML. The molecular genetic cause of FPD/AML was first elucidated by linkage analysis to a region on chromosome 21q (Song, Sullivan et al. 1999). Contained within this region is the master regulator of haematopoiesis, *RUNX1*. To date, variants have been identified as the genetic cause of FPD/AML throughout the coding region of *RUNX1* but a clustering of variants within the Runt homology domain are most likely detrimental (Johnson, Lowe et al. 2016). FPD/AML shares phenotypic similarities with Jacobsen syndrome; platelet counts are often mild-moderate but variable between individuals with the same genetic aetiology of disease and a reduction in dense granule secretion is often observed as a secondary qualitative abnormality (Stockley, Morgan et al. 2013). The major clinical complication of this disorder, however, is in fact not the bleeding but the propensity to develop myelodysplasia or leukaemia (Walker, Stevens et al. 2002). The net effect of mutation is a

haploinsufficiency of RUNX1, or a potential dominant-negative effect, that disrupts the formation of complexes with core binding factor beta (CBFB) to disrupt the regulation of genes necessary for HSC maintenance, maturation and differentiation (Hart and Foroni 2002; Jalagadugula, Mao et al. 2010).

1.5.1.7. *FLI1* – Paris Trousseau type thrombocytopenia/Jacobsen (11q23 del)

Jacobsen syndrome refers to patients with a terminal deletion of the long arm of chromosome 11, del11q23. A large majority of patients with this syndromic disorder display a bleeding diathesis of reduced platelet number and a subpopulation of platelets that have an abnormal response to thrombin and contain giant α-granules (Breton-Gorius, Favier et al. 1995). This bleeding diathesis is known as Paris Trousseau type thrombocytopenia. The link between the hemizygotic loss of *FLI1* by chromosomal deletion was quickly established with this bleeding diathesis due to the strong role of FLI1 in haematopoiesis (Raslova, Komura et al. 2004). Variants have since been identified in *FLI1* in patients causing Paris Trousseau type thrombocytopenia without the presence of the 11q deletion (Stockley, Morgan et al. 2013). Mutations are predominantly missense and phenotypically platelets present similarly to those affected by *RUNX1* mutation as dense granule secretion is disrupted and expression of MYH10 is present as a biomarker of genetic variation (Antony-Debre, Bluteau et al. 2012; Stockley, Morgan et al. 2013). However, patients are also prone to secondary symptoms such as hearing loss and cataracts.

1.5.1.8. *ETV6* – Thrombocytopenia 5 (THC5)

Like FLI1 and SP1, ETV6 is a member of the ETS family of transcription factors and contains a consensus of three helixes and four beta sheets arranged in a DNA binding helix-loop-helix motif (Sharrocks 2001). Also like RUNX1 and FLI1, its involvement in the formation of myelodysplasia and leukaemia by translocation and chromosome segment deletion is well documented (Harrison, Moorman et al. 2005; Wall, Rayeroux et al. 2012). Recently however, its presence as a causative gene of thrombocytopenia has been identified through small scale whole exome sequencing of patients with unidentified thrombocytopenia and malignancy (Zhang, Churpek et al. 2015). These results have since been confirmed and further variants have been identified within the coding region clustered around the ETS domain. Variants result in aberrant ETV6 localisation, disruption of transcriptional repression and defective megakaryocyte maturation (Topka, Vijai et al. 2015; Melazzini, Palombo et al. 2016). As with FPD/AML, the most significant clinical complication is a predisposition to leukemic malignancy, however platelet count is also mildly affected but platelet size remains largely normal (Melazzini, Palombo et al. 2016).

1.5.1.9. *GFI1B* - Novel thrombocytopenia and Grey Platelet Syndrome

The transcriptional repressor GFI1B exerts its effect through epigenetic modification by histone recruitment. It is expressed, like many other transcription factors, early on during haematopoiesis in HSCs and the first stages of progenitor cells that retain multipotency (Moroy, Vassen et al. 2015). The main role of GFI1B, like the closely related GFI1, is mediated by binding DNA with zinc finger domains and recruiting

chromatin modulators using a 21-amino acid SNAG (Snail1/GFI) domain (Moroy, Vassen et al. 2015). Genetic variations tend to localise around zinc finger domains and lead to a phenotype similar to mutations in *GATA1* (Stevenson, Morel-Kopp et al. 2013). Indeed knock out mouse models of both genes are embryonic lethal due the arrest of differentiation of both megakaryocyte and erythroid lineages (Saleque, Cameron et al. 2002). The defining feature of *GFI1B* variants though is a reduction in platelet α-granules observable by electron microscopy. This phenotype explains the reduced expression of α-granule markers such as fibrinogen and P-selectin that is observed in patients (Stevenson, Morel-Kopp et al. 2013).

1.5.1.10. *NBEAL2* – Grey Platelet Syndrome

The term Grey Platelet Syndrome was originally used to describe recessive platelet based bleeding disorders with a complete absence of platelet α-granules giving platelets a unique grey appearance (Raccuglia 1971). As molecular genetics has progressed the genetic aetiology has been attributed to homozygous or compound heterozygote mutations in the Neurobeachin-like 2 gene (*NBEAL2*) (Albers, Cvejic et al. 2011). This is confirmed by antisense morpholino injection into one cell stage zebrafish embryos that phenocopies abrogation of thrombocyte formation with the presence of spontaneous bleeding (Albers, Cvejic et al. 2011). The functional consequence of how genetic variation can lead to a reduction in platelet count and an absence or severe reduction in α-granules and their associated content is yet to be fully established however. This does though suggest a role for BEACH domain proteins in granule packing and formation as other BEACH domain genes have also been previously associated with granule deficiency disorders (Nagle, Karim et al. 1996; Castermans, Volders et al. 2010).

1.5.1.11. *FYB* – Novel thrombocytopenia

Genetic variations in *FYB* were recently identified as the likely candidate causative mutations behind a recessive microthrombocytopenia in a large consanguineous family (Hamamy, Makrythanasis et al. 2014). To date only two variants have been identified; c.1385_1386del and c.393G>A, both are within large families of Middle Eastern origin (Hamamy, Makrythanasis et al. 2014; Levin, Koren et al. 2015). *FYB* encodes adhesion and degranulation-promoting adaptor protein (ADAP), a cytosolic adaptor protein expressed in platelets. It has an important role in platelet activation through inside-out $\alpha_{IIb}\beta_3$ signalling and collagen induced activation through $\alpha_2\beta_1$ (Kasirer-Friede, Moran et al. 2007; Jarvis, Bihan et al. 2012). Genetic variation therefore affects platelet activation, but interesting platelets appear to be pre-activated due to overexpression of P-selectin and PAC-1. Thrombocytopenia is thought to be as a result of a combination of increased clearance of pre-activated platelets and dysmegakaryopoiesis caused by actin dysregulation (Hamamy, Makrythanasis et al. 2014).

1.5.2. Disorders affecting proplatelet formation/platelet release

Most commonly genes that fall within this category function to control the cytoskeleton rearrangement process involved with thrombopoiesis. In the process of platelet production by proplatelet extension, the tight regulation of actin and tubulin polymerisation is crucial to ensure normal platelet release.

1.5.2.1. *TUBB1* – TUBB1-related macrothrombocytopenia

As mentioned previously the main process regarding proplatelet extension is the polymerisation and subsequent sliding of microtubules. β1-tubulin, encoded by

TUBB1, is the main isoform within human megakaryocytes and its expression underpins this process. Genetic variants in *TUBB1* were first identified to have a positive impact on the reduction of cardiovascular risk (Freson, De Vos et al. 2005). Individuals with a p.Pro43Gln variant displayed reduced ATP secretion but were less likely to contract cardiovascular disease. Although impairing the function of platelets, this variant did not affect their number. It was Kunishima *et al,* who first identified the causative heterozygous variant, p.Arg318Trp, in a patient with congenital macrothrombocytopenia (Kunishima, Kobayashi et al. 2009). Mutated β1-tubulin is expressed but is unstable, forming punctate aggregates that do not form heterodimers with α-tubulin and are not included in microtubule assembly (Kunishima, Nishimura et al. 2014). This leads to a diminishment in proplatelet production with fewer proplatelet extensions and larger swellings around proplatelet tips, thought to allow the shedding of few non-uniform sized platelets into the circulation. The effect of TUBB1 on the formation of platelets is also noted through the presence of SNP markers of disease severity in alternate disorders. The p.Arg307His SNP is present in a large number of patients with Bernard-Souiler Syndrome (BSS) and is associated with an increased bleeding risk (Basciano, Matakas et al. 2015).

1.5.2.2. *WAS* – Wiskott-Aldrich Syndrome, X-linked thrombocytopenia

Like β1-tubulin, F-actin is present throughout platelets and allows proplatelet branching. This actin polymerisation process is promoted by nucleation factors such as WASp, encoded by *WAS,* which is selectively expressed in haematopoietic cells and also aids in the formation of the IMS through the WASp/WAVE pathway

(Schulze, Korpal et al. 2006). WASp is a multifaceted, multi-domain protein associated with several clinical disorders. Genetic variations follow a genotype-phenotype correlation dependant on the functional consequence and position of genetic alteration (Zhu, Watanabe et al. 1997; Lutskiy, Rosen et al. 2005). Nonsense mutations, frameshift causing insertions/deletions and complex mutations that lead to a null allele often cause the more phenotypically severe Wiskott-Aldrich syndrome (WAS) (Wengler, Notarangelo et al. 1995). This X-linked thrombopathy first described in 1937 by Alfred Wiskott and later reported by Robert Aldrich (Wiskott 1937; Aldrich, Steinberg et al. 1954) is classified by microthrombocytopenia, eczema and proneness to infection characterised by defects in T and B cell function. Life expectancy is usually less than 10 years and patients are prone to succumb to haematological malignancies. On the other hand, mostly missense variants, that occur within the first four coding exons of *WAS* which encodes the EVH1/WH1 domain, cause a slightly less severe bleeding diathesis known as X-linked thrombocytopenia or THC1 (Villa, Notarangelo et al. 1995). EVH1/WH1 is the binding site of the WASp-interacting protein (WIP) which, when constitutively bound, maintains the protein in a stable inactive state (Ramesh, Anton et al. 1997; Lim, Misra et al. 2007). Loss of this binding results in early degradation of WASp. This reduces effective WASp levels but function is still retained negating the lymphoid abnormalities associated with WAS.

1.5.2.3. *MKL1* – Immunodeficiency with thrombocytopenia

Myocardin-like protein 1, encoded by *MKL1*, is a recently reported novel gene causative of primary immunodeficiency with a reduction in platelet count (Record, Malinova et al. 2015). The production of F-actin was shown to be affected by a

homozygous recessive mutation and the primary phenotype was a severe immunodeficiency but a variable mild to moderate thrombocytopenia was also observed (platelet counts 50-150x10^9/L). MKL1 elucidates an effect on the actin cytoskeleton through its transcriptional co-factor, serum response factor (SRF) and knock-out (KO) of either MKL1 or SRF in mice phenocopies thrombocytopenia and megakaryocyte maturation and migration dysfunction (Halene, Gao et al. 2010; Smith, Thon et al. 2012).

1.5.2.4. *MYH9* – MYH9-related disease

Myosins, in particular IIA and IIB, are also involved in proplatelet formation by interacting with actin filaments generating contractile forces (Pertuy, Eckly et al. 2014). Mutations in *MYH9,* which is translated into the heavy chain of non-muscle myosin IIA (NMMMHC-IIA), affect the most proximal part of platelet production in the budding of preplatelets and subsequent separation into platelets. A disruption in the Rho-Rho kinase-myosin IIA pathway leads to an absence of NMMHC-IIA reactivation which is needed for the induced fragmentation of proplatelets (Chen, Naveiras et al. 2007; Chen, Boukour et al. 2013; Spinler, Shin et al. 2015). This leads to characteristic large platelets and a reduction in circulating platelet count both of which are present from birth. MYH9-related disease brings together the previously classified May-Hegglin syndrome, Sebastian syndrome, Fechtner syndrome and Epstein syndrome. Although phenotypically different, particularly in the presence of myosin aggregates known as Döhle-like body leukocyte inclusions, they are all derived from mutations in *MYH9* (Seri, Pecci et al. 2003). Like with variations in *WAS,* a genotype-phenotype correlation is observed with mutations in the ATPase

domain more frequently associated with nephropathy and/or hearing loss as well as the haematological phenotype (Dong, Li et al. 2005).

1.5.2.5. *TRPM7* – Novel macrothrombocytopenia with atrial fibrillation

Mg^{2+} causes an inhibitory effect on platelets and this role has been successfully utilised in the treatment of coronary artery disease (Ravn, Kristensen et al. 1996; Shechter 2003). Transient receptor potential melastatin-like 7 channel (TRPM7) is a ubiquitously expressed cation channel with a fused alpha kinase domain. TRPM7 is permeable to extracellular Mg^{2+} and negatively gated by intracellular Mg^{2+}, and as such PF4 specific deletion of TRMP7 is rescued by Mg^{2} supplementation (Paravicini, Chubanov et al. 2012; Stritt, Nurden et al. 2016). Impaired channel function causes macrothrombocytopenia in mice and this is phenocopied in patients with two missense variants located outside of the α-kinase domain. The reduction in platelet count is caused by cytoskeletal alterations resulting in impaired proplatelet formation, this may be the result of perturbed NMMHC-IIA by deregulated Mg^{2+} homeostasis (Stritt, Nurden et al. 2016).

1.5.2.6. *ACTN1* – Bleeding disorder, platelet type-15

α-actinin, encoded by the *ACTN1* gene, is a member of the actin-crosslinking family that facilitates actin cytoskeleton rearrangement. Along with *ACTN4,* it is expressed in platelets and megakaryocytes, distinguishing itself from the muscle specific isoforms (Haudek, Slany et al. 2009). It exerts its effect through an actin binding domain (ABD), binding filamentous actin and allowing bundling which is believed to be important for cytoskeleton remodelling upon proplatelet formation. Variants, that

are mostly localised within this ABD or the C-terminally located calmodulin-like domain, have a gain of function effect, increasing actin bundling (Gueguen, Rouault et al. 2013; Kunishima, Okuno et al. 2013; Bottega, Marconi et al. 2015; Murphy, Lindsay et al. 2016). This increase in bundling is without regulation, leading to a disorganised actin-based cytoskeleton which is not circumferential and parallel to the cell periphery in foetal-liver derived mouse megakaryocytes (Kunishima, Okuno et al. 2013). The platelets that are produced in patients with bleeding disorder, platelet type-15 are often increased in size and mildly decreased in quantity (Bottega, Marconi et al. 2015).

1.5.2.7. DIAPH1 – Novel macrothrombocytopenia with hearing loss

DIAPH1, which encodes the diaphanous related formin 1 protein, regulates microtubule stability and promotes actin assembly by forming a homodimer (Goode and Eck 2007; Chesarone, DuPage et al. 2010). DIAPH1 aids in the process of cytoskeleton remodelling upon competitive binding of Rho GTPases to a diaphanous auto-regulatory domain (DAD). The DAD domain, located near the carboxyl terminal, is usually bound to a diaphanous inhibitory domain (DID) inhibiting its activity. Two unrelated individuals were identified with a truncating nonsense c.3637C>T; p.Arg1213* transition that results in the loss of the RRKR motif in the DAD domain, rendering the protein in an inhibited state (Stritt, Nurden et al. 2016). A disruption in the process of DIAPH1 mediated cytoskeleton remodelling leads to a defect in proplatelet formation, similar to variants within *ACTN1.* Megakaryocytes also form heterogeneous clusters and patients show a mild thrombocytopenia with an increase in platelet volume characteristic with a defect in cytoskeleton remodelling.

1.5.2.8. *FLNA* – Isolated macrothrombocytopenia

In platelets, filamin A acts as a cellular tether, connecting the underlying actin cytoskeleton, by a N-terminal ABD, to principal plasma membrane receptors such as GP1ba and the integrin $\alpha_{IIb}\beta_3$ (Falet 2013). This tether allows platelet-vessel wall interactions to mediate cytoskeleton remodelling upon activation and subsequent downstream signalling. Filamin A, encoded by *FLNA,* has previously been known to be the genetic cause of dominant X-linked bilateral periventricular nodular heterotopia (BPNH) (Sole, Coupry et al. 2009). Many FLNA related BPNH patients often present with haematological complications including haemorrhaging and thrombocytopenia. Due to FLNAs involvement in platelet function an unaffected related family member of two patients suffering from BPNH was sequenced due to presence of lifelong thrombocytopenia ($80x10^9$/L) and a history of menorrhagia (Nurden, Debili et al. 2011). A heterozygous point mutation, c.5407G>A, p.Glu1803Lys, was detected in a highly conserved region. X-inactivation was random but irregular megakaryocyte differentiation and a gradient FLNA expression pattern was noted in *in vitro* megakaryocyte cultures, supporting mutations in FLNA as the potential cause of a novel macrothrombocytopenia.

1.5.2.9. *PRKACG* – Bleeding disorder, platelet type-19

FLNAs role, although not fully understood carries importance as its proteolysis is tightly regulated through phosphorylation. To avoid targeted degradation FLNA is phosphorylated at p.Ser2152 by cAMP-dependant protein kinase A which is comprised of two catalytic subunits, the y-isoform of which is encoded by *PRKACG* (Jay, Garcia et al. 2000). To date, only one variant has only been found encoding *PRKACG* by whole exome sequencing of a single pedigree (Manchev, Hilpert et al.

2014). Paediatric patients displayed a severe thrombocytopenic count, requiring transfusion for life threating bleeding episodes. Platelets were also giant and megakaryocyte clusters were observed in bone marrow aspirations. Megakaryocytes derived from CD34+ patient progenitors also displayed rapid degradation of FLNA and a reduction in the percentage of proplatelet-bearing megakaryocytes.

1.5.2.10. *SRC* – Novel thrombocytopenia with bone pathologies

The SRC proto-oncogene, non-receptor tyrosine kinase, encoded by *SRC*, is a plausible pharmaceutical target due to the presence of activating mutations within several human cancers (Irby, Mao et al. 1999). SRC protein levels are high within platelets, and SRC-dependent phosphorylation plays an important role in platelet signalling pathways downstream from cell surface receptors (Golden, Nemeth et al. 1986; Senis, Mazharian et al. 2014). So far a single heterozygous variation has been identified in a highly conserved residue within the C-terminal tail of the SRC kinase domain (Turro, Greene et al. 2016). The variant was identified using a unique approach comprising human phenotype ontology-term based filtering with whole genome sequencing (WGS). The mutated p.Glu527Lys allele predominates in patient platelets leading to elevated phosphorylation. Patients show a hyper-cellular bone marrow and patient-derived megakaryocytes show a reduced number of proplatelet extensions leading to a reduction in platelet number, a variable platelet size and alpha granule content.

1.5.2.11. *ANKRD26* – ANKRD26-related thrombocytopenia/THC2

ANKRD26-related thrombocytopenia or Thrombocytopenia 2 (THC2) is a disorder unique due to its genetic cause. Unlike all other forms of IT, THC2 is not as a result

of disruptive mutations with the coding region of *ANKRD26,* instead mutations lie within a tight cluster within the 5'-UTR of the gene (Pippucci, Savoia et al. 2011). The majority of mutations lie within a 22 nucleotide stretch from c-113 to c-134 disrupting binding of RUNX1 and FLI1 which normally functions to silence expression during late megakaryopoiesis (Noris, Perrotta et al. 2011; Bluteau, Balduini et al. 2014). Prolonged expression of ANKRD26 causes increased circulating TPO and an increase in MPL-related TPO signalling via the MAPK pathway, leads to disrupted proplatelet formation (Bluteau, Balduini et al. 2014). The typical clinical presentation is a moderate reduction in platelet count, decreased platelet α-granules and a reduction in surface GP1a. Much like variants in *RUNX1* and *FLI1,* there is also an increased propensity to developing myeloid malignancies (Noris, Favier et al. 2013). Recently the first missense mutation in *ANRKD26,* c.473A>G that segregates with disease and shares phenotypic similarities to THC2 patients was also reported (Al Daama, Housawi et al. 2013).

1.5.2.12. *GP1BA, GP1BB* and *GP9* – Bernard-Soulier Syndrome

The products of four distinct genes (*GP1BA, GP1BB, GPIX and GPV*) assemble within the maturing megakaryocyte in a 2:2:2:1 stoichiometry to form the platelet membrane receptor GPIb/IX/V (Berndt, Shen et al. 2001). During haemostasis, the GP1b/IX/V receptor complex plays a critical role through binding of VWF to arrest platelets and form a thrombus. In BSS, surface expression is decreased or absent, or the function can be reduced (Lopez, Andrews et al. 1998). Although phenotypically characterised for over 60 years, genetic variations were first identified in the *GP1BA* gene just over 25 years ago (Bernard 1948; Ware, Russell et al. 1990). To date, genetic variations have been found to be the causative defect of BSS in three genes

of the GP1b/IX/V complex; *GP1BA, GP1BB* and *GP9*. Inheritance can be recessive or autosomal dominant causing two variably different forms of BSS, biallelic and monoallelic, respectively. Both diseases are characterised by a reduction in platelet count and the presence of giant platelets, often larger than red blood cells (Savoia, Kunishima et al. 2014). Monoallelic BSS is often limited to this phenotype whereas biallelic recessive BSS often causes a complete lack of response to sub-endothelial VWF which can be measured by *in vitro* aggregation response to ristocetin (RIPA) (Savoia, Kunishima et al. 2014). Outside of its role in haemostasis, the GP1b/IX/V complex also has an important role as a structural component of the cytoskeleton by binding FLNA, and the cytoskeleton signalling proteins calmodulin, 14-3-3ζ protein or the p85 subunit of phosphoinositide 3-kinase (Nurden, Debili et al. 2011; Andrews and Berndt 2013). This interaction aids in the relay of extracellular signals in proplatelet formation.

1.5.2.13. *ITGA2B* and *ITGB3* – Glanzmann Thrombasthenia

Another platelet surface receptor with a major role in haemostasis that is affected by genetic variation is the integrin GPIIb-IIIa/$\alpha_{IIb}\beta_3$, the receptor for fibrinogen. Variants in the two encoding genes, *ITGA2B* and *ITGB3,* lead to qualitative deficiencies in the receptor causing the disorder known as Glanzmann thrombasthenia (GT) (Nurden, Fiore et al. 2011). Although still considered rare, GT is one of the more common bleeding diathesis (George, Caen et al. 1990). As with BSS, the main phenotypical attribute is mucocutaneous bleeding arising from a lack of platelet aggregation, however presentation is variable and often platelet count and size are unaffected. A number of patients have been reported however with heterozygous variants within either gene that leads to a reduction in platelet count and a marked increase in

platelet size (Peyruchaud, Nurden et al. 1998; Ghevaert, Salsmann et al. 2008). Platelet count is thought to be affected by aberrant proplatelet formation due to gain of function mutations leading to a partially activated receptor, however dominant-negative variants have been reported in association with macrothrombocytopenia as well (Kunishima, Kashiwagi et al. 2011; Nurden, Fiore et al. 2011).

1.5.2.14. *CYCS* – CYCS-related thrombocytopenia/THC4

One gene whose involvement in platelet production currently, through the intrinsic apoptosis pathway, is controversial is Cytochrome C (*CYCS*). CYCS initiates apoptosis through the activation of caspase 9 and subsequently caspase 3, a dysregulation of which was originally suggested as a feature of platelet production (Jiang and Wang 2004; Morison, Cramer Borde et al. 2008). To date, two variants (p.Tyr49His [p.48 in original publication] and p.Gly42Ser [p.41 in original publication]) have been observed (Morison, Cramer Borde et al. 2008; De Rocco, Cerqua et al. 2014). Intramedullary naked megakaryocyte nuclei, indicative of dysregulated megakaryopoiesis and premature platelet release, were identified in bone marrow aspirates. Furthermore, mouse lung fibroblasts transduced with *Cycs* variants show increased apoptotic activity in response to staurosporine. It was therefore first believed that apoptosis had a fundamental role in the timing of pro-platelet release, hypothesising that inhibition of actin polymerisation, a process required for appropriate proplatelet branching, activates the apoptotic pathway (Avanzi, Izak et al. 2015). However, platelet production can proceed regardless of both the intrinsic and extrinsic apoptotic pathways (Josefsson, Burnett et al. 2014).

1.5.3. Disorders of platelet clearance/unknown functional disorders

The fundamental processes governing the production of platelets are easily affected by deleterious mutations leading to thrombocytopenia. On the contrary, the main process regarding the intrinsic control of platelet lifespan and clearance is relatively untouched by genetic variation. There are two potential reasons that may explain this; that variants within the main pro-survival gene *BCL2L1* are not viable or that they are considerably rare and even though the field of haemostatic genetics has progressed they have just not been identified yet. However, a number of genes with congenital mutations do exist that have a unique effect on platelet clearance by pre-activating platelets. In addition, a number of genes have recently been identified where the functional effect of genetic variation has yet to be established.

1.5.3.1. *ADAMTS13* – Thrombotic thrombocytopenia purpura, Upshaw-Schulman Syndrome

Variants in the metalloproteinase encoding gene *ADAMTS13* are the genetic cause of a rare form of life-threatening systemic IT known as Upshaw-Schulman syndrome, or alternatively known as thrombotic thrombocytopenia purpura (TTP) (Levy, Nichols et al. 2001; Schiff, Roberts et al. 2004). Endogenous circulating ADAMTS13 cleaves VWF at the p.Tyr1605=Met1606 peptide allowing for the removal of circulating VWF from haemostatic vessels avoiding unwarranted platelet activation. In TTP, genetic variation renders the metalloproteinase function of ADAMTS13 inactive allowing for the accumulation of large hyperactive VWF multimers. These have the capacity to trigger platelet aggregation causing subsequent removal within the spleen.

Phenotypically this disorder presents with a reduction in platelet count due to this removal but also haemolytic anaemia, neurological symptoms and renal dysfunction which can threaten a patient's life (Levy, Nichols et al. 2001).

1.5.3.2. *VWF* – Von Willebrand disease type 2B

The physical importance of VWF is known throughout haemostasis. Variants within the encoding *VWF* gene lead to functional deficiency causing von Willebrand disease (VWD). VWD encompasses a myriad of disorders, most of which do not affect the platelet count; however, type 2B does lead to thrombocytopenia. Causative variants of type 2B lead to a gain of function and an increased affinity of VWF for GP1b on the platelet surface (Othman 2011). Continuous interaction between the receptor and ligand disrupts megakaryopoiesis leading to the production of giant platelets (Nurden, Debili et al. 2006). In addition, large multimers of VWF that bind platelets are present in the circulation in a direct comparison to TTP (Nurden, Debili et al. 2006). In VWF knock-in mice these complexes are taken up efficiently by macrophages in the liver and spleen indicating an increased clearance as well as disruption to production (Casari, Du et al. 2013).

1.5.3.3. *STIM1* – Stormorken-York Syndrome

Increased platelet clearance is also the proposed mechanism behind the thrombocytopenia observed in Stormorken-York syndrome. Caused by gain of function variants in *STIM1,* patients present with a reduction in platelet count due to pre-activation and subsequent removal within the spleen (Misceo, Holmgren et al. 2014; Markello, Chen et al. 2015). The process of calcium uptake enables platelet activation. Upon sensing extracellular Ca(2+), STIM1 undergoes a conformational

change allowing it to interact and open ORAI1, a Ca(2+) release-activated Ca(2+) pore. Genetic variants within STIM1 lead to a gain of function, causing a constitutively active calcium channel and subsequent pre-activated platelet state.

1.5.3.4. *ABCG5* and *ABCG8* – Thrombocytopenia with sitosterolemia

Thrombocytopenia as a secondary symptom of sitosterolemia is a unique situation where a role in the normal balance of platelet production/platelet clearance is not affected. Sitosterolemia is a rare, autosomal recessive disorder caused by mutations in two adjacent ATP-binding cassette transport genes; *ABCG5* and *ABCG8.* Both encoded proteins usually function to remove dietary sterols from the circulation by acting as a heterodimeric efflux pump. When mutated, there is a decrease in secretion of dietary sterols which leads to an accumulation in the circulation. Heterodimeric efflux pumps for sterols are not present on blood cells or platelets and therefore accumulated high levels are toxic, resulting in platelet death and subsequent thrombocytopenia. Interestingly, in ABCG5/ABCG8 KO mice, which are fed on high-plant-sterol diets, their platelets are also noted to be hyper-active and show impaired function.

1.5.3.5. *GNE* – GNE-related myopathy with thrombocytopenia

Like thrombocytopenia with sitosterolemia, variants in *GNE* have been noted to cause a disorder with a secondary symptom of thrombocytopenia. To date, two compound heterozygous variations have been reported in the gene encoding glucosamine (UDP-N-Acetyl)-2-epimerase (GNE) (Izumi, Niihori et al. 2014; Zhen, Guo et al. 2014). GNE functions as an enzyme in the sialic acid biosynthesis

pathway and is expressed within all cells of the haematopoietic lineage. Previous mutations have been associated with either sialuria or hereditary inclusion body myopathy, however a defining feature of severe platelet counts less than $45x10^9$/L was identified in two unrelated pedigrees (Seppala, Lehto et al. 1999; Eisenberg, Avidan et al. 2001). All patients also presented with severe body myopathy as a primary symptom but patients did not display signs of myopathy until mid-adolescence/early adulthood. The link between how variants in *GNE* can cause a reduction in platelet count is yet to be established.

1.6. Genetic studies into inherited thrombocytopenia

Due to the relative heterogeneity of ITs and the lack of definitive gold-standard platelet function test, DNA-based approaches have been vital in the improvement of clinical diagnosis. However, genetic studies have thus far been mainly focused around single disorders, utilising targeted sequencing mainly around an area of tight linkage to disease. DNA sequencing has revolutionised medical research since the invention of Sanger sequencing and the polymerase chain reaction (PCR) (Sanger, Nicklen et al. 1977; Mullis, Faloona et al. 1986). Traditional methods are now being replaced by massively parallel next generation sequencing of whole exomes and genomes, but their use in IT diagnosis is still limited and not without challenges.

1.6.1. Pre-genomic era

Many aforementioned disorders were often strongly phenotypically and clinically classified before the progression of genetic diagnostics became the clinical norm. This is indeed true for the three of the most prevalent disorders; Bernard-Soulier syndrome, Glanzmann Thrombasthenia and MYH9-related disorders. All disorders

were originally classified due to the presence of secondary symptoms in addition to a reduction in platelet count (George, Caen et al. 1990; Seri, Pecci et al. 2003; Andrews and Berndt 2013). Platelet function testing, mostly in the form of LTA, also played a strong role in the determination of disease. This allowed disorders to become characterised and distinguished, allowing for rapid genetic analysis later on in a number of individuals with a shared phenotype. A diagnostic approach like this follows a reverse genetics methodology, analysing the functional consequence and subsequently determining the genetic cause. It has allowed for a complete diagnosis of disorders and a move towards a situation where they are now routinely tested for in haematology labs throughout the UK. The tests utilise phenotypic assays; blood smears, aggregometry and flow cytometry, often in combination with single gene sequencing for efficient and accurate diagnosis.

1.6.2. Post-genomic era

As the field of molecular genetics has progressed, a number of studies have moved away from targeted sequencing of an area of disease linkage. More commonplace is utilising whole exome and whole genome sequencing when a patient's genotype is not known. To date though this has largely been contained to small scale studies of single individuals or a small pedigree with a shared phenotype. While effective, and responsible for the discovery of a number of recent causative genes of IT, it is still often applied to patients with a strong phenotype. On the contrary, few studies have focused around the methodology of forward genetics; utilising genetic searching to find a variant and subsequently determining the effect of the variant and how it might be linked to the disease phenotype.

Panel sequencing refers to a strategy of designing singular baits or probes to a known region of interest allowing for parallel amplification and sequencing of multiple genes. It builds upon the fundamental methodology of Sanger sequencing but often now utilising next generation sequencing technology. This strategy has been adopted by a number of large scale studies that have begun into the genetic study of IT. Many are global and are pushing the boundaries of our knowledge but they are not without limitation. One of the largest studies into the genetics of IT utilised a combination of targeted panel sequencing with single gene Sanger sequencing to compile the results from 500 individuals (Balduini, Pecci et al. 2012). The study highlighted the correct prevalence of a number of forms of IT suggesting they are more common than previously thought. The most defining feature of the study though, was that 51% of index cases were without a genetic cause of disease. Over the past four years, since this study, a number of new forms of IT have been identified suggesting that this number would likely diminish. However, targeted sequencing is restrained by only being able to identify variants in genes that were previously known to cause disease.

To combat this, a number of groups worldwide have started studies using the potential lack of restriction benefits of whole exome and whole genome sequencing to study IT. Each group has a unique approach to the problem and many form sections of larger consortiums looking at the molecular genetic effects of bleeding diathesis as a whole. Whole exome sequencing (WES) is becoming increasingly popular due to decreasing costs, improvements in efficiency and in ease of data handling (Warr, Robert et al. 2015). Since the first report of selectively sequencing all exomes successfully, WES has progressed into new capture methods, sample library

preparations and parallel sequencing techniques, but all retain the fundamental goal of sequencing all coding exons (Ng, Turner et al. 2009).

The UK-GAPP study, which this thesis forms part of, takes a novel approach. The study aims to categorise completely a patient's disease through a combination of extensive platelet phenotyping and broad whole exome sequencing to determine their genotype.

The BRIDGE Consortium-Bleeding and Platelet Disorders (BPDs) study (B20) takes an alternative approach to genetic diagnosis in patients with a broad range of platelet and non-platelet based bleeding disorders. By combining whole genome and exome sequencing with annotation with Human Phenotype Ontology (HPO) terms they aim to characterise novel genetic variants in a large patient cohort (Robinson, Kohler et al. 2008; Maclachlan, Watson et al. 2016). Similarly, groups in Italy and the USA are progressing with projects incorporating next generation sequencing into the study of IT as well.

1.7. Thesis Hypothesis and Aims

By combining extensive phenotyping with next generation sequencing in the form of WES I hypothesise that we will determine novel genetic variants in previously undiagnosed patients with IT; allowing for an expansion of our current knowledge of the fields surrounding platelet production, lifespan, function and death.

The aims of this thesis are to build upon previous work within the UK-GAPP study specifically looking at the molecular genetics behind IT. As no large scale WES studies have been applied to this disease subtype before, the work will be novel in its approach. Through this work I aim to try to determine novel genetic causes of disease and try to determine how genetic variation can lead to a fundamental loss of function in platelet formation, function or lifespan. The overall aim of my thesis is therefore;

- To elucidate novel causative mutations in patients with inherited thrombocytopenia.

I will approach this in a number of ways, including:

- Develop and use methodology to fully determine a patient's phenotype and platelet function.
- Utilise whole exome sequencing to determine novel candidate mutations in patients with unknown genetic aetiology of disease.
- Validate the effect of genetic variation and how detrimental mutation can cause a patient's bleeding diathesis.

2. Materials and Methods

2.1. Materials

All general lab consumables and plastic ware were purchased from StarLab (StarLab, UK). Unless otherwise stated all materials and reagents were available fro m Sigma Aldrich (Sigma Aldrich, UK).

Material/Software	**Supplier**
4% citrate concentrated solution	Sigma Aldrich, UK
AggroLink	ChronoLog/LabMedics, UK
Agilent SureCall v3.5	Agilent Technologies, UK
Agilent SureDesign v3.5.4	Agilent Technologies, UK
Chromas	Technelysium Pty Ltd, Australia
Citrate blood collection tubes	BD, UK
ExonPrimer	Helmholtz Center Munich, Germany
Human Protein Atlas	Uppsala Universitet, Sweden
Human Splicing Finder	Aix Mareseille Universite, France
IGV	Broad Institute, USA
MIT CRISPR database	MIT, USA
Mutation Surveyor	SoftGenetics, USA
Mutation Taster	Charite, Berlin
Novoalign	Novocraft technologies, Malaysia
Polymorphism Phenotyping v2	Harvard, USA
Primer3	Open Source
Protein Variation Effect Analyser	J.Craig Venter Institute, USA
Shandon™ColorFrost™ Plus microscope slides	ThermoFisher, UK
Siliconized glass cuvettes	LabMedics, UK
Sorting Intolerant From Tolerant	J.Craig Venter Institute, USA
Spray-coated KDEDTA blood collection tubes	BD, UK
Sysmex XN1000 reagents (all)	Sysmex, UK

Table 2.1.1 Materials and software used throughout this project and their suppliers.

Database	Website	Last accessed
1000G project	www.internationalgenome.org/	Oct-16
ClinVar	www.ncbi.nlm.nih.gov/clinvar/	Oct-16
dbSNP139	www.ncbi.nlm.nih.gov/projects/SNP/	Oct-16
EVS	evs.gs.washington.edu/EVS/	Oct-16
ExAC	exac.broadinstitute.org/	Oct-16
HGMD	www.hgmd.cf.ac.uk/ac/index.php	Oct-16

Table 2.1.2 Databases used throughout this project.

2.2. Study approval

The UK-GAPP study was approved by the National Research Ethics Service Committee of West Midlands–Edgbaston (REC reference: 06/MRE07/36) and participants gave written informed consent in accordance with the Declaration of Helsinki. This study was registered at www.isrctn.org as #ISRCTN 77951167. The UK-GAPP study is also included in the National Institute of Health Research Non-Malignant Haematology study portfolio (ID9858).

2.3. Patient recruitment

All patients included in the UK-GAPP study were recruited either through regional Comprehensive Care Haemophilia centres across the UK or by referring haematologists specialising in bleeding disorders. Patients were recruited following a clinical history and subsequent diagnosis of a suspected platelet-based bleeding disorder. Patients were identified as suffering from inherited thrombocytopenia after whole blood platelet counts were found to be below $150x10^9$/L in routine haematological analysis. Cases were included within the study that ranged between $150\text{-}200x10^9$/L if it was deemed that there was a strong family history of shared phenotypic disease between affected individuals or if previous platelet counts were below $150x10^9$/L prior to enrolment. Genetic pedigrees were taken when available. Selection for recruitment to the UK GAPP study was dependent upon meeting specific criteria as set out by the study.

Patients were invited to participate in the study if they satisfied all of the inclusion criteria:

a. Aged 0-85 years old
b. Abnormal bleeding symptoms compatible with a platelet function disorder
c. Results from coagulation factor tests all within local laboratory reference intervals
d. Absence of demonstrable acquired platelet dysfunction.

Patients were excluded from the study if they;

a. Had recently undergone major surgery
b. Had chronic renal failure requiring dialysis
c. Had severe anaemia (Hb<8g/dl)
d. Had recently had blood/platelet transfusion or taking medications known to affect platelet function.

Patients were also excluded if they have a known form of inherited thrombocytopenia. Patients were routinely tested for the presence of BSS and MYH9-related disease upon suspicion of an inherited platelet based bleeding disorder as standard clinical practice. Clinical testing varied dependant on the referring haematological centre but comprised of initial analysis by blood film to determine the presence of giant platelets and leukocyte inclusion bodies in all cases. Results were then confirmed by genetic analysis if available. Patients were also excluded if bleeding was thought to be as a result of ITP (a highly variable platelet count was observed over a sustained period) or other non-platelet based bleeding disorders including von Willebrand disease (excluding type 2B) and inherited coagulation factor deficiencies.

2.4. Healthy controls

Healthy donor volunteers were at least 18 years of age and considered healthy if they did not have a history of bleeding symptoms or did not require long-term medical therapy. They were also only included within the study if they had refrained from drugs known to influence platelet function in the previous two weeks and had not donated blood in at least that same time period.

2.5. Sample collection

Sample collection varies by age of consenting patient. In all cases, blood was collected simultaneously in both patient and control at the site of patient sample collection if possible. If not possible, the control sample was collected at the site of analysis in a similar time scale to the patient sample. One control sample was used for multiple patient samples collected on the same day at the same collection site. Blood was transported from all centres at ambient temperature (approximately 20°C) by courier to the laboratory in Birmingham. The UK-GAPP study has previously shown that the results in PRP in samples transported in this way are indistinguishable from those locally collected (Dawood, Wilde et al. 2007).

For adults, fifty millilitres of blood in 3.2% trisodium citrate was collected by venepuncture. For paediatric patients (<18 years), up to 20ml, dependant on the age of the consenting patient, of blood in 3.2% trisodium citrate was collected by venepuncture. In most cases, blood was collected into BD Vacutainer® 2.7mL plastic citrate blood collection tubes (BD, UK, #363083), however, collection tubes did vary occasionally due to specific protocols at each referring haematological centre. Control samples that were collected in house at the site of analysis were collected by

venepuncture into a 1:9 ratio with 4% citrate concentrated solution (Sigma-Aldrich, UK, #S5770).

For all patients, an additional 2-4mL of ethylenediaminetetraacetic acid disodium salt solution (EDTA) anti coagulated blood was collected for whole blood cell counts. Blood was collected by venepuncture into BD Vacutainer® plastic whole blood tubes with spray-coated KDEDTA (BD, UK, #367835) for a concentration of 1.8mg of EDTA per millilitre of blood.

2.6. Preparation of whole citrated blood

Platelet rich plasma (PRP) was utilised in both methods of platelet function analysis used in this thesis and was originally used to determine MPV up until October 2013. PRP was obtained by centrifugation of whole blood for 20 minutes at 200g in a Sorvall Legend XTR centrifuge (ThermoFisher, UK, #75004520). Platelet counts in PRP were measured with a Coulter Z_2 analyser (Beckman Coulter Ltd, UK) to determine the method of platelet function analysis to use and to measure MPV. 5µL of prepared PRP was diluted in 10mL diluent, measured three times and an average value taken for both platelet count and MPV. After October 2013, only platelet count in PRP was measured.

PRP from patients presumed to suffer from macrothrombocytopenia was prepared differently. Presence of macrothrombcytopenia was originally determined by clinical report, and since October 2013, macrothrombocytopenia was determined by an MPV >13fL in whole blood cell counts on the Sysmex XN-1000™blood analyser. In both instances, PRP was obtained by centrifugation for 30 minutes at 100g to try to eliminate loss of larger platelets in the final PRP layer.

2.7. Preparation of wash platelets

Washed platelet were prepared for use within platelet spreading assays. 20mL of acid-citrate-dextrose (ACD) (Sigma Aldrich, UK, C3821) and 50mL of modified Tyrode's buffer (134mM NaCl, 2.9mM KCl, 0.34mM $Na_2HPO_412H_2O$, 12mM $NaHCO_3$, 20mM HEPES, 1mM $MgCl_2$ all reagents available from Sigma Aldrich) containing 0.045g glucose (pH 7.3) were pre-warmed to 30°C. 5mL of warmed ACD was added to 50mL fresh blood (containing 5mL sterile sodium citrate as anticoagulant) and centrifuged for 20 minutes at 200g in a Sorvall Legend XTR centrifuge. The supernatant PRP was removed and 10µg prostaglandin I2 (PGI2) (10µL of stock solution at 1mg/mL [Sigma Aldrich, UK. #P6188]) was added and mixed before centrifugation at 1000g for 10 minutes. The supernatant was discarded and the remaining pellet resuspended in 1mL of modified Tyrode's buffer containing 150µL of ACD before being made up to a final volume of 28mL containing 25mL modified Tyrode's and 3mL ACD. Platelet count was determined using the Coulter Z_2 analyser. The platelets were re-pelleted by addition of 10µg PGI_2 and centrifuged at 1000g for 10 minutes. An appropriate volume of modified Tyrode's buffer containing ACD was added to resuspend the pellet to a final platelet concentration in solution of $2x10^7$/mL.

2.8. Platelet counts in whole EDTA-anticoagulated blood

Whole blood cell counts were performed on EDTA-anti coagulated blood using the Sysmex XN-1000™ haematology analyser (Sysmex, UK) from January 2014. Prior to this date, platelet and white blood cell counts in whole blood were determined at the haematological centre of study consent. Repeat patients were measured by the Sysmex XN-1000™ if a repeat sample was received after January 2014.

Counts on the Sysmex XN-1000™ were taken in manual operating mode for both sealed and open tubes depending on the previous method of sample collection. The analyser utilises the properties of fluorescent flow cytometry to determine accurate whole blood cell counts. There are five standard channels of analysis, however for this study only two were used. The WNR channel evaluates white blood cell count (WBC), basophils and nucleated red blood cells. The PLT-F channel exclusively measures platelets. An accurate platelet count is determined using oxazine; an RNA binding dye. A measurement of platelet production (Immature Platelet Fraction - IPF) and the mean platelet volume (MPV) is also in this channel. Platelet counts were checked against both the impedance (PLT-I) and optical (PLT-O) channels that are generated. If the values differed significantly, samples were re-run. Quality control material (XN check [Sysmex, UK]) was tested on a daily basis to ensure instrument performance throughout the study. A normal range was determined by analysis of 40 healthy volunteers in an alternate study; ranges are represented as the mean value +/-2 standard deviations as seen in section 3.6.

2.9. Platelet function testing

Platelet function testing was approached on an individual basis dependant on the platelet count in PRP. Patients with a platelet count in PRP exceeding $1.5x10^8$/mL were analysed by light transmission aggregometry (LTA) with lumiaggregometry. Patients with a platelet count $<1.0x10^8$/mL were analysed by a custom designed flow cytometry assay as outlined in section 2.8.2. The reason for this was that previous work within the laboratory determined that LTA was insensitive for patients with thrombocytopenia and a platelet count lower than $1.0x10^8$/mL (Dawood, Wilde et al.

2007). Patients with a borderline platelet count between $1.0x10^8$/mL and $1.5x10^8$/mL were measured by both assays in case of discrepancies.

2.9.1. Light transmission aggregometry with lumiaggregometry

LTA with lumiaggregometry was performed in adjacent studies originally by Dr Gillian Lowe and subsequently by Miss Sian Drake and Dr Rashid Al Ghaithi. All patients were analysed as mentioned in previous studies published by the UK-GAPP study group (Dawood, Wilde et al. 2007; Dawood, Lowe et al. 2012).

Patient and control samples were run simultaneously in all cases and subjected to the same methodology and agonist conditions. All aggregations were performed using a dual Chrono-log 460VS lumiaggregometer (Chronolog, US) and measured for light transmittance from baseline in native undiluted PRP upon agonist stimulation. ATP secretion from dense granules was measured using the same lumiaggregometer utilising the luciferase assay (Holmsen, Holmsen et al. 1966; Dawood, Wilde et al. 2007). Undiluted autologous platelet poor plasma (PPP) was used as a reference, and to set the aggregation scale, and was obtained after further centrifugation of whole blood samples at 1000g for 10 minutes in the Sorvall Legend XTR centrifuge (mentioned in more detail in section 2.9). All aggregations were performed within 6 hours of preparation of the PRP; priority was given to time-sensitive agonists which were used first e.g. ADP. PRP samples (400µl) were pre-warmed in siliconized glass cuvettes (LabMedics, UK) at 37°C for 120s and stirred at 1200rpm for 60s prior to agonist addition. Originally, a chart recorder was used, where percentage aggregation was calculated by the observed change in optical density compared to the full scale aggregation as set by the PPP control. In more

recent testing, responses were assessed using the AggroLink computer interface (ChronoLog/LabMedics, UK).

Platelets from patients and healthy volunteers were exposed to the following agonists in Table 2.9.1.

Agonist	Concentrations used	Supplier
ADP	3, 10, 30 and 100µM	Sigma-Aldrich, UK, #A2754
Adrenaline	10 and 30µM	Sigma-Aldrich, UK, #E-4375
Arachidonic acid	0.5, 1 and 1.5mM	Cayman Chemical,USA,#10006607
PAR-1 peptide	30 and 100µM	Alta Biosciences, Birmingham, UK
PAR-4 peptide	100, 250 and 500 µM	Richard Farndale, Cambridge, UK
Collagen	1, 3 and 10µg/mL	Nycomed, Austria, #NYAR 1130630
Collagen-related peptide (CRP)	1 and 3µg/mL	In-house synthesis
Ristocetin	1.5 and 2mg/mL	Helena Laboratories, USA, #5372

Table 2.9.1 Agonists used for routine LTA assessment. PAR-1 peptide [PAR-1 specific thrombin receptor activating peptide with sequence SFLLRN]. PAR-4 peptide [PAR-4 receptor-specific peptide with sequence AYPGKF]

A restricted panel of agonists was used initially and expanded or repeated if deemed necessary due to variable responses. If responses to ADP or arachidonic acid, two receptors involved in Gi signalling and the TxA_2 pathway, were deemed abnormal PRP was also exposed to the thromboxane receptor agonist U46619 (1 and 3mM) (Cayman Chemicals, USA, #CAY16450).

All agonists were diluted in phosphate buffer saline (PBS) with the exception of collagen and CRP which were diluted with their respective diluents.

Secretion was measured by lumiaggregometry by adding 30µl of CHRONO-LUME™ (Chronolog, USA, #395) to the highest concentration of each agonist tested. CHRONO-LUME™ contains 0.2mg luciferin, 22,000 units d-luciferase plus magnesium sulphate, human serum albumin, stabilizers and buffer per vial. It was added during PRP pre-warming following manufacturer's instructions. ATP secretion was quantified by adding a known amount (4nmol) of ATP standard (Chronolog, USA). As with aggregation most recently, secretion was assessed using the AggroLink computer interface.

2.9.2. Flow cytometry

An in-house flow cytometry assay was originally developed by Dr Gillian Lowe and was routinely used for the assessment of platelet function in patients with a platelet count in PRP of $<1.5x10^8$/L. Patient and control PRP was subjected to the same conditions, agonists and methodology simultaneously.

Flow cytometry analysis focuses on three main areas; (i) the resting quantitative levels of cell surface receptors, (ii) alpha granule release upon activation by a panel

of agonists measured by quantitative P-selectin (CD62P) cell-surface expression and (iii) the binding of fibrinogen to activated platelets.

(i) Resting cell surface levels of CD41 (integrin α-IIb, encoded by *ITGA2B*), CD42b (glycoprotein 1b α-chain, encoded by *GP1BA*) and GPVI were measured. 45µl of native PRP was added to individual reactions containing 5µL monoclonal mouse anti-human CD41/FITC clone 5B12 (Dako, Denmark, #47088) or 5µL FITC mouse anti-human CD42b (BD, UK, #555472), for a final concentration of antibody in PRP of 1:10. The solutions were incubated for 40 minutes at room temperature (~20°C). 45µl was added to 5µl 1:1000 diluted in PBS unconjugated anti-GPVI antibody (Prepared in house by Dr Eelo Gitz) and incubated for 30 minutes at room temperature (~20°C), after 15 minutes 2.5µL of anti-mouse IgG (whole molecule) F(ab')2 fragment-FITC (Sigma Aldrich, UK, #F2883) was added. All reactions are stopped after the allocated times with the addition of a five-fold volume (250µL) of ice-cold PBS. Reaction controls including a blank (void of any antibody) and an isotype control (5µL of Isotype reagent mouse IgG1/FITC [Dako, Denmark, #X0927]) were ran simultaneously on both patient and control samples.

(ii) Alpha granule release upon agonist stimulation was measured as follows. 45µL of native PRP was added to 5µL of FITC mouse anti-human CD62P (BD, UK, #555523) and incubated at room temperature (~20°C) for 40 minutes. Individual solutions of antibody and PRP were then exposed to the following agonists separately; ADP (3 and 30µM), CRP (0.3 and 3µg/mL) and PAR-1 peptide (10 and 100µM) and the solutions incubated for 2 minutes at 37°C. The reaction was subsequently terminated with the addition of a fivefold excess of ice cold PBS. Blank and IgG isotype controls were run simultaneously in tandem without the addition of any agonists.

(iii) Fibrinogen binding was measured by determining the relative levels of polyclonal rabbit anti-human Fibrinogen/FITC (Dako, Denmark, #F0111) to the cell surface receptor GpIIb/IIIa acting as a marker of integrin activation. 45µL of PRP, diluted 1:10 in PBS, was added to 5µL of FITC-conjugated fibrinogen. The same panel of agonists were used to determine fluorescent fibrinogen binding as P-selectin expression. All reactions were incubated at 37°C for 2 minutes before the reaction was terminated by the addition of a fivefold excess of ice cold PBS. In addition to all individual agonist reactions, a blank control was run in tandem.

All agonists used were the same as those used in LTA with lumiaggregometry so were sourced from the same suppliers.

All reactions including stimulated and resting cell surface reactions from both patient(s) and control were then analysed using an Accuri™ C6 flow cytometer (BD, UK). Samples were run ungated, at slow flow rate, 14µL/min, 10µm core, and 10,000 events were recorded per sample.

Mean fluorescent intensity (MFI) was calculated from the 10,000 recorded events and compared for each reaction against the parallel travel control. All reactions were also compared to responses from a panel of 20 healthy volunteers for each reaction.

2.10. Sample storage – Protein and buffy coat

Remaining PRP, after use for platelet function testing, was processed for storage of platelet based protein. An appropriate volume of PRP containing $3x10^8$ platelets, as determined from the mean platelet count in PRP, was added to labelled polypropylene Falcon™ 15mL conical centrifuge tubes. 1µL of PGI2 (1mg/mL stock solution) was added to each sample and mixed, and the samples immediately

centrifuged at 2000g for 10 minutes creating a cell pellet. The supernatant was removed and discarded by pipetting and the cell pellet resuspended in 1mL PBS before being transferred to a labelled 1.5mL Eppendorf tube. 1µL of PGI2 was then added, the sample mixed, and immediately centrifuged in a ThermoFisher Heraeus Pico 21 microcentrifuge (ThermoFisher, UK) at 13,000rpm for 2 minutes. Following centrifugation, the supernatant was removed and the remaining cell pellet resuspended in 600µL 1x sample buffer (SB). Samples were stored at -80°C until needed.

1x sample buffer; 60mM Tris-HCl pH 6.8 (Sigma Aldrich, UK, T6455), 2% SDS (20% solution) (Sigma Aldrich, UK, #L3771), 10% glycerol (Sigma Aldrich, UK, #G5516), and made up to 10mL with dH_20. Brilliant Blue R (Sigma Aldrich, UK, #B7920) was added to visualise loaded protein in electrophoresis gels.

Following removal of PRP from patient and control whole blood samples after centrifugation, samples were subsequently centrifuged at 1000g for 10 minutes in a Sorvall Legend XTR centrifuge. Further centrifugation separated the remaining whole blood into PPP and a red cell mass separated by a thin layer of white blood cells known as the buffy coat. The buffy coat was removed by pipetting and added to labelled 1.5ml Eppendorf tubes. Buffy coat was stored at -80°C until needed. Control buffy coat was not stored.

2.11. Preparation of platelet lysates from washed platelets

Prior to preparation of platelet lysates 2x lysis buffer was prepared. 2x lysis buffer contained; 300mM NaCl, 20mM Tris, 2mM EGTA, 2mM EDTA, and 4mL NP-40 detergent (IGEPAL® CA-630) (Sigma Aldrich, UK, #I8896). 2x lysis buffer was made

up to 200mL with dH_2O and adjusted to a pH of 7.5. On the day of preparation, 10mL of extraction buffer was prepared containing; 9.536mL of 2x lysis buffer and the addition of the following protease inhibitors (250µL sodium vanadate, 100µL leupeptin, 100µL AEBSF, 10µL aprotinin, and 4µL pepstatin). An equal volume of extraction buffer was added to prepared washed platelets and incubated on ice for 40 minutes with mixing by vortex every 10 minutes. Following incubation cells and lysis buffer were spun at 4°C at 12000g for 20 minutes. Following centrifugation, the supernatant was collected and used for subsequent experimentation.

2.12. Preparation of lysates for protein analysis from cell lines

Three cell lines were utilised within Western blot (WB) analysis as positive and negative controls of expression.

DAMI (Megakaryoblast, human) (ATCC, US, #CRL-9792)

OAT1 HEK 293T/17 (referred to as HEK293T herein) (ATCC, US, #CRL-11268G-1)

HEL (human erythroleukaemic cell line) (Mike Tomlinson, Birmingham, UK)

Cell line lysates were prepared in the same fashion as washed platelet lysates.

2.13. DNA extraction from buffy coat and whole blood

DNA extraction from both buffy coat and whole blood was completed using the Gentra Puregene Blood DNA extraction kit (Qiagen, UK, #158389).

For buffy coat, 750µL of thawed, mixed buffy coat was dispensed into a 15mL polypropylene Falcon™ tube. Three volumes of RBC lysis solution (2.25mL) was added and inverted to mix. The mix was left to incubate for 10 minutes at room

temperature (~20°C) and inverted once during incubation. The non-lysed cells were pelleted by centrifugation at 2000g for 5 minutes. The supernatant was discarded by pouring leaving approximately 100-200µL residual liquid left. The cell pellet was resuspended in this residual supernatant by vortexing the tube vigorously for 30 seconds. 3mL of cell lysis solution was then added to the resuspended pellet and the mixture vortexed for 30 seconds to lyse the cells. If cell clumps were visible, the mixture was incubated at 37°C for up to 24 hours, the mixture was periodically vortexed throughout the period to aid in lysis. Once the solution was homogenous, 15µL of RNase A solution was added and the solution mixed by inverting 25 times. The solution was re-incubated at 37°C for another 30 minutes before being cooled to room temperature. 3mL of protein precipitation solution was then added and the solution was vortexed for 20 seconds at high speed. The precipitated protein was pelleted by centrifugation at 2000g for 5 minutes. The DNA containing supernatant was transferred into a clean labelled 15mL polypropylene Falcon™ tube before 3mL of 100% isopropanol was added. The solution was mixed by inverting gently 50 times before the DNA was pelleted by centrifugation at 2000g for three minutes. The supernatant was discarded and the tube drained on a paper towel. 3mL of freshly made 70% ethanol was added and the tube inverted to wash the DNA pellet. The DNA was re-pelleted by centrifugation at 2000g for one minute. Again, the supernatant was discarded and the pellet left to air dry inverted on a paper towel. Once dry, 100-500µL of DNA hydration solution was added depending on the size of the DNA pellet as judged by eye and the mixture was vortexed for 10 seconds. The solution was incubated at 55°C overnight to allow the DNA to go into solution.

DNA extraction of whole blood followed the same process with the exception of an original starting volume of 2mL of fresh whole EDTA-anticoagulated blood added to a three times volume of RBC lysis solution.

Stored buffy coat was freeze thawed a maximum of twice before being discarded to avoid degradation of the stored DNA.

All DNA was stored at -80°C, DNA to be used frequently was aliquoted to avoid excess free-thawing.

2.14. DNA quantification

DNA was quantified depending on the following uses. DNA extracted for whole exome sequencing was quantified using Qubit® 3.0 fluorometric quantification (ThermoFisher, UK). 3µL of DNA sample was quantified each time according to the manufacturer's Qubit® dsDNA Broad Range (BR) assay kit instructions, and quantity and quality was determined on a Qubit 3.0 Fluorimeter (ThermoFisher, UK, #Q32850 for BR kit). Standards were re-run each day samples were analysed.

DNA for any other use other than whole exome sequencing was quantified using a NanoPhotometer® P360 Spectrophotometer (Implen, UK). Samples were vortexed to mix and 1µl of sample was tested and repeated two additional times. The average concentration and 260/280 values were recorded. Readings were blanked with 1µL of DNA hydration solution.

2.15. Next generation sequencing

2.15.1. Whole exome sequencing

Whole exome sequencing was performed externally by collaborators in King's College London headed by Prof Michael Simpson. <1µg of quantified genomic DNA was transported at ambient room temperature in sealed 1.5mL Eppendorf tubes. DNA was subsequently enriched for coding regions and intron/exon boundaries using the SureSelect Human All Exon 50Mb Kit (Agilent Technologies, UK) following the manufacturer's instructions. Sequencing was performed on an Illumina HiSeq 2500 (Illumina, UK) with 100bp paired-end reads. Sequences were aligned using Novoalign (Novocraft technologies, Malaysia) that aligns based upon a Needleman-Wunsch algorithm with affine gap penalties to the reference genome (hg19). Duplicate reads and reads mapping to multiple locations were excluded from downstream analysis. Single nucleotide variations and small insertions/deletions were identified and filtered for quality using the SAMtools software package and in-house software tools (Li, Handsaker et al. 2009). All calls with a read coverage of <4 were excluded from further analysis. Novelty of filtered variants was determined by comparison to known variants present within dbSNP139, the 1000 Genomes Project, EVS (evs.gs.washington.edu/EVS/) and the results from over 1200 in house control exomes analysed using the same protocol.

Copy number variants (CNVs) were detected using ExomeDepth (Plagnol, Curtis et al. 2012). CNVs were analysed using bioinformatics tailored to IT and will be discussed in section 4.3.4.

Variant called and post aligned sequencing bioinformatics was approached in a unique way and will be discussed in results section 4.3.1 as it forms part of this thesis.

2.15.1.1. Analysis of 5'UTR *ANKRD26* variants

Raw aligned reads of all exomes were analysed separately using the Integrative Genomics Viewer software (Robinson, Thorvaldsdottir et al. 2011) to determine the presence of variants within the 5' untranslated region (UTR) of *ANKRD26*.

2.15.1.2. Conservation and pathogenicity prediction

Conservation at the site of variation was determined using phyloP and phastCons scores. Scores were precomputed and offered by UCSC and coupled with the pathogenicity prediction software MutationTaster. PhlyoP values range between -14 and 6 and measure conservation at individual columns, ignoring the effects of neighbouring bases. Sites predicted to be conserved are assigned positive scores, those that are considered fast-evolving are assigned negative scores. phastCons values vary between 0 and 1 and reflect the probability that each nucleotide belongs to a conserved region, based upon the multiple alignment of genome sequences from 46 different species. The closer the value is to 1, the more probable the nucleotide is conserved. Only the value for the site of variation will be noted, the values for flanking bases are computed but not considered.

Pathogenicity was predicted using three separate *in silico* based pathogenicity prediction software. These were; MutationTaster (Schwarz, Cooper et al. 2014), Sorting Intolerant from Tolerant (SIFT) with Protein Variation Effect Analyzer (PROVEAN) (Kumar, Henikoff et al. 2009; Choi, Sims et al. 2012), and Polymorphism Phenotyping v2 (PolyPhen-2) (Adzhubei, Schmidt et al. 2010).

MutationTaster uses a Bayes classification to predict the disease potential of an alteration. Individual models are used dependant on the nature of the variant in question and the predictions questioned with outputted probability value of certainty. Variants are predicted as either disease causing or polymorphisms, if the variants are known they are described automatically.

SIFT deciphers tolerated from deleterious sequences dependant on the sequence homology in multiple aligned sequences. It works on the premise that protein evolution is correlated with protein function. Positions with a normalised probability of less than 0.05 are predicted to be deleterious, those greater than or equal to 0.05 are tolerated.

PROVEAN introduces a delta alignment score based on the reference and variant versions in comparison to sequence homologs. The threshold is set at -2.5 for binary classification, scores below the threshold are deleterious, those above, neutral.

PolyPhen-2 determines pathogenicity dependant on the sequence, phylogenetic and structural information characterising the substitution. Two pairs of datasets predict the pathogenicity score independently, for this study HumVar scores were used which uses all human disease-causing mutations from UniProtKB as a positive control together with common human nsSNPs (minor allele frequency (MAF)>1%) with no involvement in disease, which are treated as non-damaging.

Splice site predicted was performed using Human Splicing Finder (www.umd.be/HSF3 Last accessed February 2017) which combines algorithms based on either PWM matricies or maximum entropy principle to predict donor or acceptor splice site changes (Desmet, Hamroun et al. 2009).

2.15.1.3. Variant classification

Candidate variants were classified following the guidelines for the interpretation of sequence variants as recommended by a joint consensus of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (Richards, Aziz et al. 2015). Dominant inheritance was assumed for all patients unless exception was noted from patient provided family histories. All criteria for the determination of predicted pathogenicity were maintained as stated with the exception of segregation within multiple affected family members. As the majority of variants observed in this study are novel, classification of segregation was determined as follows:

When the variant was present in either; two affected related individuals and not present in at least one unaffected related family member or one affected individual and at least two unaffected related individuals, it was classed as supporting evidence of pathogenicity.

When the variant was present in either; three affected related individuals and not present in at least one unaffected related individual or at least three affected individuals across at least two families it was classified as moderate evidence of pathogenicity.

When the variant was present in more than three affected related individuals across at least three generations and not present in at least two controls it was classified as strong evidence of pathogenicity.

2.15.2. Thrombocytopenia specific panel sequencing

A thrombocytopenia panel was designed and performed as an initial NGS sequencing/pre-screen before whole exome sequencing in collaboration with the Regional Genetics laboratory at Birmingham Women's Hospital.

The panel was designed using the Agilent SureDesign v3.5.4 (Agilent Technologies, UK) design software. The original design included the 30 genes present in Table 2.15.1.

Gene
ABCG5
ABCG8
ADAMTS13
ANKRD18A
ANKRD26
CYCS
FLI1
FLNA
FYB
GATA1
GFI1B
GP1BA
GP1BB
GP5
GP9
HOXA11
ITGA2B
ITGB3
MKL1
MPL
MYH10
MYH9
NBEAL2
ORAI1
RBM8A
RUNX1
SLFN14
STIM1
TUBB1
WAS

Table 2.15.1 Genes including within the first version of the Agilent SureSelect^{QXT} IT specific sequencing panel

The 30 genes encompassed; genes previously identified to cause IT as well as some of their related genes and novel genes identified within this study. Sequencing probes/baits were designed to cover the following regions; all coding exons, the 5'UTR and 3'UTR regions and ±10 bp flanking from the intron-exon boundary. Sequencing baits were designed with 2x density so that each desired region was covered by at least two overlapping probes. Baits were also designed with the strictest masking stringency settings possible. SureDesign masks repetitive sequences dependant on three masking tools; RepeatMasker, WindowMasker and Uniqueness 35 track. The design software uses combinations of all three tools to create three masking stringencies which vary in their inclusiveness of repeat regions. If baits could not be found in the highest stringency possible, stringency was decreased until they could be found. 18 genes were covered entirely by the most stringent of settings, eight genes were covered by most and moderate stringency settings and the remaining four genes were covered in sections by a combination of all three stringency settings. Balanced boosting of GC-rich probes was used which replicated the amount of probes within a GC-rich region by a defined factor to improve capture of difficult genomic fragments. The final design incorporated 3,309 probes with an overall size of 212.189kbp.

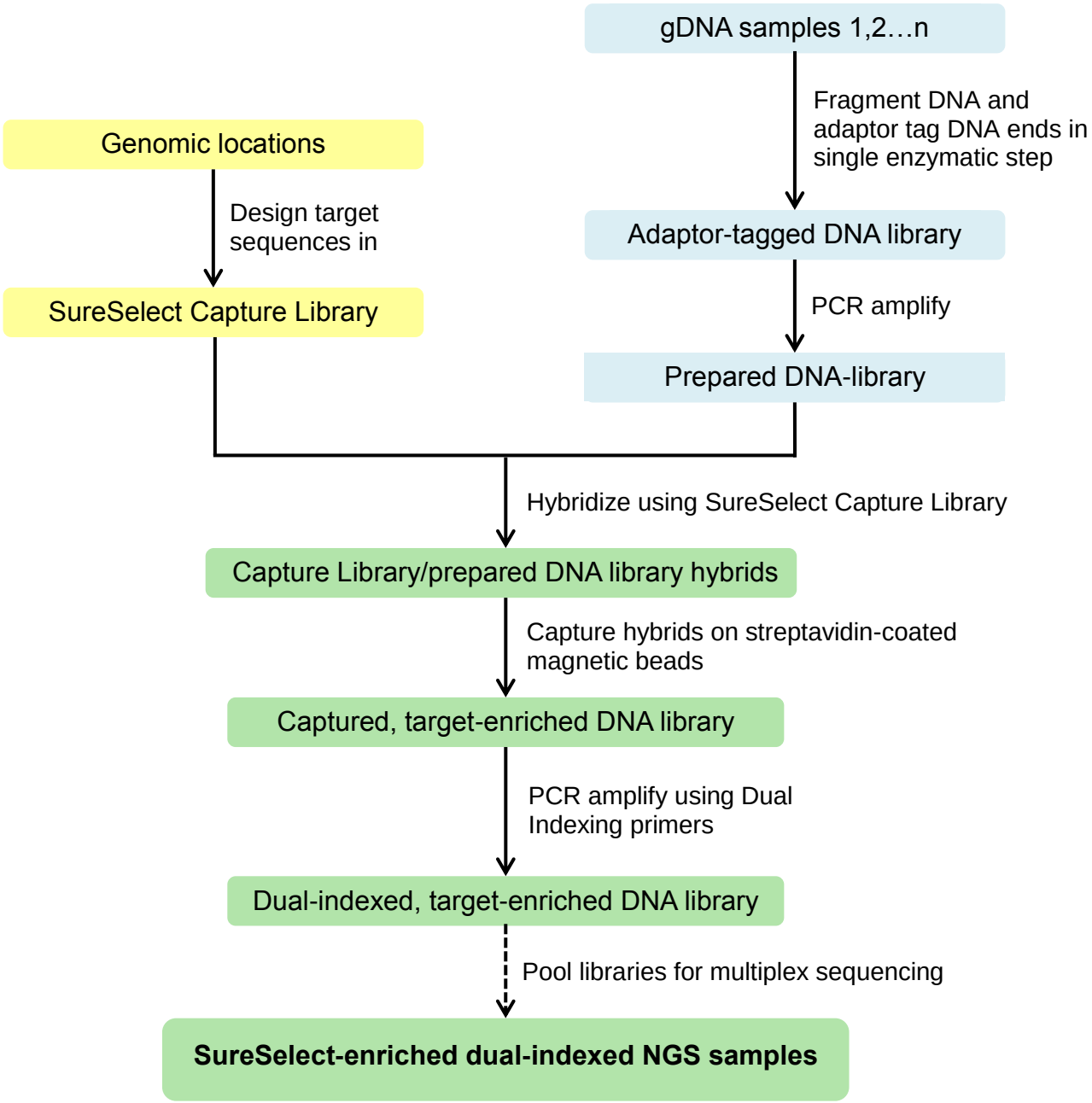
For subsequent sequencing beyond the first group of 16 patients, an improved design was utilised to include new genes implicated in inherited thrombocytopenia. The second version of the design included all probes from the first design with the addition of baits designed to sequence the following genes; *ACTN1, ETV6, PF4 and PRKACG.* All genes, with exception of PRKACG, were covered by probes with the most stringent masking settings. Probes designed for amplification and sequencing

of *PRKACG* included 4 probes with the least stringent masking settings applied. Overall the new design incorporated 3447 probes covering 221.305kbp.

Target enrichment was performed for all designs using Agilent SureSelect^{QXT} NGS target enrichment kit for Illumina multiplexed sequencing (Agilent Technologies, UK). Sample preparation followed the workflow outline in the accompanying manufacturer's instructions (Figure 2.15.1). Methodology followed the supplied user manual throughout. Due to the relatively small size of the capture library, all size related steps followed the methodology for capture libraries <3 Mb. In the preparation for hybridization, 750ng of gDNA, prepared to a volume of 12µL was used.

A maximum of 16 samples were prepared per run. DNA samples were quantified using BR and High Sensitivity (HS) Qubit® dsDNA fluorometric quantification kits and were analysed using a Qubit® 2.0 Fluorimeter (ThermoFisher, UK, #Q32854 for HS kit) in initial sample preparation. Purification steps utilised Agencourt AMPure XP magnetic capture beads (BeckmanCoulter, UK, #A63880). DNA quantity and quality was assessed at two separate points throughout the protocol using an Agilent 2200 Tapestation system and associated D1000 and high sensitivity D1000 screen tape (#5067-5582 for D1000, #5067-5584 for HS D1000), and reagents (including ladder and sample buffers) (#5067-5583 for D1000 reagents, #5067-5585 for HS D1000 reagents) (Agilent Technologies, UK). Dynabeads MyOne Streptavidin T1 magnetic beads were used for hybrid capture (ThermoFisher, UK, #65601). Index tags were added using the SureSelect^{QXT} P7 and P5 dual indexing primers. All thermocycling steps were performed using a Bio-Rad DNA Engine Tetrad® 2 Thermal Cycler (Bio-Rad, UK). Magnetic separation was achieved using DynaMag™-96 Side magnet (ThermoFisher, UK).

Figure 2.15.1 Sample preparation workflow for the IT specific next generation sequencing panel using Agilent SureSelectQXT capture methodology



Samples were pooled for multiplexed sequencing so that each index-tagged sample was in equimolar amounts in the pool. For each sample the following formula was used to determine the amount of index sample to use.

$$\text{Volume of Index} = \frac{V\,(\text{f}) \times C\,(\text{f})}{\# \times C\,(\text{i})}$$

V (f) = Final desired volume of pool

C (f) = Desired final concentration of all DNA in pool

= is the number of the indexes

C (i) = Initial concentration of each sample

A final desired volume of pool of 20µL was used and a final concentration of 4nM. In all cases, 16 indexes were pooled.

To achieve an optimal cluster density, a final concentration of 10pM of DNA was used. DNA was firstly denatured by the addition of 5µL of 0.2M NaOH to 5µL of 4nM pooled library and allowed to incubate at room temperature (~20°C) for five minutes. 990µL of pre-chilled HT1 hybridisation buffer was added to achieve a 20pM solution. 300µL of the 20pM solution was added to 300µL HT1 to achieve a final concentration of 10pM in 600µL.

Sequencing was performed using an Illumina MiSeq (Illumina, UK) and MiSeq v2 300 Cycle Reagent Kits (Illumina, UK, #15033626). Sample sheets were designed to allow for the use of custom primers and no adaptor trimming. Sequencing followed a Nextera XT sample preparation kit and amplicon chemistry.

Sequence alignment, annotation, categorisation and variant calling was performed using the SureCall v3.5 software (Agilent Technologies, UK) and the GenAligners 3.0 alignment tool (Agilent Technologies, UK). Post aligned, annotated and categorised

sequence data was analysed using a personalised bioinformatics pipeline that will be discussed in section 5.3.

2.16. Polymerase chain reaction

Polymerase chain reaction (PCR) was used throughout the project for the amplification of genomic DNA for subsequent Sanger sequencing. All PCRs were performed following a standardised in house protocol. All reactions formed a total volume of 25µL, consisting of: 100ng DNA, 5µmol of both forward and reverse primers, 12.5µL of Sigma RedTaq PCR ready mix (Sigma Aldrich, UK, #R2523) and UltraPure™ DNase/RNase-free distilled water (ThermoFisher, UK, #10977035). MyTaq™ Red Mix (Bioline, UK, #BIO-25043) was also used at the same volume as an alternative DNA polymerase ready mix if amplification was unsuccessful with Sigma RedTaq. Genomic DNA was used at a concentration of 20ng/L in all reactions and was prepared from extracted DNA stocks and diluted with UltraPure™ water. Where 20ng/µl was not available, DNA was used neat and the volumes of water and DNA adjusted to achieve 20ng/µl. All reactions were performed in parallel with a negative UltraPure™ water control.

Oligonucleotide primers were designed using Primer3 software (http://primer3.ut.ee/ Last accessed October 2016) (http://ihg.helmholtz-muenchen.de/ihg/ExonPrimer.html Last accessed October 2016) and validated using *In-Silico* PCR software (http://genome.ucsc.edu/cgi-bin/hgPcr Last accessed October 2016). Oligonucleotides were designed for an optimum annealing temperature of 60°C, length of 20bp and presence of a GC clamp. All oligonucleotides are available from Sigma Aldrich and sequences are available upon request.

PCR was conducted with the universal cycling conditions of denaturation, annealing and hybridisation as follows;

- 94°C for 3 minutes
    - 94°C for 1 minute
    - 60°C for 1 minute
    - 72°C for 1 minute

    } x 30 cycles
- 72°C for 5 minutes

All reactions were performed using the DNA Engine Tetrad® Peltier Thermal Cycler (BioRad, UK). PCR products were visualised after running using electrophoresis in a 1% ethidium bromide stained 1% agarose tris base, acetic acid, EDTA (TAE) buffer gel and viewed on an ultraviolet transilluminator system (Syngene Genegenius Bioimaging, UK). Sizes of bands were determined by loading all gels with 5µl of 1kb DNA ladder (NEB, UK, #N3232).

2.17. Sanger sequencing

Amplified DNA from PCR reactions was purified and concentrated using microCLEAN solution (Microzone, UK, #2MCL). PCR products and *micro*CLEAN solution were applied in equal volumes (2.4µL) to individual wells of a 96 well plate and the DNA pelleted by centrifugation at 2218g for 40 minutes. The plate was inverted and centrifuged for 30 seconds at 34.6g to remove any residual liquid. All centrifugation steps were performed using a Hettich Universal 320 plate centrifuge (Hettich, Germany).

4pmol of the appropriate forward and reverse primers specific to the gene of interest were used in conjunction with the Big Dye® Terminator v3.1 Cycle Sequencing Kit (ThermoFisher, UK, #4337454) for sequencing reactions. The sequencing was

performed under the following PCR cycling conditions for optimal elongation and random termination of fragments:

- 96°C for 30 seconds
- 50°C for 15 seconds
- 60°C for 4 minutes

} x 30 cycles

DNA was precipitated by the addition of 2µL 0.125M EDTA (Sigma Aldrich, UK, #E7889) and washed twice, firstly in 30µL 100% EtOH and secondly in 90µl 70% EtOH in water. DNA was denatured with 10µL of Hi Di Formamide (ThermoFisher, UK, #4311320) at 94°C for two minutes before snap chilling on ice. Subsequent sequencing analysis was performed using an ABI 3730 automated sequencer (Applied Biosystems®, UK).

Sequencing results were analysed using MutationSurveyor® software available from Soft Genetics® or alternatively viewed using the Chromas sequence chromatogram viewer. Variants occurring in regions not covered by >20 fold read coverage were discarded from further analysis. Areas with excessive background signals were repeated for clarity. If amplification and subsequent sequencing was unsuccessful, primers were redesigned flanking the region of interest and the process repeated.

2.18. PCR based X-inactivation assay

A PCR based X-inactivation assay was completed by collaborators at the Regional Genetics Laboratory, Birmingham Women's Hospital. Skewed X-inactivation was assessed using a PCR based assay that amplifies the first exon, containing a CAG repeat, of the androgen receptor gene on chromosome Xq11-12. 500ng of DNA was either digested overnight at 37°C using a methylation-sensitive restriction enzyme (H), HpaII (NEB, UK, #R0171), or left undigested (U). PCR was performed using

forward and reverse primers flanking the androgen receptor CAG repeat and the following cycling conditions;

- 95°C for 5 minutes
    - 95°C for 1 minute
    - 59°C for 1 minute
    - 72°C for 1 minute

    } x 26 cycles
- 72°C for 10 minutes
- 60°C for 30 minutes

Samples were run on the ABI3130XL genetic analyser (Applied Biosystems, UK) with a 12 second injection and standard formamide mix to determine length of fragments, indicating number of CAG repeats, and quantity of that fragment present in the sample. Skewing was determined by the use of a formula that calculates the ratio of the allele peak areas from the H samples (HpaII digested) to those of the U samples (undigested genomic DNA). Samples were run in parallel with a female control (50:50 random skewing) and a positive female control (100:0 completely skewed).

2.19. Western blot analysis

Western blots were utilised to determine protein expression of the encoded candidate genes of interest. Western blots were performed on stored platelet protein or patient platelet lysates if available, details regarding specific assays are outlined in the appropriate results sections.

All samples were denatured for five minutes at 105°C. 5µL of Colour Prestained Protein Standard, Broad Range (NEB, UK, #P7712S) ladder, made up to 20µl with SB, and 20µl of denatured samples were loaded into wells of a precast polyacrylamide 10 well Bolt™ 4-12% Bis-Tris Plus Gels (Invitrogen, UK, #NW04120).

Bolt® Mini Gel Tanks (Invitrogen, UK, discontinued) were used with 20x Bolt® MES SDS Running Buffer (ThermoFisher, UK, #B0002) diluted to 1x in dH_2O. Running buffer also contained 500µL Bolt® Antioxidant applied to the front chamber of gel tanks. Gels were run at 70V for 10 minutes and the voltage increased to 120V for >30 minutes depending on the size of the band of interest.

Protein bands were transferred from polyacrylamide gels to polyvinylidene difluoride (PVDF) membranes using the BioRad Trans-Blot® Turbo™ RTA mini transfer kit (BioRad, UK, #1704272) and the Trans-Blot® Turbo™ Transfer System (BioRad, UK). PVDF membranes were activated in 100% methanol for 2 minutes prior to transfer.

Membranes were blocked for 1 hour at room temperature (~20°C) in 5% skimmed milk powder in TBS-T (Sigma Aldrich, UK, #70166).

All primary antibodies were diluted in 5% skimmed milk powder in TBS-T and incubated overnight at 4°C on a rolling incubator. Primary antibodies and concentrations used are detailed in the appropriate results sections. In all cases Western blots were probed with an anti-GAPDH antibody loading control (Abcam, UK, #ab9485). Following primary antibody incubation, membranes were washed 3x in TBS-T.

Membranes were incubated at room temperature for one hour in secondary antibodies. The following secondary antibodies were used for WB analysis, dependant on the species of host the primary antibody was generated within; Amersham ECL Mouse IgG, HRP-linked whole Ab (from sheep) (GE healthcare, UK, #NA931)

Amersham ECL Rabbit IgG, HRP-linked whole Ab (from donkey) (GE healthcare, UK, #NA934)

Rabbit anti-Goat IgG (H+L) Secondary Antibody, HRP (ThermoFisher, UK, #81-1620).

Following incubation with secondary antibodies, WBs were washed 3x in TBS-T before staining with Pierce™ ECL Western Blotting Substrate (ThermoFisher, UK, #32106) following the manufacturer's instructions. WBs were visualised using Amersham Hyperfilm ECL (GE healthcare, UK, #28906836).

2.20. Platelet spreading and immunohistochemistry

Platelet spreading was performed on washed platelets prepared from whole blood. Shandon™ColorFrost™ Plus microscope slides (ThermoFisher, UK, #6776214) were prepared the day before experimentation. Slides were firstly washed in 100% EtOH and rinsed with PBS before being coated with 400µL 100µg/mL human fibrinogen (Enzyme Research Laboratories, UK) diluted in PBS. Slides were incubated overnight at 4°C and blocked the following day by the removal of fibrinogen and addition of 300µl of 5mg/ml denatured Bovine Serum Albumin (BSA) in PBS for 1 hour at room temperature. Slides were washed 3 times with PBS to remove any BSA. PBS was removed by aspiration and 300µL of washed platelets at a concentration of $2x10^7$/mL was added and subsequently incubated at 37°C, 5% CO_2. After 45 minutes, the slides were rinsed in 1x sterile PBS. Cells were then fixed by the addition of 300µL 10% formalin (Sigma Aldrich, UK, #HT501128) for 10 minutes. Slides were washed 3x with PBS.

Cells were permeabilised using PBS + 0.1% Triton X-100 (Sigma Aldrich, UK, #X100) for five minutes at room temperature and subsequently washed 3x in PBS. Slides were then stained with either Alexa Fluor® 488 Phalloidin (ThermoFisher, UK, #A12379) for one hour stain was removed by washing in 3x PBS. Slides stained with phalloidin were incubated in the dark. Slides requiring secondary antibodies were stained with secondary antibody in PBS + 0.5% BSA for one hour in the dark. All slides were washed 3x in PBS and once in dH_2O before mounting with 4µL hydromount (National Diagnostics, UK, #HS-106). Slides were stored at 4°C until imaging.

2.21. *TTF2* guide RNA *in silico* guide design and cloning

Guide sequences were generated using the MIT CRISPR database (crispr.mit.edu Last accessed July 2016) and targeted towards the C-terminus of *TTF2* spanning the stop codon. Guide sequences were selected as 20bp fragments which immediately preceded at the 3' with the universal 5'-NGG protospace adjacent motif (PAM) site for Cas9 recognition. An additional G-C base pair was added at the 5' end of the guide sequence to allow guanine as the first base of the U6 transcript. This does not adversely affect targeting efficiency. Guide sequences were also designed to include 4 bp overhangs complementary to the overhangs created by digestion with BsmBI. Four guide sequence complementary pairs were selected on the basis of inverse likelihood of off-target binding and are shown below in Table 2.21.1.

Guide	Orientation	Sequence
TTF2 guide A Top	5’	*CACC*G*CATCTAACCTCCTGTGGATA*
TTF2 guide A bottom	3’	CGTAGATTGGAGGACACCTAT*CAAA*
TTF2 guide B Top	5’	*CACC*G*CAGGAGGTTAGATGCCAAAAAGG*
TTF2 guide B bottom	3’	CGTCCTCCAATCTACGGTTTTTCC*CAAA*
TTF2 guide C Top	5’	*CACC*G*TTTTGGCATCTAACCTCCTGTGG*
TTF2 guide C bottom	3’	CAAAACCGTAGATTGGAGGACACC*CAAA*
TTF2 guide D Top	5’	*CACC*G*AGATGCCAAAAAAGGACTCTGAGG*
TTF2 guide D bottom	3’	CTCTACGGTTTTTTCCTGAGACTCC*CAAA*

Table 2.21.1 TTF2 CRISPR RNA guide sequences

The highest off target sites occurred within guide D and included 395 off target sites 43 of which occur within the gene. All guides were of above a quality score of 50. Guide sequences were suspended at 10µM and annealed to form a phosphorylated oligo duplex by incubation at 37°C with the addition of 2µL 10x T4 DNA ligase buffer (NEB, UK, #B0202S), 0.5µL T4 Polynucleotide kinase (NEB, UK, #M0201S), and 15.5µL H_20.

BPK1520 guide vectors were used to express the U6 promotor and small guide RNAs (Figure 2.21.1) (Addgene, UK, #65777). We thank the depositing K Joung lab for use of the BPK1520 vector. The vector backbone was digested by incubation with BsmBI (NEB, UK, #R0580S) and NEB 10x 3.1 Buffer (NEB, UK, #B7203S) overnight at 37°C. Vectors were also dephosphorylated by treatment with 0.5µl Alkaline Calf Intestinal Phosphatase (NEB, UK, #M0290S).

2µg of digested vector per guide was ligated with 0.5µM oligo duplexes at room temperature for 30 minutes with 1µL 10x T4 DNA ligase buffer and 0.5µL DNA ligase (NEB, UK, #M0202S).

2µL of ligated vector were transformed into 50µL Alpha-Select Silver Efficiency cells (Bioline, UK, #BIO-85206). Cells were first incubated with the DNA on ice for 30 min before heat shock at 40°C for 40 seconds, followed by a further two minutes on ice. Cells were suspended in 950µL S.O.C medium (ThermoFisher, UK, #15544034) and incubated on a rotary shaker for one hour at 37°C. Following incubation, 50µL of cell containing medium was plated on agar plates containing 100µg/mL ampicillin (Sigma Aldrich, UK, #A9393) and incubated at 37°C overnight. The following day, colonies were picked and grown overnight on a rotary shaker at 37°C in 10mL liquid ampicillin containing Lysogeny broth (LB).

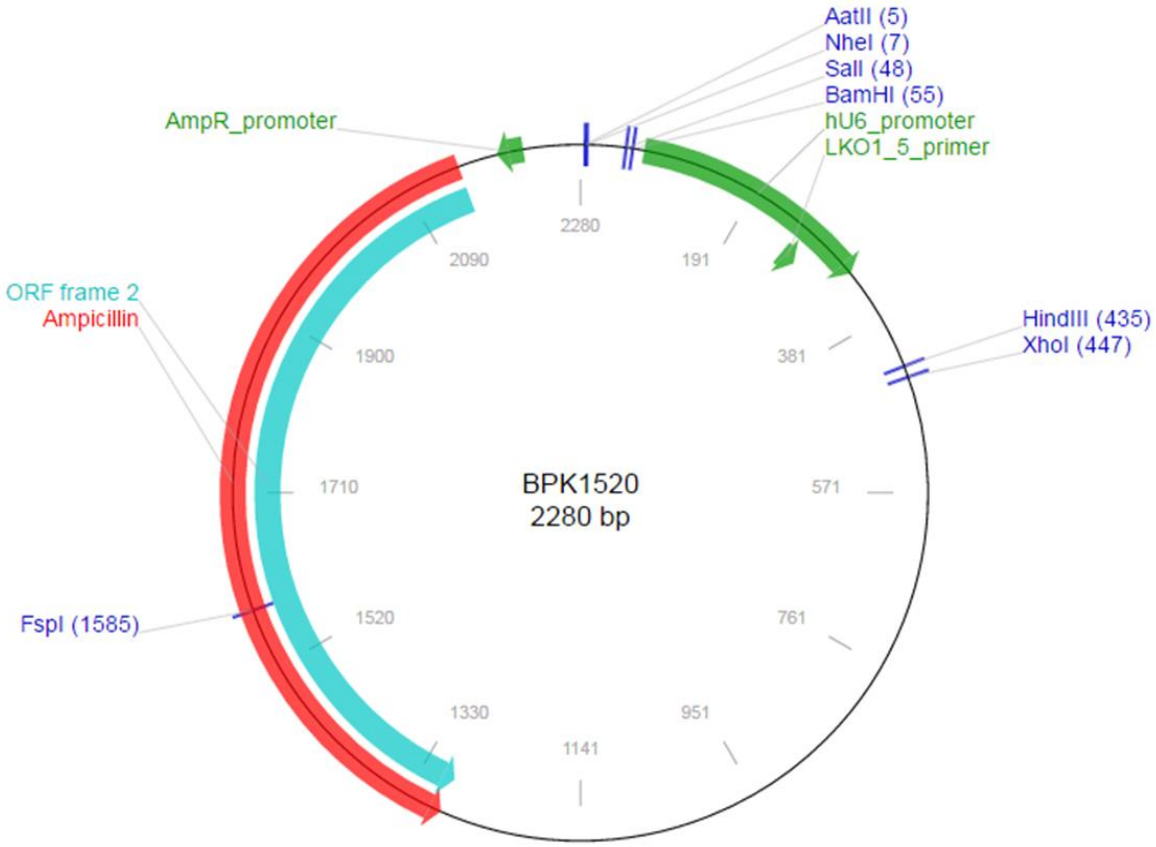


Figure 2.21.1 BPK1520 vector map indicating restriction sites, open reading frames and site of AmpR promoter.

3. Patient phenotypical presentations and platelet function testing

3.1. Summary of background to this research

Understanding a patient's phenotype is critical to the correct diagnosis of IT. The presence of secondary qualitative platelet function defects commonly associated with IT disorders has allowed many forms to be extensively phenotypically categorised. Disorders such as Bernard Soulier Syndrome and Glanzmann Thrombasthenia can often be distinguished based upon platelet function testing alone, due to the absence of CD42b (GP1b) and CD41 (GPIIb/IIIa) cell surface expression, respectively. Additionally, determining an accurate morphology of patient platelets can also help determine the aetiology, where disorders such as WAS are commonly associated with small platelets and MYH9-related disease and BSS are often associated with giant platelets. As a result, platelet function testing now forms a part of routine diagnosis when many forms of IT are suspected from a patients bleeding history. To date though, there is still a lack of a gold standard platelet function test with the capability to efficiently diagnose all forms of IT. The reason for this is that certain disorders present without a qualitative secondary defect or that the defects are heterogeneous dependant on a patient's individual genotype. It is even thought that phenotypic presentation can vary in individuals with the same genetic aetiology of disease, suggesting that secondary factors may have an impact on platelet function. Variable phenotypic presentation of disease suggests that instead of focusing on one method of diagnosis, patients should be considered individually and categorised dependant on a number of factors that may influence their clinical symptoms.

The UK-GAPP study works on this ethos and aims to completely categorise a patient's individual disease by a combination of extensive phenotyping and platelet function testing with next generation sequencing-based genotyping. The UK-GAPP study is a multinational study into rare inherited bleeding disorders which incorporates three research groups at the Universities of Birmingham, Bristol and Sheffield. Patients are recruited from 28 haematology referral centres in the UK which accounts for more than 2/3rds of the overall number of Haemophilia Comprehensive care centres (Figure 3.3.1). The study also recently started recruiting patients from Our Lady's Children's Hospital Crumlin, Dublin.

3.2. Aims

This section aims to build upon previous work by the UK-GAPP study group to completely categorise individual patient's phenotypes. Patients were tested following a unique workflow designed for patients with IT. I aim to combine whole blood cell counts with platelet function testing to help understand the pathophysiology of an individual's disease and aid in the subsequent genetic analysis.

3.3. Patients with inherited thrombocytopenia recruited to the UK-GAPP study

The UK-GAPP study has been recruiting patients with suspected inherited bleeding disorders for 10 years. Overall in excess of over 900 patients and healthy volunteers have been recruited to the study. Of those 900 patients, 153 patients have been recruited with suspected or sustained IT that meets the inclusion criteria as detailed in the methods section (Chapter 2). The 153 patients included 112 index cases/probands and 41 related affected family members. The work in this thesis is

based on 95 patients who have undergone platelet phenotyping in addition to genotyping by either WES or a next generation thrombocytopenia specific gene panel. These 95 patients represent those that have undergone the full workflow of the UK-GAPP study to analyse a patient's aetiology of disease. The UK-GAPP study incorporates all steps from patient recruitment, patient phenotyping, patient genotyping and further functional work. The workflow is outlined in Figure 3.3.2**.**

Figure 3.3.1 Map of the UK-GAPP study network

Red pins indicate all referring Haemophilia Comprehensive Care Centres and shown in blue are the research centres part of the study.

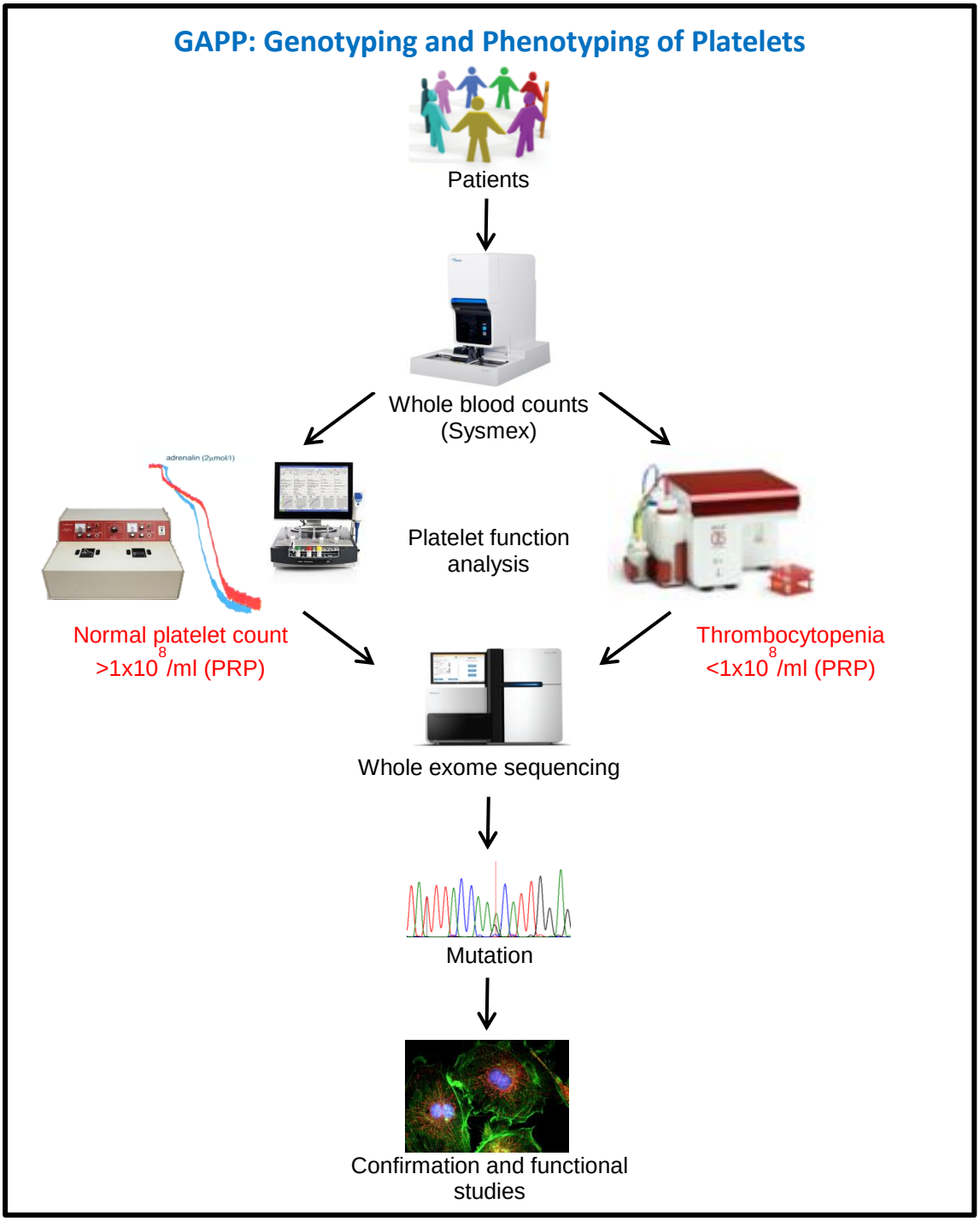


Figure 3.3.2 UK-GAPP patient workflow

3.4. Patient characteristics – Age, gender and ethnicity

All 95 patients that were fully analysed were recruited from participating UK haematology centres. Referral centres included both paediatric and adult clinics and as such patient age varies considerably across the cohort of tested individuals. Patient age was provided for 74 individuals. An average age of 30 was observed across the whole study of 74 individuals (range 2-79 years) (Figure 3.4.1).

Patient gender was known in 87 individuals. The cohort was 63% female (n=55) and 37% male (n=32) (Figure 3.4.1).

All patients recruited to the study were of European (non-Finnish) ethnicity. This is with the exception of patients 35.I and 35.II who were two related cousins of South Asian ethnicity. Downstream post-sequencing analysis was adjusted to factor in the difference in known inheritance in these patients.

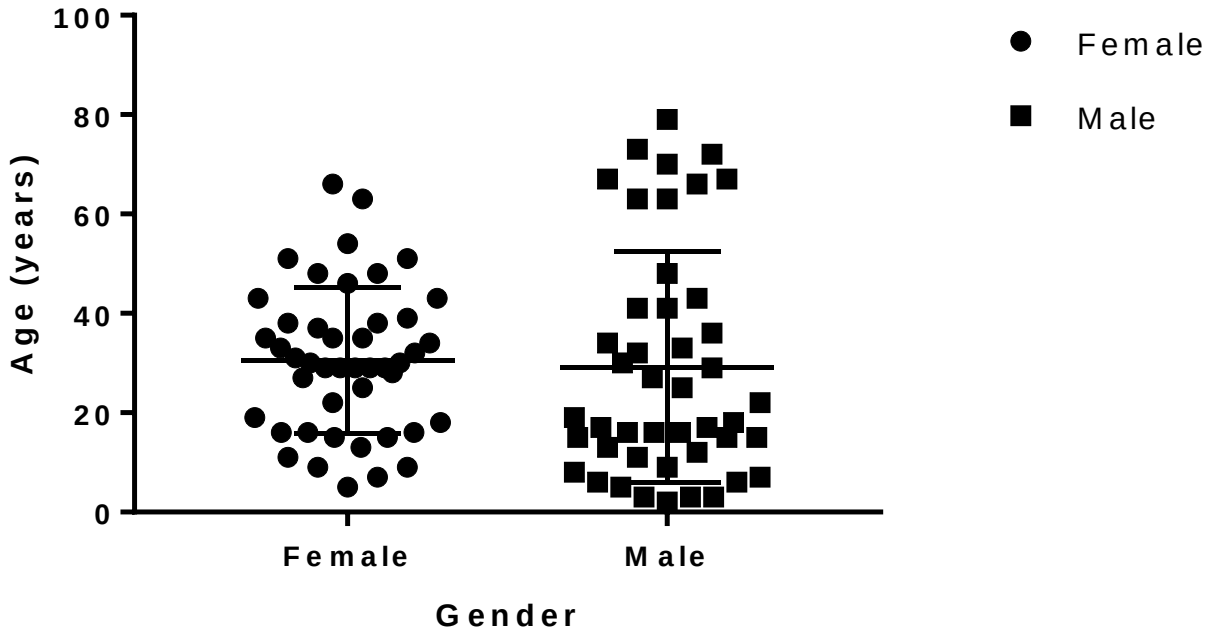


Figure 3.4.1 Age and gender distribution, including mean and standard deviation of 73 patients in which the data was available.

3.5. Patient bleeding symptoms and treatment

A patient bleeding history was taken at the point of examination and inclusion into the study. Most patients suffered from mild bleeding symptoms including cutaneous bruising, bleeding and epistaxis in addition to more severe bleeding symptoms in some cases. Detailed patient clinical symptoms related to bleeding were available for all patients and are displayed in Table 3.5.1.

Most common clinical symptoms were related to easy cutaneous bleeding/bruising; overall 86% of patients (79/91) said they had previously suffered from such symptoms. Epistaxis and prolonged oral cavity bleeding, especially after dental extractions and treatment, was the next most common symptom, present in 33% (30/91) and 12% (11/91) of patients, respectively. Of the 46 known female patients who had undergone menarche, menorrhagia was present in 41% (19/46). Several female patients also suffered from post-partum haemorrhage following vaginal child birth that was resolved using uterovaginal packing in most cases, but some cases required the use of antifibrinolytics and in rare cases a blood transfusion.

Two patients suffered from clinical symptoms not directly related to a bleeding diathesis which could be associated as a secondary symptom to a platelet function related disorder as follows.

Patient 14.I, a 29-year-old female at enrolment, presented with frequent epistaxis requiring cauterisation, menorrhagia, easy bruising and prolonged healing following minor cuts. She had also suffered previously with occasional haematochezia and oral cavity bleeding requiring dental extraction that required packing. In addition, the patient had undergone surgery for congenital cataracts aged eight months and

surgery to correct strabismus and detached retina aged 24. The patient is registered as partially sighted and suffers from tinnitus and hearing loss. Blood films were prepared following routine assessment and were negative for any neutrophil inclusions or the presence of giant platelets that may be attributed to MYH9-RD or Paris-Trousseau/Jacobsen syndrome, disorders often associated with secondary symptoms of hearing loss and cataracts.

Patient 71.I, a 15 year old female at the time of study inclusion, presented with menorrhagia, lethargy, joint aches and swelling, and gum bleeding and ulceration. She was initially investigated under the rheumatology team at Leeds General Infirmary for a possible diagnosis of systemic lupus erythematosus (SLE). Antinuclear antibody (ANA) results were negative in clinical testing and the patient was subsequently managed by the Regional Centre for Paediatric Haematology at Leeds Children's Hospital where she was enrolled within the study.

Patient 63.1 has also previously been tested clinically for antinuclear antibodies (ANA). Initial results dated 2011 were observed to be "weakly positive" for the presence of ANA but have since been negative when re-tested.

Developmental delay and skull abnormalities secondary to related hydrocephalus were noted in patient 41.II. Patient 37.II, a 3 year old boy, also presented with developmental delay.

The majority of patients were treated with methods to stem symptomatic episodes; cauterization for repeat epistaxis, packing for oral cavity bleeding and uterovaginal packing following post-partum haemorrhage. Rarely patients required more intensive therapy to treat bleeding episodes including; antifibrinolytics (including tranexamic

acid), DDAVP and platelet, RBC, plasma and whole blood transfusions. All treatments were used to stem bleeding episodes and treatment was stopped after recovery. In a few cases with severe menorrhagia, treatments such as tranexamic acid and other antifibrinolytics were utilised on a regular basis to improve symptoms.

Four patients received HLA matched blood product transfusions on a regular basis. RBC transfusions were used to treat recurrent bleeding episodes in patient 22.II. Whole blood transfusions were used in patient 25.I following an initial diagnosis of ITP but the patient did not respond to the treatment. Regular transfusions are still used to date in patients 35.I and 35.II, with patient 35.I receiving weekly HLA matched transfusions and patient 35.II receiving transfusions up to twice a month.

Ten patients were initially diagnosed with ITP before being reclassified as suffering from IT. All patients subsequently received treatment for the presence of ITP before correct diagnosis.

Family	Patient	Age	Gender	Platelet count ($x10^9$/l)	MPV (fL)	IPF (%)	Secondary defect	Bleeding Phenotype
1	I	35	F	73	9.6		Yes (Fibrinogen)	Cutaneous bruising/bleeding, menorrhagia
2	I	5	F	50	8.6		No	Cutaneous bruising
3	I		F	101	10.7		Yes (Secretion and cyclooxygenase)	Haematuria
4	I	35	F	80	10.3		Yes (Cyclooxygenase)	Cutaneous bruising, menorrhagia
	II	33	F	50	12		Yes (Cyclooxygenase)	Cutaneous bruising, epistaxis, purpura
	III	8	M	98	10.5	3.2		Cutaneous bruising, epistaxis, purpura
5	I			88	11.7			
	II			76	9.3			
6	I	19	F	142	11.8+		Yes (Secretion and Gi)	Oral cavity bleeding, epistaxis, menorrhagia
	II	48	F	157	11.4+		Yes (Secretion)	Cutaneous bruising, oral cavity bleeding
7	I			92	8.8		Yes (Secretion)	Cutaneous bruising, epistaxis, bleeding into joints
	II			100	8.6		Yes (Secretion)	Cutaneous bruising, life threatening bleeding following surgery
8	I	39	F	200	13.9+	24.5+	Yes (Secretion)	Cutaneous bruising/bleeding
	II	54	F	107	N/A	53.5+	Yes (Secretion)	Cutaneous bruising/bleeding
	III	28	F	85	N/A	55.1+	Yes (Secretion)	Cutaneous bruising/bleeding
9	I		M	123	13	11.3+	Yes (Secretion)	Cutaneous bruising
10	I	57	M	110			Yes (Secretion)	Cutaneous bruising, excessive bleeding following surgery

	II	36	M	100	8.9		Yes (Secretion)	Cutaneous bruising, epistaxis
11	I	22	F	70	10.7+		No	Cutaneous bruising, menorrhagia, post-partum haemorrhage
	II	29	F	70	10.1		No	Cutaneous bruising, epistaxis, haematuria, menorrhagia, post-partum haemorrhage
12	I	30	F	50	10.4		Yes (Secretion)	Cutaneous bruising, menorrhagia
13	I	29	F	147	12.7+	10.7	Yes (CD41)	Cutaneous bruising, hematomas, oral cavity bleeding, menorrhagia
14	I	29	F	35	11.4+		No	Cutaneous bruising, epistaxis, menorrhagia, haematochezia and oral cavity bleeding
15	I			55				Cutaneous bruising, epistaxis
16	I	9	F	62			Yes (Secretion)	Cutaneous bruising
	II	38	F				Yes (Secretion)	Cutaneous bruising, epistaxis, menorrhagia
	III	17	M	146			Yes (Secretion)	Cutaneous bruising, epistaxis
17	I	19	M	100	8		Yes (Secretion)	Excessive cutaneous bleeding
18	I	46	F	163	9.1		No	Cutaneous bruising, epistaxis, haematoma
	II	17	M	45	11.9+		Yes (GPVI)	Cutaneous bruising, epistaxis
19	I	16	F	139	8		Yes (Secretion)	Epistaxis, haematoma
20	I	3	M	90	7.6-		Yes (Secretion)	Cutaneous bruising, petechiae
21	I	33	M	130	7.1-		Yes (Secretion and Gi)	Cutaneous bruising, epistaxis, oral cavity bleeding
	II	6	M	70	7.5-		Yes (Complex)	Cutaneous bruising, epistaxis, oral cavity bleeding
22	I							Excessive bruising/bleeding

23	I	15	M	110	8.1		Yes (Secretion)	Cutaneous bruising/bleeding, petechiae, haematoma
24	I	16	M	100	9		No	Cutaneous bruising/bleeding
	II	37	F	100	9.2		No	Cutaneous bruising/bleeding
25	I		F	145	9.4		Yes (Secretion)	Cutaneous bruising, epistaxis, hematomas, menorrhagia
26	I	2	M	89	13+	17.5+	Yes (Secretion and Gi)	Cutaneous bruising
27	I		F	63	11.9	19.1+	Yes (Secretion)	Cutaneous bruising, epistaxis, haematoma
	II		F	83	11.9	24.3+	Yes (Secretion)	Cutaneous bruising, epistaxis, haematoma
28	I	30	M	74	11.2		Yes (Secretion and Gi)	Cutaneous bruising/bleeding, haematuria
	II	31	F	62	12.7+	20.8+	Yes (Secretion and Gi)	Cutaneous bruising/bleeding, menorrhagia, postpartum haemorrhage, haematoma
	III		F	109	11		Yes (Secretion)	Cutaneous bruising, haematoma, menorrhagia
29	I	16	F	119	11.1		Yes (Secretion)	Cutaneous bruising/bleeding
30	I	51	F	104	9.6	26.8+	No	Menorrhagia, post-partum haemorrhage
	II	79	M	133	8.6		No	Epistaxis
31	I	70	M	11	13.4+		Yes (Secretion and Fibrinogen)	Cutaneous bruising
32	I	43	F	43	14+		No	Cutaneous bruising, menorrhagia, oral cavity bleeding
33	I	29	F	100	10.3		Yes (Secretion)	Cutaneous bruising/bleeding, epistaxis, oral cavity bleeding
34	I	16	M	25	8.5		Yes (Secretion)	Cutaneous bruising

35	I	16	F	15	9.4		Yes (Secretion)	Haematomas
36	I	38	F	137	7.9		Yes (Secretion)	Cutaneous bruising, epistaxis, menorrhagia
	II	35	F	186	8.6		No	Cutaneous bruising, menorrhagia, haematuria
	III	63	M	80	8.2		Yes (Secretion)	Cutaneous bruising
37	I	30	F	20	9.7		Yes (Secretion)	Cutaneous bruising, epistaxis, oral cavity bleeding
	II	3	M	17	N/A	56.4+	Yes (Secretion)	Easy bruising, bleeding after surgery.
38	I	51	F	15	9.5	20.2+	No	Cutaneous bruising
39	I	16	M	66	9.9	1.8	Yes (Secretion)	Cutaneous bleeding
40	I	67	M	93	14.4+	20.5+	Yes (Secretion)	Cutaneous bleeding, epistaxis
41	I	7	F	15	10.4	87+	Yes (Secretion and other)	Cutaneous bruising, epistaxis, haematomas
	II	3	M	14	15+	83+	Yes (Secretion and other)	Cutaneous bleeding
42	I	32	F	104	13.3+	17+	No	Menorrhagia
43	I	66	F	130	9.7		No	Cutaneous bruising, epistaxis, menorrhagia
44	I	73	M	70	12	12+	N/A	Haematuria, bleeding following surgery
45	I	63	F	100	9.9	2.8	Yes (Cyclooxygenase)	Haematomas, ovulation bleeding
46	I	6	M	147	12.4		Yes (Secretion)	Cutaneous bruising
47	I	29	F	53	11.7	9.1	No	Cutaneous bruising, haematuria, oral cavity bleeding
48	I	5	M	125	9.1		Yes (Secretion)	Cutaneous bruising, petechiae
49	I	41	M	30	N/A	9.4	Yes (CD42b)	Cutaneous bruising
50	I	41	M	30	N/A	59.4+	Yes (CD42b)	Cutaneous bleeding
51	I			162	9.2		Yes (Secretion and	Cutaneous bleeding, oral cavity

							GPVI)	bleeding
52	I	43	F	131	8.7	14.7+	Yes (Secretion and Gi)	Cutaneous bleeding, epistaxis, menorrhagia, Gi bleeding, oral cavity bleeding
53	I	27	F	104	9.1		No	Cutaneous bleeding
54	I	12	M	101	10	39.8+	Yes (Secretion)	Cutaneous bleeding
55	I		F	30	8.6	2.3	No	Cutaneous bruising, oral cavity bleeding, menorrhagia
56	I	15	F	48	10.2	45.2+	No	Cutaneous bleeding, epistaxis
57	I	11	F	153	12.1	9.2	Yes (Secretion)	Cutaneous bruising/bleeding
58	I	9	F	82	8.6	6.4	Yes (Secretion and Fibrinogen)	Cutaneous bruising/bleeding
59	I		M	94	12.3	7.1	No	Cutaneous bruising/bleeding
60	I		F	146	13.4+	15.8+	No	Cutaneous bruising, oral cavity bleeding, menorrhagia
61	I		F	76	9.7	3.4	Yes (CD41 and other)	
62	I	34	F	138	13.8+	17.5+	No	Cutaneous bleeding
63	I	13	F	37	14.6+	16+	Yes (Secretion)	Cutaneous bruising, petechiae, epistaxis
64	I		F	105	14.5+	23.1+	Yes (Secretion)	Cutaneous bruising, epistaxis
65	I	35	F	52	14.9+	19.4+	Yes (other)	Cutaneous bruising/bleeding
66	I	18	F	87	10	1.8	Yes (Secretion)	Cutaneous bruising/bleeding
67	I	22	M	40	13.1+	15.7+	No	Cutaneous bruising/bleeding, epistaxis
68	I	72	M	50	8.4		Yes (Secretion)	Cutaneous bruising, epistaxis
69	I	48	F	92				Cutaneous bruising, haematomas
70	I		F	92	10.2		No	
71	I	15	F	76	9.4	13.5+	No	Oral cavity bleeding, menorrhagia

72	I	M	101	13.9+	14.4+	No	Cutaneous bleeding

Table 3.5.1 Phenotypic symptoms of all patients recruited to the UK-GAPP study with IT of unknown aetiology. Family members 41.I and 41.II are cousins born from consanguineous relationships. MPV – Mean platelet volume. IPF- Immature platelet fraction. M – male, F – Female. + denotes an elevated MPV/IPF, - denotes a reduction in MPV when compared to controls.

3.6. Platelet counts and morphology

Platelet counts were measured by different haematology cell counters dependant on the date of inclusion into the study, the differing methods are outlined in methods section 2.7.

Platelet counts were determined at the haematological referral centre in 62 patients and 33 patients were analysed using the Sysmex XN-1000. All platelet counts are shown in Table 3.5.1. Platelet counts are also displayed in comparison to travel controls and a panel of 40 healthy volunteers who have been analysed by the Sysmex XN-1000 in Figure 3.6.1.

Platelet counts were available for all patients with the exception of patients 16.II and 22.II as follows.

A platelet count was unavailable for patient 16.II at the time of enrolment; the patient shared a bleeding phenotype similar to affected family members and was included in further study due to clinical history of thrombocytopenia.

Platelet counts, morphology and platelet function testing were unavailable in patient 22.I due to the enrolment of the patient posthumously. This patient will be discussed in further detail in section 4.4.

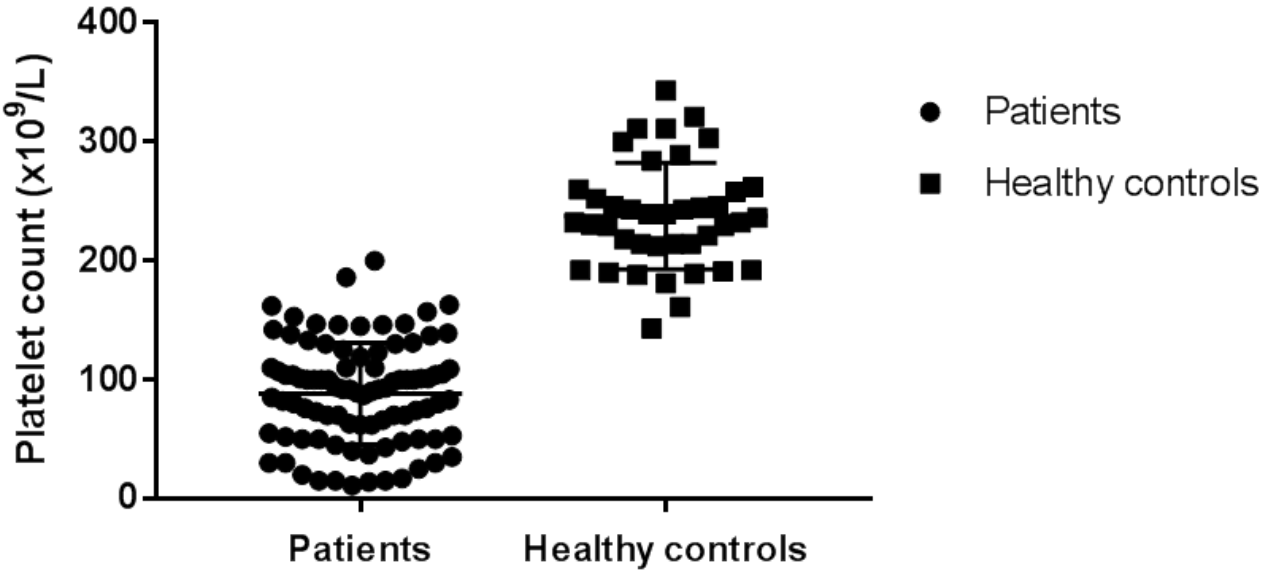


Figure 3.6.1 Circulating platelet counts in whole blood in all patients recruited to the UK-GAPP study compared to a panel of 40 healthy controls analysed by the Sysmex XN-1000.

Overall platelet counts ranged from 15-200x10^9/L across all patients analysed. An average platelet count across all individuals of 88x10^9/L was observed (n=91) (normal range to 2SD 147-327x10^9/L as measured on the Sysmex XN-1000, n=40). Six patients (patients 6.I, 8.I, 18.I, 36.II, 51.I and 57.I) were included within the study with a platelet count exceeding 150x10^9/L at the time of enrolment. All six patients had previous platelet counts <150x10^9/L prior to enrolment (n=2) or showed a shared phenotype to related affected family members (n=4).

Mean platelet volumes were assessed by two different methods. 88 patients were assessed for mean platelet volume in total, this is comprised of 48 patients analysed by the Beckman Coulter counter and 40 patients analysed by the Sysmex XN-1000. All MPV values are shown in Table 3.5.1. An average MPV of 10.64fL was observed across all patients (range 7.1-14.9fL) (normal range to 2SD 7.8-12.69fL, n=40). Values are denoted with + and – depending on the presence of macro or microthrombocytopenia, respectively, when compared to a range of healthy volunteers specific to the method of measurement used. In total 18 patients presented with a macrothrombocytopenia and three; two related individuals 16.I and 16.II, and 15.I presented with a microthrombocytopenia. Mean platelet volume was unable to be established in five patients due to abnormal platelet distribution when analysed by the Sysmex XN-1000. This is denoted with N/A in Table 3.5.1 and is often due to large/clumped platelets exceeding the predefined threshold size and appearing within the RBC range.

Immature platelet fraction (IPF) measurements were available for 38 patients. Immature platelet fractions ranged from 1.8-87% of the total platelet count (normal range 1.3-10.8%). An average IPF of 23% was observed across all 28 patients.

Twenty-eight patients presented with an IPF that exceeded the normal range. In 13 individuals, both an increased MPV and IPF were observed simultaneously.

3.7. Whole blood white cell counts

White cell counts were measured in whole EDTA anti-coagulated blood as outlined in methods section 2.7. In a two-stage reaction, cells were initially lysed and then labelled with a fluorescent marker specific to white blood cells, basophils and nucleated red blood cells.

White blood cells were measured in 25 patients and were present within a normal range in all patients (normal range ± 2SD 3.78-10.11x10^9/L, n=40).

3.8. Platelet function testing

Platelet function testing was performed on 86 of the 95 recruited patients. Nine patients were unavailable for testing either due to the lack of enough material to perform testing or recruitment of the patient specifically for genetic testing only and sent as a DNA sample. As mentioned in the methods section, platelet function testing was split dependant on the platelet count in PRP. 21 patients were tested by light transmission aggregometry with lumiaggregometry only. 46 patients were tested by a flow cytometry assay developed in house. 19 patients had a borderline count in PRP of >1 and <1.5x10^8/mL and were thus tested by both assays.

3.8.1. Light transmission aggregometry

In total, platelet function was analysed in 40 patients by LTA with lumiaggregometry (Table 3.5.1). Overall, a qualitative defect secondary to the reduction in platelet count was observed in 82.5% of patients (31/40). Nine patients therefore showed no

alteration in platelet function as analysed by light transmission aggregometry. Two patients, who were analysed by both methodologies, revealed normal responses to all doses of agonists when measured by LTA but presented with defects in cell surface receptors when analysed by flow cytometry.

One patient, 44.I, showed a lack of response to all agonists when tested by LTA. The patient had a platelet count in PRP of $1.02x10^8$/mL which is within the cut off for testing by LTA. However, upon analysis a lack of response was observed to all agonists and the results were difficult to interpret, potentially due to the low platelet count. Analysis by flow cytometry was unavailable on the day of testing.

Defects in dense granule secretion, as determined by lumiaggregometry measuring ATP secretion in response to 100µM PAR1 stimulation, were most common, occurring in 27 patients (60%) (representative trace shown in Figure 3.8.1). When available, secretion was measured as a concentration of ATP (nmol)/$1x10^8$ platelets. These values are displayed in Figure 3.8.2 simultaneously with travel controls tested on the same day and a panel of healthy volunteers. Average secretion values of those available were 0.69 nmol ATP/$1x10^8$ platelets for patients and 1.24 nmol ATP/$1x10^8$ platelets for travel controls. The 5^{th} percentile for secretion in response to PAR1 peptide 100µM lies at 0.65 nmol ATP/$1x10^8$ platelets in previously studied healthy volunteers (n=40). When comparing secretion values in the patient population to those in the control a significant difference (*p* 0.0008) was observed.

Seven patients showed a transient aggregation in response to all doses of ADP tested, indicative of a defect in Gi signalling or potentially defective alpha granule

secretion (representative trace shown in Figure 3.8.3). All patients, with the exception of 21.I also showed a subsequent lack of response to all doses of adrenaline.

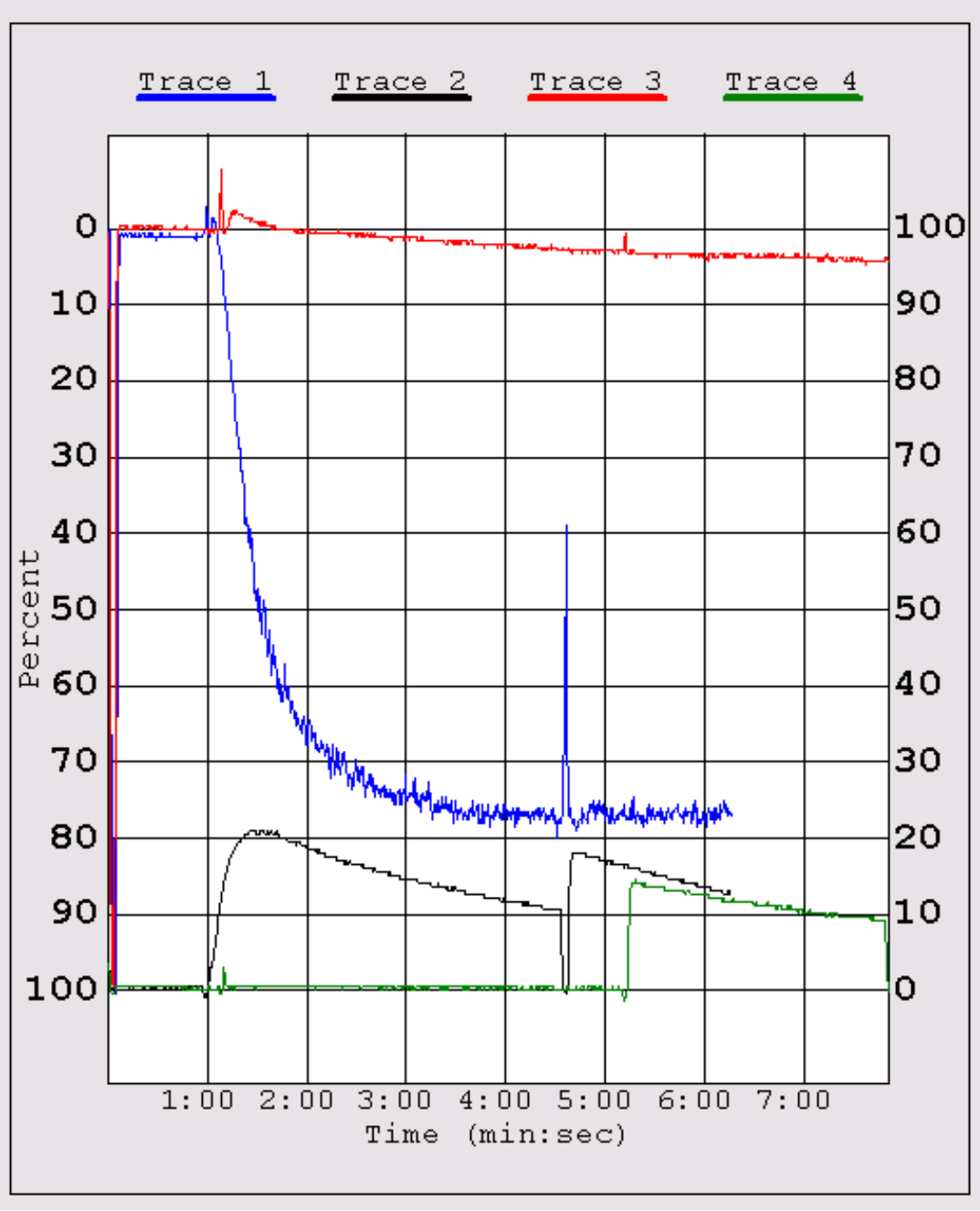


Figure 3.8.1 Representative trace of a patient dispalying a defect in secretion in response to PAR-1 peptide 100μM (Patient 27.I). Red – Patient aggregation, blue – control aggregation, green – patient secretion, black – control secretion. Patient secretion (green) shows no response to stimulation by PAR-1 peptide when compared to control secretion (black).

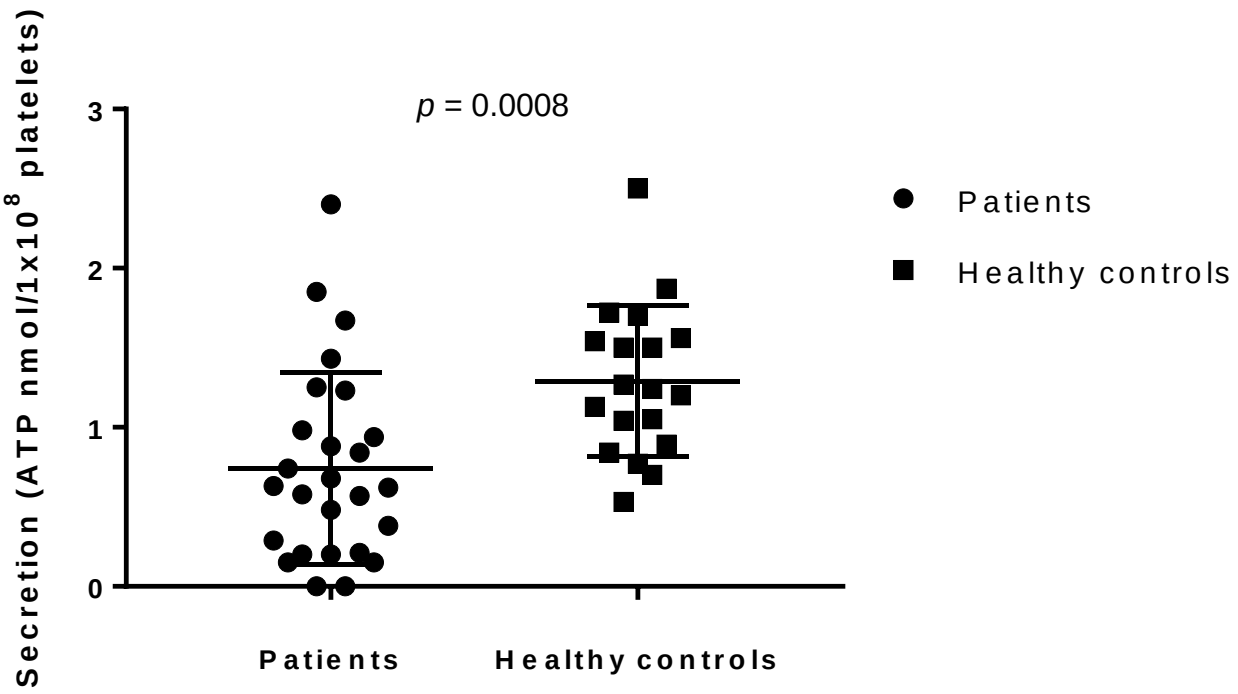
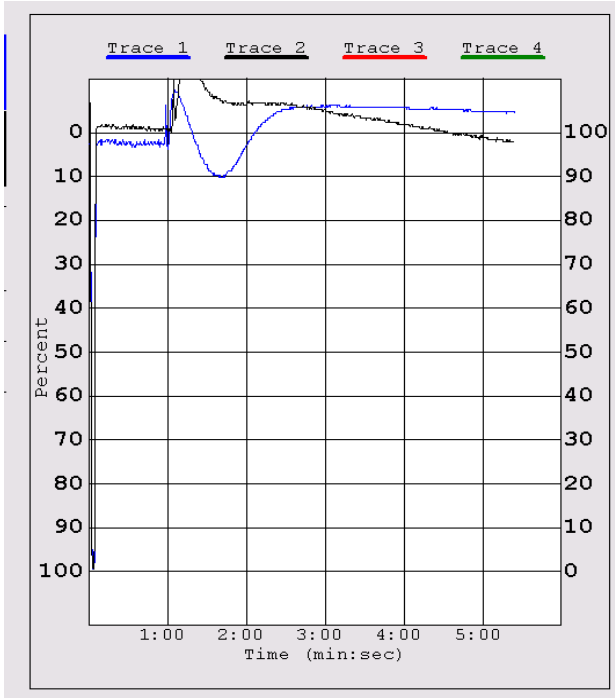
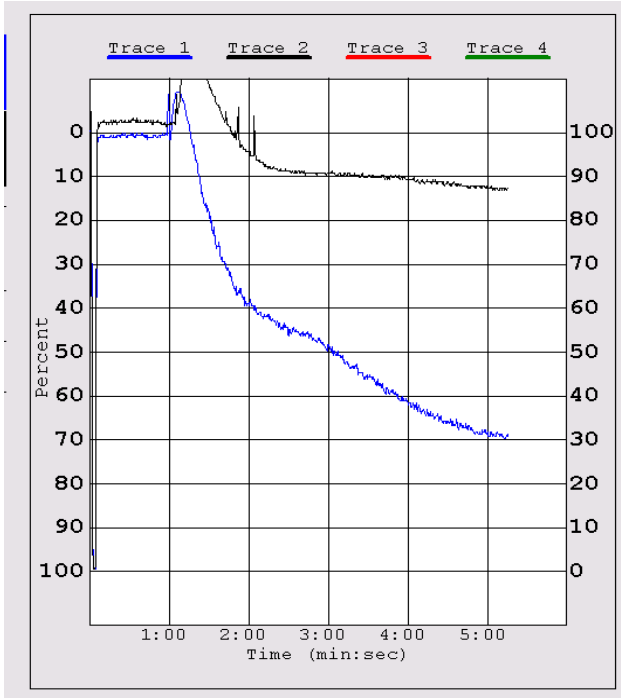


Figure 3.8.2 Secretion values in response to PAR1 100μM in 24 patients compared to 17 travel healthy controls. Secretion values calculated from aggregation traces as a concentration of ATP (nmol)/1x10^8 platelets. Using an unpaired t-test the two-tailed P value was statistically significant at 0.008.

ADP 3µM



ADP 10µM



ADP 30µM

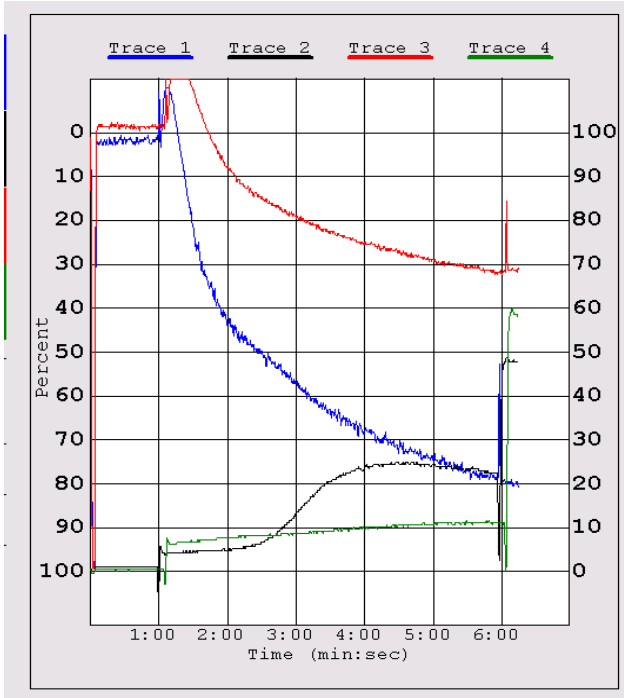


Figure 3.8.3 Representative traces displaying a defect in Gi signalling correlating as a lack of response to all doses of ADP (3, 10 and 30µM) (Patient 52.I). Black/red – patient aggregation, blue – control aggregation, green – patient secretion, black (in 30µM only) – control secretion. Patient aggregation traces (black in resonse to 3 and 10µM and red in response to 30µM ADP stimulation) show a lack of full or transient aggregation in response to agonist stimulation.

An absence of response to all doses of arachidonic acid was noted in three patients, 4.I, 4.II and 45.I. The suggestion of a defect in the cyclooxygenase pathway in all three patients showed a complete lack of response at low dose 0.5mM AA and partial responses at high dose 1mM AA.

Aggregation in response to U46619, a thromboxane A2 mimetic, was measured in two patients. Aggregation was normal when compared to the travel control in patient 22.I, however a lack of aggregation was noted in patient 45.I indicating a potential receptor defect.

Patient 18.II showed inhibition of aggregation to all concentrations of collagen and CRP tested. This observation was not shared within the affected mother, 18.I, and indicates a potential defect in the collagen signalling pathway through the GPVI receptor.

Patient 22.II presented with a complex platelet function phenotype when tested by LTA. The patient shared the observation of a defect in dense granule secretion with the patient's maternal uncle, 22.I. However, in addition, the patient displayed an impairment in response to ADP which resulted in reversible aggregation, primary wave only when stimulated with adrenaline and reduced aggregation to AA, PAR-1 and collagen. The patient was further tested with a combination of ADP (30μM) and PAR-1 peptide (100μM) and full aggregation was observed indicating that the complex defect is not a result of insensitivity of LTA when measuring patients with a low platelet count.

3.8.2. Flow cytometry

46 patients were analysed by an in house developed flow cytometry protocol, either solely or in combination with LTA. A secondary qualitative defect was observed in 50% of patients (23/46). 23 patients showed no observable defect when analysed by flow cytometry and when compared to travel controls and a range of healthy volunteers.

As with patients analysed by LTA, defects in secretion were the most common secondary qualitative defect in platelet function. 33 patients showed a reduction in the secretion overall, 23 of which showed a reduction in secretion when only tested by flow cytometry. Unlike LTA, secretion defects presenting in flow cytometry results indicate a predicted defect in alpha granule secretion due to a reduction in P-selectin expression upon stimulation. An example of this is shown in Figure 3.8.4. The 10 patients with a defect in secretion who were analysed by both methodologies point to a pan-secretion defect due to a reduction in both alpha and dense granule secretion.

Three patients, 1.I, 31.I and 58.I, showed a reduction in response to all agonists of fluorescent fibrinogen binding (Figure 3.8.5). A reduction in response to all agonists may be indicative of a defect in the integrin $\alpha_{IIb}\beta_3$ complex, however CD41 (encoded by *ITGA2B*, which makes up the α_{IIb} section of the complex) surface receptor levels were normal when compared to a travel control in all three patients. Two of the three patients, 31.I and 58.I, also displayed a reduction in P-selectin expression as well as a reduction in fluorescent fibrinogen binding.

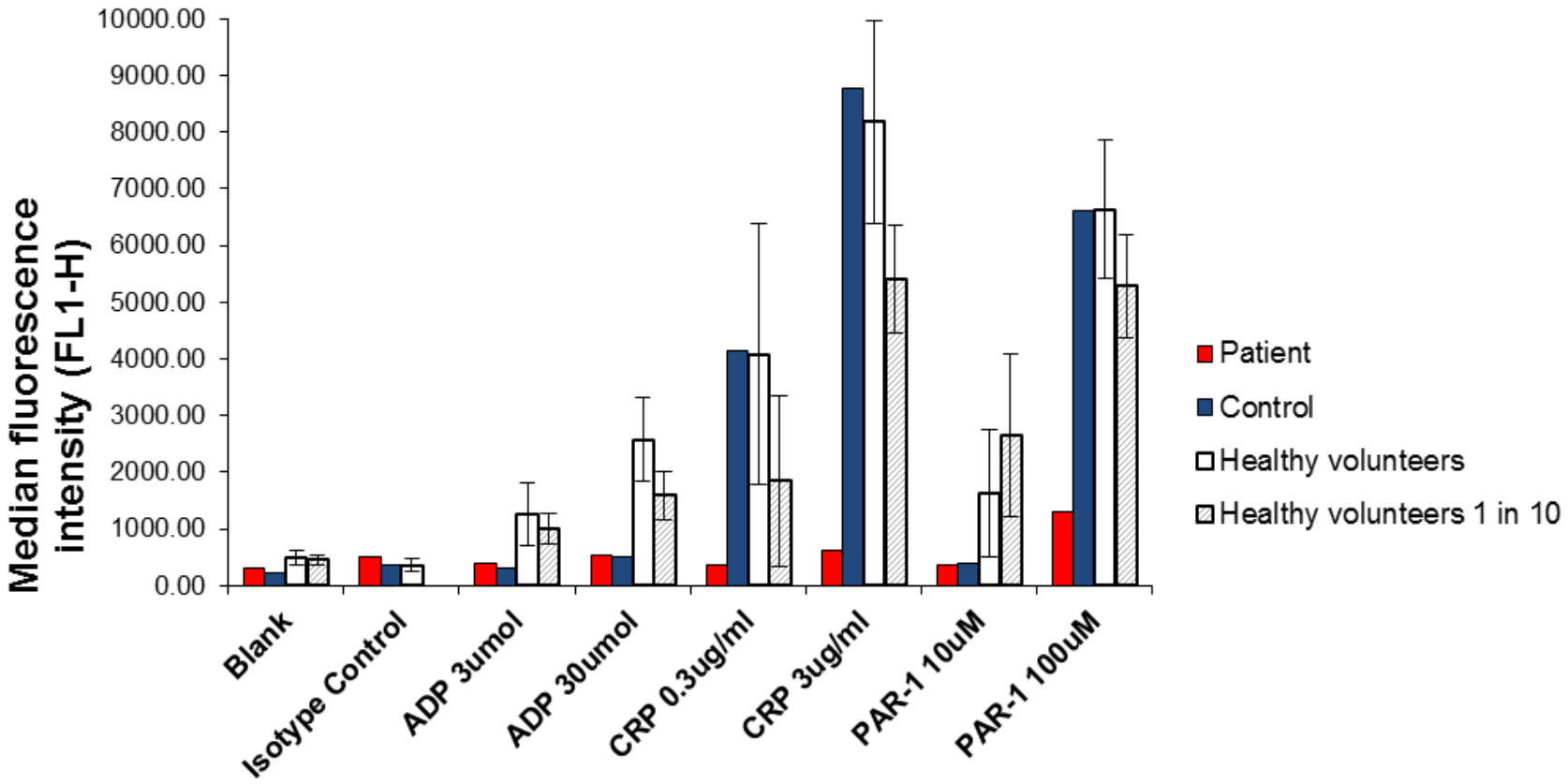


Figure 3.8.4 Representative example of a defect in alpha-granule secretion as analysed by a reduction in P-selectin expression in flow cytometry analysis (Patient 64.I). In response to all doses of agonist, median fluorescent intensities for patient agonist stimulations (red) are reduced when compared to the travel control (blue).

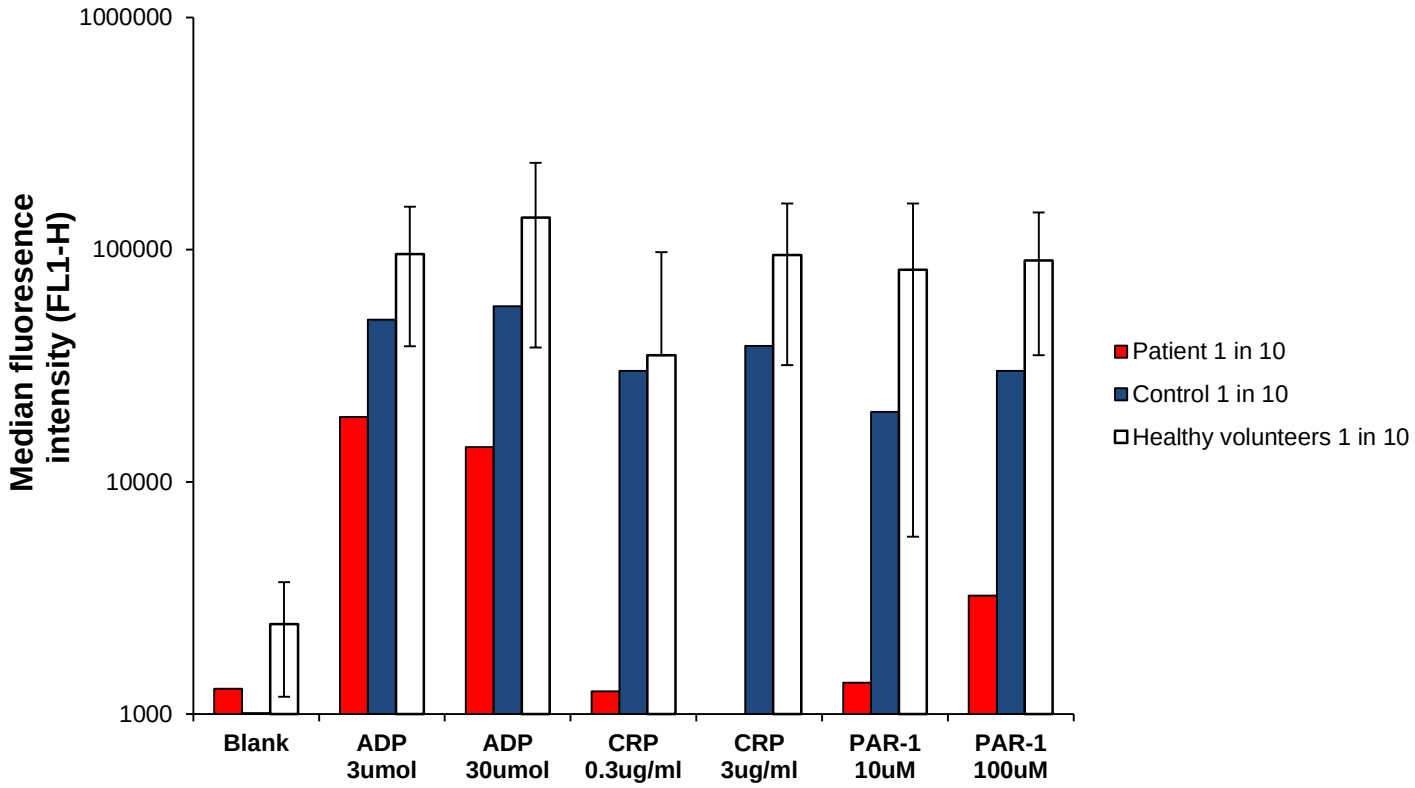


Figure 3.8.5 Representative example of defect in fluorescent fibrinogen binding in flow cytometry analysis (Patient 58.I). In response to all doses of agonist, median fluorescent intensities for patient agonist stimulations (red) are reduced when compared to the travel control (blue).

Surface receptor levels were disrupted in a total of five patients. Two patients, 13.I and 61.I, showed marked reductions in surface receptor levels of CD41, encoded by *ITGA2B* (Figure 3.8.6). Two patients showed a marked reduction to around 50% of the travel control in median fluorescent cell surface levels of CD42b, encoded by *GP1BA* (Figure 3.8.6). One patient, 51.I, showed a mild reduction in cell surface receptor levels of GPVI in addition to a phenotype with a suggested alpha granule secretion defect (Figure 3.8.6).

Two patients showed agonist-specific defective responses when analysed. Patient 61.I showed a reduction in P-selectin expression and fluorescent fibrinogen binding, specifically when stimulated with both low and high dose CRP (0.3µg/mL and 3.0µg/mL, respectively). This phenotype was observed in addition to a reduction in cell surface CD41 receptor levels. Patient 65.I also presented with a unique reduction in response to PAR-1 peptide at both low and high doses (10 and 100µM). Like patient 61.I, the defect was observed, upon stimulation, in both P-selectin expression and fluorescent fibrinogen binding.

Patient 4.II was tested by flow cytometry only due to a low platelet count in PRP that did not exceed the cut off for LTA. As a potential defect in the cyclooxygenase pathway was observed in the affected sibling, patient 4.I, the patient was also analysed by Optimul® 96 well platelet aggregometry assay that was being tested in the laboratory at the time of patient recruitment. Indeed the patient shared the same reduced aggregation in response to doses of arachidonic acid as her affected sibling. The patients flow cytometry results were normal.

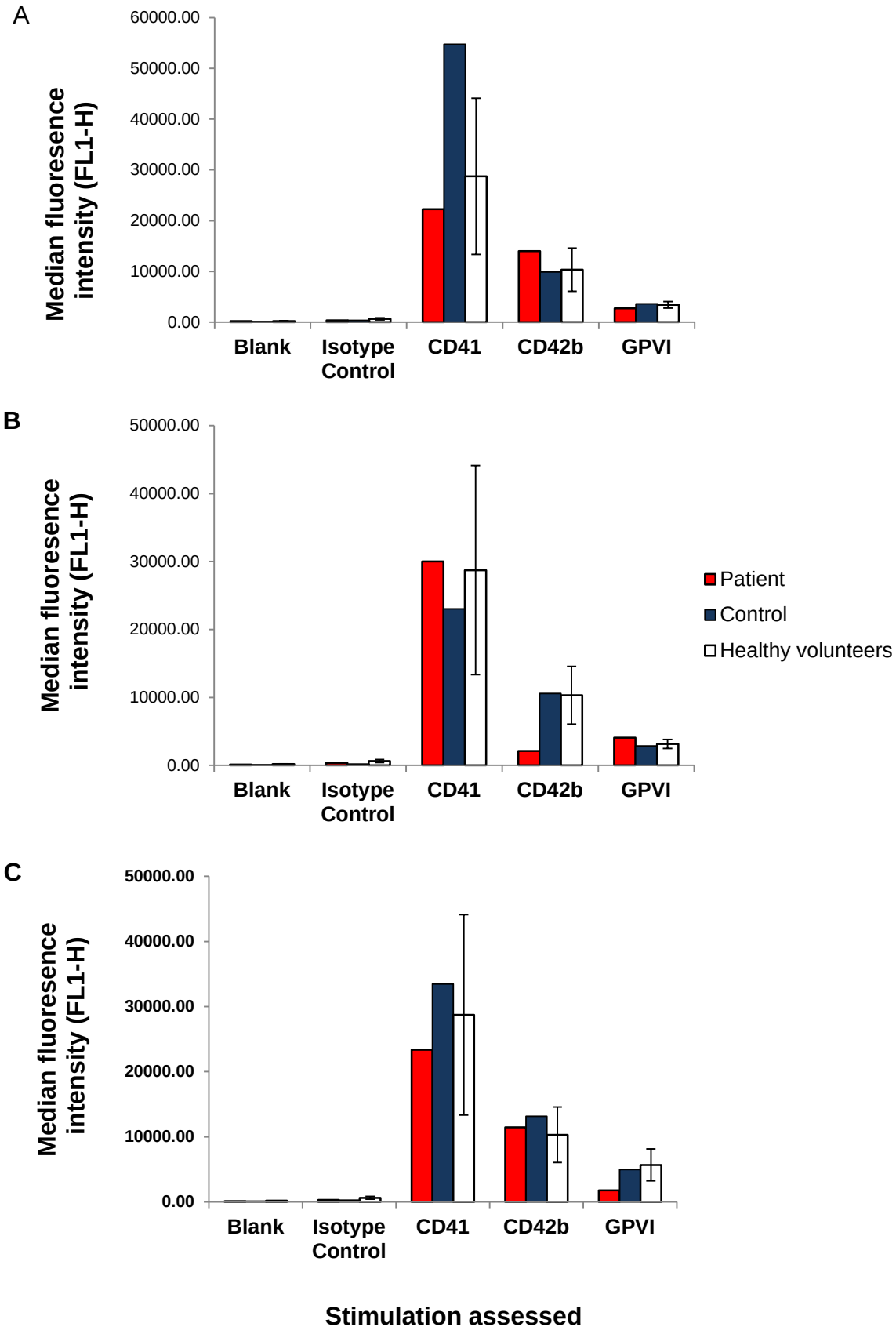


Figure 3.8.6 Examples of a reduction in cell surface levels of: (A) CD41 (Patient 61.I), (B) CD42b (Patient 50.I), and (C) GPVI (Patient 51.I). In response to all doses of agonist, median fluorescent intensities for patient agonist stimulations (red) are reduced when compared to the travel control (blue).

3.9. Discussion

This section covers the patients' characteristics, phenotypes and the presence or absence of secondary qualitative defects in addition to a reduction in platelet count.

Patient characteristics were in line with a multicentre UK based study. Patient age varied considerably among all consenting patients due to enrolment of consenting patients from both paediatric and adult haematological centres and inclusion criteria of <85 years of age. Due to the relatively silent nature of many cases of IT that only present upon haemostatic challenge, the average age of patients included within the study is significantly into adulthood at 30 years. As such the 25th percentile lies at 16 years and the 75th percentile at 39 years. This is in line with previous studies into IT with clinical detailing (Perez Botero, Pruthi et al. 2016). It therefore could be predicted that the paediatric cases that present with IT before the age of 10 would be of a severe life-threatening composition. However, only two of the 12 patients aged under the 10 were regularly receiving intensive therapy to combat bleeding episodes. In addition, only four patients were recruited with related affected adult family members.

A slight skewing towards a higher percentage of female patients (63%) to male patients (37%) was observed and this is in line with previous large scale studies into the patient characteristics of IT (Noris, Biino et al. 2014). Due to the increased prevalence of regular haemostatic challenge in women of menstrual age, this is of no surprise.

Most patients suffered from mild bleeding symptoms relating to their IT. As mentioned, cutaneous bleeding often leading to bruising was the most commonly

described symptom in patient provided questionnaires (example in appendix). In rare cases bleeding episodes were potentially life threatening; however, these were often in response to extensive haemostatic pressure such as in childbirth and surgery. Many of the patients who suffered from these events fall below the $30x10^9$/L platelet cut off, above which blood loss as a result of haemostatic challenge rarely requires treatment (Burrows and Kelton 1990). The prevalence of mild-bleeding related symptoms is therefore considerably higher within our cohort of patients potentially due to the lack of significant haemostatic challenge. However, unlike previous studies into IT, all patients within our cohort displayed a defined bleeding phenotype (Latger-Cannard, Philippe et al. 2016).

Four patients in total presented with non-bleeding symptoms. All patients were recruited under the belief that known causes of IT that are routinely tested for within UK haematological centres have been excluded. As such the presence of congenital cataracts and hearing loss in patient 9.I but a lack of a definitive MYH9-RD or Paris-Trousseau diagnosis is an interesting case to pursue with next generation sequencing.

Treatment for the clinical complications of bleeding episodes followed UKHCDO guidelines and therapy mostly consisted of minimally-invasive methods to stem bleeding and allow clotting to occur naturally, albeit at a slower rate (Bolton-Maggs, Chalmers et al. 2006). In rare cases, pharmacotherapy and blood product transfusions were required to return platelet counts to normal circulating levels and minimise the clinical impact of bleeding episodes. All followed previous management guidelines (Bolton-Maggs, Chalmers et al. 2006). Only two patients required regular transfusions to maintain a healthy circulating platelet count and avoid severe

bleeding episodes. These patients, 41.I and 41.II, were noted for a particular low platelet count and extensive bleeding history despite their young age.

Interestingly, a subset of 10 patients were initially diagnosed with ITP upon discovery of a low platelet count. As mentioned, ITP treatment is often severe, aiming to increase the circulating platelet count through often the first-line use of corticosteroids such as prednisolone (Force. 2003). Prednisolone improves platelet count by suppressing systemic reticulo-endothelial phagocytic function and reducing antibody production but it is often associated with severe adverse effects relating to patient wellbeing and health (Mizutani, Furubayashi et al. 1992; Provan, Stasi et al. 2010). Accurate molecular diagnosis of patients upon presentation of a low platelet count is crucial therefore to avoid unnecessary treatment. Several studies have worked on trying to improve separation of the two disorders dependant on clinical presentation, however as many of the symptoms overlap it remains difficult and may benefit from progressing molecular genetic diagnostic techniques (Drachman 2004). None of the patients recruited to the study initially diagnosed with ITP underwent a splenectomy.

Platelet counts varied from mild to severe forms of thrombocytopenia. Platelet count is often thought to be correlated to the genetic aetiology of disease and this will be determined in the following section (Chapter 4). The range in platelet counts matched previous studies into the platelet diameters of patients with known forms of IT (Noris, Biino et al. 2014). Platelet counts are often misinterpreted in cases of macrothrombocytopenia. Indeed, in our cohort, cases of increased platelet size are often associated with a severely low platelet count. Platelet counts determined by impedance or optical analysis are often clouded by the inclusion of reticulocytes in

cases of giant platelets. However, analysis using the Sysmex XN-1000 PLT-F channel should circumvent any underestimation by determining the platelet count specifically by the binding of a fluorescent RNA specific dye.

Unlike previously published studies, patients with a platelet count between 150-200x10^9/L in whole blood were retained for analysis with our study if they met specific criteria. Platelet counts are unique to an individual and vary considerably among family members and also due to ethnicity, age and gender. It is therefore possible that a reduction in platelet count due to genetic mutation can occur in an individual in whom a normal circulating platelet count would be within the upper region of normal if unaffected. If such an individual is affected by a defect within platelet production, function or consumption, then their platelet count would be reduced but may not lie below the traditional cut-off value of 150x10^9/L. Patients with a borderline count between 150-200x10^9/L, therefore, may still provide us with insight into the genetic aetiologies of IT and are an interesting subset of patients yet to be studied.

Mean platelet volumes of patient platelets were also within normal ranges previously reported in patients with known forms of IT (Noris, Biino et al. 2014). MPV is often an indicator of the genetic cause of disease, with macrothrombocytopenias often related to mutations in cytoskeletal regulatory genes and microthrombocytopenias usually confined to disorders such as WAS and CYCS-related thrombocytopenia. Our cohort consists of 19 patients with an increase in MPV and three with a decrease below a range of normal healthy volunteers.

Immature platelet fractions are a non-invasive indication of platelet lifespan. An increase in IPF is often associated with a decreased platelet survival which can be

directly related to an increased consumption or defect in platelet maturation. 17 patients presented with an increased IPF above a reference range established in 40 healthy volunteers. IPF was consistent across affected family members when available and was often correlated with an increase in MPV and a severely reduced platelet count.

Although many patients within our cohort suffered from a moderate to severe reduction in platelet count, platelet counts were often disproportionate to the bleeding phenotype observed. Secondary qualitative defects in platelet function are often additive to a patients bleeding phenotype; increasing their risk of bleeding episodes and the severity and duration of haemostatic events.

A secondary qualitative defect in platelet function was observed in a total of 71% of patients (61/86) that were tested by LTA, flow cytometry or a combination of both assays. As mentioned previously, the most common platelet function defect was within either alpha or dense granule secretion. Secretion plays a crucial role in the amplification of activation within platelets. Without a functioning secretion mechanism, platelets have the ability to bind to exposed subendothelial matrix proteins and tether, but lack a positive feedback loop to aid in the formation of a stable platelet-fibrin plug/clot. This defect in the positive feedback mechanism of platelets may explain the increased bleeding risk and duration of haemostatic events observed in our cohort of patients when compared to their platelet counts.

A defect in dense granule secretion is often a secondary qualitative defect observed in patients with genetic variants in genes such as *RUNX1* and *FLI1.* Conversely, defects in alpha granule secretion measured by flow cytometry, may be an indicator

of a reduced platelet alpha granule content associated with Grey platelet syndrome and variants in *GFI1B* and *NBEAL2.* Due to the similar presentation in LTA of patients with secretion defects as those who have ingested aspirin prior to analysis, it is difficult to confirm or deny the defect without performing genetic sequencing.

Variants in *ITGA2B* and *GP1BA* are associated with GT and BSS, respectively. Variants in both genes can cause a reduction in platelet count and a disruption to platelet function through the production of non-functional, or the lack of functional presentation of, cell surface receptors. Both disorders are often highly syndromic and easily identified either through routine analysis or a detailed patient clinical history and therefore are unlikely to present within our study. However, four patients have been noted with reductions in the presentation of the associated cell surface receptors as indicated by flow cytometry. This is a strong marker of both diseases but requires genetic confirmation to be able to claim causality.

3.10. Further work

As platelet counts vary within an individual, multiple counts over a pre-defined period would be the optimum way of determining a patient's platelet count. The same can be said for measurements of platelet diameters and morphology. For patient samples this is often a difficult feat both logistically and temporally. Different methods of platelet diameter measurement can also introduce unwanted variance within a sample set. However, by extending the parameters of a study to be more inclusive to those with borderline counts and morphology one can mitigate the risk of excluding patients that are not within a pre-defined cut-off on the day of analysis.

Without a gold standard platelet function test it is often difficult to determine the presence or absence of secondary qualitative defects in each individual. Further testing such as including broad function assays such as microchip flow chamber systems or the PFA-100 to analyse thrombus formation under shear could provide a unique tactic to furthering platelet phenotyping (Harrison 2005; Hosokawa, Ohnishi et al. 2011). In addition, further conformational tests could be utilised after initial testing such as transmission electron microscopy for patients with suspected alpha and dense granule deficiencies.

Although furthering our knowledge of platelet function can often help to describe a patient's phenotype and explain the clinical bleeding diathesis observed, it lacks an ability to clinically diagnose a patient. This is especially apparent in a cohort of patients with a presumed disorder of unknown genetic aetiology. To fully understand a patient's inherited disease a combination of phenotyping and genotyping is required.

3.11. Key findings

- Average patient age and patient gender distributions are in line with previously published studies surrounding haematological bleeding disorders
- Patients present with broad ranging platelet counts and variance in platelet diameters and morphology consistent with the heterogeneity of IT.
- 70% of patients display with a secondary qualitative defect in platelet function which may explain the disproportionate bleeding observed in patients when compared to platelet counts.

4. Patient genotyping by whole exome sequencing

4.1. Summary of background to this research

Due to limitations of platelet function testing discussed in chapter 3, patient genotyping is an ideal way to clinically diagnose patients with IT of unknown aetiology.

Molecular genetic sequencing has progressed rapidly since the invention of Sanger sequencing and the subsequent completion of sequencing of the human genome (Sanger, Nicklen et al. 1977; Venter, Adams et al. 2001; Consortium 2004). Sanger sequencing has evolved away from radiography and gel electrophoresis into utilising fluorescence and capillary array electrophoresis, thereby improving efficiency and efficacy, however it still suffers from limitations in sensitivity. The emergence of next generation sequencing techniques with improved sensitivity has aided in a shift in usage from small scale applications of sequencing areas of disease linkage into large scale studies in whole exome and genome sequencing. This progression has been adopted clinically, allowing for rapid diagnosis dependant on patient genotype in a number of disease areas.

Part of this increase in clinical usage is due to decrease in cost, especially of next generation sequencing techniques such as whole exome and whole genome sequencing. The completion of the reference genome sequence cost nearly $3 billion to complete, and during the past decade the price of genome sequencing has plummeted to an estimated $1000 dollars for a genome at 30x coverage (Warr, Robert et al. 2015). Not only that, but the time taken to sequence and analyse the

subsequent data has also diminished allowing for allocation within a clinical time-frame.

Improvements and the emergence of multiple next generation sequencing techniques has helped research into IT progress from a pre-to a post-genomic era, incorporating sequencing into fundamental research. As mentioned previously, this has aided in the discovery of novel causes of disease and has allowed for clarity in the molecular aetiology behind a number of previously well-classified disorders. However, the application of next generation sequencing remains solely in the confirmation of disorders that are already well characterised. The exceptions to this rule occur in single pedigrees with a strong, shared phenotype and often clinical symptoms in addition to bleeding complications. The question arises then whether a disease subset such as IT would benefit from direct next generation sequencing in the form of whole exome sequencing and whether this will aid in the discovery of novel causative genes not previously implicated in disease.

4.2. Aims of this chapter

This results chapter will aim to genotype 95 patients with suspected IT of unknown aetiology that have been entered into the UK-GAPP study. All patients were analysed using WES to try and determine a variant in a gene likely causative of disease. The results will grant insight into a patient's disease, aiding in their overall diagnosis and potentially offering us novel candidate genes with which to pursue functional characterisation.

4.3. Whole exome sequencing overview

Whole exome sequencing was performed on 69 patients following patient phenotyping and platelet function testing as detailed in Chapter 3 and outlined in Figure 3.3.2. One patient, 47.I, was analysed by WES following analysis by the inherited thrombocytopenia specific next generation sequencing panel. This patient will be discussed in further detail in both chapters and the rationale for analysing the patient using both methodologies will be explained in section 5.7.

4.3.1. Development of a thrombocytopenia gene specific bioinformatics pipeline

A bioinformatic pipeline was created to determine candidate variants from WES data (Figure 4.3.1). Alignment, annotation and variant calling as well as novelty determination were performed as described in section 2.13.1 by collaborators in King's College London. Bioinformatics proceeded on variant called files from each individual patient.

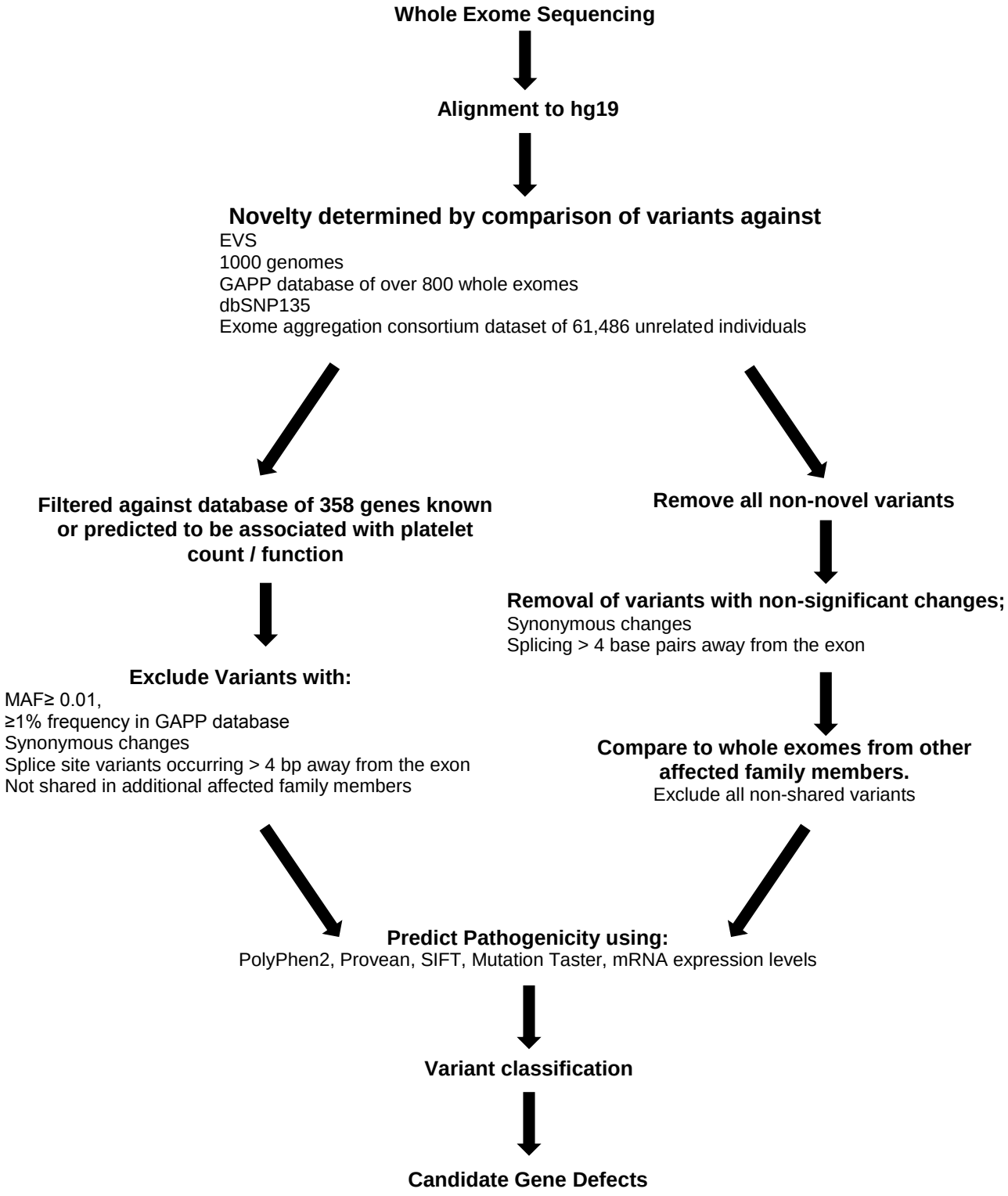


Figure 4.3.1 Bioinformatics pipeline for the analysis of WES data.

Analysis consisted of two methods; comparison to a database of genes known to be implicated in platelet formation, function, lifespan or death (which includes all genes previously known to cause IT) and analysis of novel only variants.

Comparison to a database of 358 known platelet related genes was performed initially (genes outlined in Table A within the appendix). Prior to comparison to a database of platelet related genes, variants were selected to be filtered dependant on frequency. A traditional rare variant cut off value of 0.01 or 1% of the population was used to filter out variants, variants with a MAF of >0.01 are not excluded but “rare” frequency variants are preferentially analysed first by pathogenicity prediction and classification. Also, synonymous variants not known to change the amino acid were excluded.

The database of genes was developed by members of the UK-GAPP study and is an extended version of the one first used by Stockley *et al,* (Stockley, Morgan et al. 2013). The initial list of genes was formulated based upon genes known to be involved with platelet disorders, those implicated in platelet formation and lifespan, and candidate genes from animal models and *in vitro* studies. The list has been updated following the literature and also encompasses genes from an *N*-ethyl-*N*-nitrosourea (ENU) based mutagenic screen into platelets as performed by our collaborator Professor Ben Kile.

Average coverage of all genes within the database in all exomes analysed is displayed in Table A within the appendix. Average percentage at 20x coverage was 0% for five cases as the gene IDs were not found within the targeted regions of WES. Eight genes were noted with 20x coverage <50%, this included one gene, *GP1BB*,

that has previously been identified as causing IT. Therefore, this gene was checked manually due to the overall low coverage. Overall, average coverage across all 358 genes was 89.2%.

Following comparison to the panel of platelet related genes, variants were excluded dependant on location and frequency. Variants predicted to affect the splice site occurring >5bp from the intron-exon boundary were excluded. Variants with a frequency of >0.01 or 1% in the in house UK-GAPP exome database were also excluded.

Candidate variants that met the inclusion criteria were then scrutinised for presence in affected family members, if available. Variants not present in all affected family members were also excluded.

Following comparative analysis to a panel of platelet related genes, exomes were analysed by selecting candidates from novel variants only.

Exomes were initially filtered to display only novel variants and subsequently filtered to remove splice site variants occurring >4bp from the intron-exon boundary and synonymous variants.

Like with the variants that were present within the panel of platelet related genes, novel variants that met the inclusion criteria were scrutinised for presence in affected family members who had also undergone WES. Variants that were present in all affected family members were pooled as candidates with those from the comparative analysis to a panel of platelet-related genes.

All candidate variants subsequently underwent *in silico* pathogenicity predictions to rank variants as the most likely to be disease causing. Variants occurring in genes known to previously cause IT were preferentially selected.

All candidate variants were subsequently classified as stated in section 2.14.1.3.

4.3.2. Average number of sequence variants, sensitivity and false discovery rate

WES revealed, on average, between 24,000 and 26,000 variants (single nucleotide variants (SNVs), small scale insertions/deletions, and splice site variations) in the DNA from each patient. Of these variants, an average of 201 per exome analysed were novel in the databases scrutinised (n=69). Per individual, an average of 2531 variants with a MAF of ≤0.01 were observed excluding synonymous variants.

Over 99% sensitivity and an approximate 3% false discovery rate was found by evaluating the specificity of the pipeline in calling small variations (SNVs, small scale insertions/deletions, splice site changes).

4.3.3. Coverage of exome sequencing data

Fold coverage across the entire targeted region was available for 67 patients. Patients 7.I and 7.II were unavailable for coverage analysis due to the date in which the patients were analysed by WES. On average, 70.6% of all reads were mapped to target regions, 76.6% were mapped to target reads plus 150bp. An average of 91.8% of accessible target bases were covered with at least 20 overlapping reads. The overall mean coverage of targets across all 67 patients was 119.0.

4.3.4. Copy number variants using exome data

CNVs were analysed using a bioinformatics pipeline similar to the analysis of small scale variants such as SNV, small insertions/deletions and splice site variants. CNV were initially compared to determine presence of variants within genes of the same panel of 358 platelet related genes that small variants are analysed with. Variants occurring within one of the genes within the panel of platelet related genes were then scrutinised for their novelty.

Copy number data was available for 44 exomes and unavailable for 25. On average 185.7 variants were observed across all exomes where CNV data was available. Fifteen patients were observed to have no CNVs in the genes within the panel of 358 platelet related genes.

Only two patients, 22.I and 45.I, displayed variants in genes previously known to cause IT. Patient 22.I was observed with a previously identified 18,455bp deletion (Accession number – nsv829233) in *MKL1* in addition to a 21,001bp duplication in *ANKRD18A* (nsv1020605) and a 19,898bp duplication in *PDPK1* (esv2422427). All three variants have been previously identified and the variant in PDPK1 forms part of a segmental duplication.

Patient 45.I, was observed to have two CNVs in genes previously implicated in IT. A previously unidentified duplication in *ANRKD26* was noted that spans 7,893bp. Also a 4,388bp known duplication in *FLNA* (nsv519042) was also observed.

4.3.5. Variants in known thrombocytopenia causing genes

Following bioinformatic analysis as outlined in section 4.3.1, 71% of patients (49/69) were observed to harbour a variant within a gene previously implicated in IT. This

corresponds to a potentially determined genetic aetiology of disease in 66% of index cases (31/47). All variants exceeded 30x coverage at the point of variation and passed all quality control (QC) that was performed externally by collaborators at King's College London.

All variants were present within genes of the panel of 358 platelet related genes. On average, 36 variants excluding synonymous variants, with a frequency of ≤0.01 in the EVS database were noted in the panel of 358 platelet related genes (n=69). On average, four variants were deemed significant (frequency of ≤0.01 in the in house UK-GAPP exome database and <4bp from the intron-exon boundary for splice site variants) per exome analysed (n=69). Of these, variants in genes previously known to cause IT were deemed the most likely candidates and are displayed in Table 4.3.1.

Family	Patient	Gene(s)	Genomic variation	Protein effect	Variation type	Prevalence
1	I	ACTN1	c.2647G>C	p.Gly883Arg	Missense	Novel
2	I	ANKRD26	c.-126T>G		5'-UTR variation	Known
3	I	CYCS	c.104G>T	p.Gly35Val	Missense	Novel
4	I	CYCS	c.155C>T	p.Ala52Val	Missense	Novel
	II	CYCS	c.155C>T	p.Ala52Val	Missense	Novel
	III	CYCS	c.155C>T	p.Ala52Val	Missense	Novel
5	I	DIAPH1	c.3637C>T	p.Arg1213*	Nonsense	Known
	II	DIAPH1	c.3637C>T	p.Arg1213*	Nonsense	Known
6	I	FLI1	c.992_995del	p.Asn331Thr fs*4	Frameshift deletion	Novel
	II	FLI1	c.992_995del	p.Asn331Thr fs*4	Frameshift deletion	Novel
7	I	FLI1	c.1028A>G	p.Tyr343Cys	Missense	Novel
	II	FLI1	c.1028A>G	p.Tyr343Cys	Missense	Novel
8	I	GFI1B	c.503G>T	p.Cys168Phe	Missense	0.0006011 (ExAC)(rs527297896)
	II	GFI1B	c.503G>T	p.Cys168Phe	Missense	0.0006011 (ExAC)(rs527297896)
	III	GFI1B	c.503G>T	p.Cys168Phe	Missense	0.0006011 (ExAC)(rs527297896)
9	I	GFI1B	c.814+1G>A		Splicing	Novel
10	I	GFI1B	c.814+1G>A		Splicing	Novel
		STIM1	c.1828G>A	p.Ala610Thr	Missense	0.00019 (1k)
	II	GFI1B	c.814+1G>A		Splicing	Novel
		STIM1	c.1828G>A	p.Ala610Thr	Missense	0.00019 (1k)
11	I	GP1BA	c.413G>T	p.Gly138Val	Missense	Novel
	II	GP1BA	c.413G>T	p.Gly138Val	Missense	Novel
12	I	GP1BA	c.1761A>C	p.Gln587His	Missense	0.00043 (EXaC) (rs570515282)
13	I	ITGA2B	c.2033C>T	p.Ala678Val	Missense	0.0001848 (ExAC)(rs200481952)
		ITGA2B	c.2176A>T	p.Lys726*	Nonsense	Novel
		THPO	c.609insG	p.Glu204Gly fs*123	Frameshift insertion	Novel
14	I	MYH9	c.3493C>T	p.Arg1165Cys	Missense	Known (rs80338829)
15	I	NBEAL2	c.1376delT	p.Leu459Arg fs*13	Frameshift deletion	Novel
		NBEAL2	c.6893A>G	p.Asn2298Ser	Missense	Novel

16	I	RUNX1	c.16G>A	p.Asp6Asn	Missense	Novel
	II	RUNX1	c.16G>A	p.Asp6Asn	Missense	Novel
	III	RUNX1	c.16G>A	p.Asp6Asn	Missense	Novel
17	I	RUNX1	c.236G>A	p.Trp79*	Nonsense	Novel
		ITGB3	c.349C>T	p.Arg117Trp	Missense	Novel
18	I	RUNX1	c.270+1G>T		Splicing	Novel
	II	RUNX1	c.270+1G>T		Splicing	Novel
19	I	RUNX1	c.270+1G>T		Splicing	Novel
20	I	RUNX1	c.322G>A	p.Gly108Ser	Missense	Novel
21	I	RUNX1	c.427+1G>T		Splicing	Novel
	II	RUNX1	c.427+1G>T		Splicing	Novel
22	I	RUNX1	c.505A>G	p.Thr169Ala	Missense	Novel
23	I	RUNX1	c.512A>T	p.Asp171Val	Missense	Novel
24	I	RUNX1	c.530G>A	p.Arg177Gln	Missense	Known
	II	RUNX1	c.530G>A	p.Arg177Gln	Missense	Known
25	I	RUNX1	c.568G>A	p.Gly190Arg	Missense	0.00002475 (ExAC)
		DIAPH1	c.1820insCTCCTC	p.Thr607delinsTTP	Nonframeshift insertion	Novel
26	I	SLFN14	c.652A>G	p.Lys218Glu	Missense	Novel
27	I	SLFN14	c.657A>T	p.Lys219Asn	Missense	Novel
	II	SLFN14	c.657A>T	p.Lys219Asn	Missense	Novel
28	I	SLFN14	c.659T>A	p.Val220Asp	Missense	Novel
	II	SLFN14	c.659T>A	p.Val220Asp	Missense	Novel
	III	SLFN14	c.659T>A	p.Val220Asp	Missense	Novel
29	I	TPM4	c.548C>T	p.Ala183Val	Missense	Novel
		TUBB1	c.726C>G	p.Phe242Leu	Missense	Novel
30	I	TUBB1	c.721C>T	p.Arg241Trp	Missense	0.0001071 (ExAC)(rs368923302)
	II	TUBB1	c.721C>T	p.Arg241Trp	Missense	0.0001071 (ExAC)(rs368923302)
31	I	TUBB1	c.1080_1081insG	p.Leu361Ala fs*19	Frameshift insertion	Novel
32	I	Unknown				
33	I	Unknown				

34	I	Unknown
35	I	Unknown
36	I	Unknown
	II	Unknown
	III	Unknown
37	I	Unknown
	II	Unknown
38	I	Unknown
39	I	Unknown
40	I	Unknown
41	I	Unknown
	II	Unknown
42	I	Unknown
43	I	Unknown
44	I	Unknown
45	I	Unknown
46	I	Unknown
47	I	Unknown

Table 4.3.1 Results of variants in known IT causing genes following WES analysis in 47 index cases covering 69 patients. Displayed on the patient identifiers as well as candidate genetic variants in genes known previously to cause IT if available. Shown are the genomic and protein effect of variation, type of variation and the prevelance in the previously listed databases.

In total, 36 variants were observed in 17 genes known previously to cause IT. Twenty-five index cases presented with variants in one single gene previously known to cause IT. Five index cases had two variants in genes known to previously to cause IT and one patient, 13.I, had three variants. Candidate variations were present within; *ACTN*, the 5'-UTR of *ANKRD26*, *CYCS*, *DIAPH1*, *FLI1*, *GFI1B*, *GP1BA*, *ITGA2B*, *ITGB3*, *MYH9*, *NBEAL2*, *RUNX1*, *SLFN14*, *STIM1*, *THPO*, *TPM4* and *TUBB1*.

Of the 47 index cases, five index cases presented with two variants in known thrombocytopenia causing genes. Four of these index cases displayed variants in two alternate genes. These were as follows: patient 10.I (and also the related family member 10.II) - *GFI1B*; c.814+1G>A and *STIM1*; c.1828G>A, p.Ala610Thr. Patient 17.I - *RUNX1*; c.236G>A, p.Trp79* and *ITGB3*; c.349C>T, p.Arg117Trp, patient 25.I – *RUNX1*; c.586G>A, p.Gly190Arg and *DIAPH1*; c.1820insCTCCTC, p.Thr607delinsTTP and patient 29.I – *TUBB1*; c.726C>G, p.Phe242Leu and *TPM4*; c.548C>T, p.Ala138Val. One patient displayed two variants in one gene, this was patient 15.I and the variants were both located within *NBEAL2*. The variants were compromised of a frameshift causing deletion; c.1376delT, p.Leu459Arg fs*13 and a missense variant; c.6893A>G, p.Asn2298Ser indicating the recessive inheritance of GPS.

One patient was observed with three variants in two genes previously implicated within IT. Patient 13.I presented with two variants within *ITGA2B*, consisting of a missense variant, c.2033C>T, p.Ala678Val, and a nonsense causing SNV, c.2176A>T, which is predicted to truncate the coded protein sequence; p.Lys726*. In addition, patient 13.I was also identified with a frameshift causing insertion within *THPO*; c.609insG, p.Glu204Glyfs*123. The predicted effect of the variant is a

frameshift which allows for the subsequent translation of an altered stretch of 123 amino acids and a formation of a new stop codon, truncating the wild type protein sequence by 26 amino acids (p.Glu204Glyfs*123).

4.3.5.1. Protein effects of genetic variants

Of the 36 variants observed within all cases, four variants were identified in non-coding regions of the genes. The most commonly observed variation type was a SNV leading to a missense change in amino acid sequence; this was noted in 64% of all variants observed (23/36). Three SNVs altered the amino acid sequence to introduce a premature stop codon at the point of variation. These nonsense variants occurred in index patients 5.I, 13.I and 17.I with variants in *DIAPH1, ITGA2B* and *RUNX1,* respectively.

The four variants occurring outside of the coding region consisted of three splice site variants all located within 1bp of the intron-exon boundary, and one variant upstream of the coding sequence within the 5' UTR of *ANKRD26*. This variant was noted initially by a separate project to determine the presence of variants within the 5' UTR of *ANKRD26* and was later confirmed by exome and Sanger sequencing. The variant was not present within the variant called and annotated exome data and was determined by analysing the raw aligned sequence data manually as mentioned in the methods section 2.14.1.1. This variant was the only variant to occur within the 5' UTR of *ANKRD26* in all of our patients analysed where coverage was possible.

Three splice site variants were noted in all patients. These included; a splice site variant in *GFI1B*; c.814+1G>A, which was noted in two unrelated index cases, 9.I and 10.I, as was *RUNX1*; c.270+1G>T which was noted in index cases 18.I and 19.I

and *RUNX1*; c.427+1G>T in index case 22.I. *GFI1B*; c.814+1G>A and *RUNX1*;c.427+1G>T are both predicted to alter the WT donor site affecting normal splicing. *RUNX1*;c.270+1G>T is predicted to activate an intronic cryptic donor site potentially altering splicing.

Five small scale insertions/deletions were observed in the patient dataset. Only one variant, a 6bp in frame insertion within *DIAPH1* in patient 25.I, did not cause a change in reading frame. All other variants are predicted to cause a frameshift truncating the coded protein sequence due to the introduction of a premature downstream codon. These variants are outlined in Table 4.3.1.

4.3.5.2. Novelty of variants found in patients

All variants in index cases were found to be novel in all databases scrutinised with the exception of 10 variants in total. Four of the 10 variants have been previously implicated in disease and have been determined causative of IT. These variants occurred within; patient 2.I: *ANKRD26*;c.-126T>G, patients 5.I and 5.II: *DIAPH1*;c.3637C>T, patient 14.I: *MYH9*;c.3493C>T and patients 24.I and 24.II: *RUNX1*;c.530G>A. All variants were initially analysed for novelty in the EVS, the 1000G project and our in-house UK-GAPP database of over 1200 exomes as depicted in Figure 4.3.1. Candidate variants that were not excluded during bioinformatics post sequencing analysis were also scrutinised against the ExAC database for prevalence. The ExAC database may include data from individuals with low platelet counts who were either undiagnosed or recruited through an unrelated study. Six variants were observed in patients that were not novel and were present in

the databases outlined in Table 4.3.1. All variants were present at a frequency of <0.0005 in their relative databases.

On average, less than one novel variant was expected to be observed in the causative genes in which variants were observed. This was calculated by determining the percentage of the overall target exome that a targeted gene occupies and applying the average number of novel variants. The number of variants occurring in genes known to cause IT also exceeds the expected number when extending the analysis to cover variants with a MAF of ≤0.01.

4.3.5.3. *In silico* pathogenicity prediction and conservation

Pathogenicity was predicted for all variants using the *in silico* prediction software as outlined within the bioinformatics pipeline in section 4.3.1. Results of pathogenicity predictions are shown in Table 4.3.2.

Variants predicted to alter the length of the encoded transcript (insertions, deletions and nonsense variants) as well as variants predicted to effect splice sites were not analysed by SIFT, PROVEAN and PolyPhen-2 software. Results were unavailable from using these software due to limitations in the software in predicting only SNV that lead to missense changes. Splice site variant pathogenicity was predicted as mentioned previously.

Conservation was determined by two methods and are displayed in Table 4.3.2. Conservation was in agreement between both methods used in 83% of variants (30/36)

Family	Patient	PhyloP	Phastcons	Mutation taster	PolyPhen-2	SIFT	Provean	ACMG criteria	Classification
1	I	5.969	1	Disease causing	Damaging	Damaging	Deleterious	PM2, PP2, PP3, PP1, PP4	**Likely Pathogenic**
2	I	-0.433	0	Polymorphism				PS1, PS3, PP1, PP5, PP4	**Pathogenic**
3	I	5.962	1	Disease causing		Damaging	Deleterious	PM2, PP3	**Uncertain significance**
4	I II III	5.962	1	Disease causing	Benign	Damaging	Deleterious	PM2, PP3, PP2, PS (segregation), PP4	**Likely Pathogenic**
5	I II	1.798	1	Disease causing				PM4, PS1, PVS1, PP (segregation)	**Pathogenic**
6	I II	4.943	1	Disease causing				PM2, PVS1, PP (segregation), PP4	**Pathogenic**
7	I II	3.302	0.995	Disease causing	Damaging			PM2, PP3, PP2, PP (segregation), PP4	**Likely Pathogenic**
8	I II III	5.229	1	Disease causing	Damaging	Damaging	Deleterious	PP3, PM (segregation)	**Uncertain significance**
9	I	3.444	1	Disease causing				PM2, PP3, PP2, PP4	**Uncertain significance**
10	I	3.444	1	Disease causing				PM2, PP3, PP2, PP (segregation), PP4	**Likely Pathogenic**
		-0.084	0.493	Disease causing				PP3, PP2, PP (segregation), PP4	**Uncertain significance**
	II	3.444	1	Disease causing				PM2, PP3, PP2, PP (segregation), PP4	**Likely Pathogenic**
		-0.084	0.493	Disease causing				PP3, PP2, PP (segregation), PP4	**Uncertain significance**
11	I II	0.812	0.208	Polymorphism	Damaging	Damaging	Deleterious	PM2, PP3, PP2, PP (segregation), PP4.	**Likely Pathogenic**
12	I	-0.015	0.812	Polymorphism		Damaging	Neutral	PP2, PP4	**Uncertain significance**
13	I	4.959	0.999	Disease causing	Damaging	Damaging	Deleterious	PP3, PP2, PP4	**Uncertain significance**

		1.419	0.957	Disease causing				PM2, PM4, PVS1, PP4	**Pathogenic**
				Disease causing				PM2, PVS1	**Pathogenic**
14	I	4.131	1	Disease causing	Damaging	Damaging	Deleterious	PS1, PS3, PP2, PP5, PP4	**Pathogenic**
15	I	2.687	0.997	Disease causing				PM2, PM4, PM6, BP2, PP4	**Likely Pathogenic**
		4.508	1	Disease causing	Damaging	Damaging	Deleterious	PM2, PP3, PP2, PM6, BP2, PP4	**Likely Pathogenic**
16	I II III	5.86	1	Disease causing	Damaging	Damaging	Deleterious	PM2, PP3, PM (segregation), PP4	**Likely Pathogenic**
17	I	4.077	1	Disease causing				PM2, PVS1, PM4, PP4, PP3	**Pathogenic**
		1.362	0.454	Polymorphism	Damaging	Damaging	Deleterious	PM2, PP3, PP4	**Uncertain significance**
18	I II	6.077	1	Disease causing				PM2, PM4, PM (segregation), PP4	**Likely Pathogenic**
19	I	6.077	1	Disease causing					
20	I	6.077	1	Disease causing	Damaging	Damaging	Neutral	PM2, PP3, PP4	**Uncertain significance**
21	I II	5.36	1	Disease causing				PM2, PM4, PP (segregation), PP4	**Likely Pathogenic**
22	I	4.429	1	Disease causing	Damaging			PM2, PP3, PP4	**Uncertain significance**
23	I	4.429	1	Disease causing	Damaging	Damaging	Deleterious	PM2, PP3, PP4	**Uncertain significance**
24	I II	5.36	1	Disease causing	Damaging	Damaging	Deleterious	PS1, PM2, PP3, PP (segregation), PP4	**Pathogenic**
25	I	2.485	1	Disease causing	Damaging	Damaging	Deleterious	PP3, PM1	**Uncertain significance**
				Polymorphism				PM2, BP4, PM4	**Uncertain significance**
26	I	2.338	1	Polymorphism	Damaging	Benign	Neutral	PM2, PP (segregation)	**Uncertain significance**
27	I II	0.852	0.952	Polymorphism	Damaging	Benign	Neutral	PM2, PP (segregation)	**Uncertain significance**
28	I II	2.366	0.167	Polymorphism	Damaging	Damaging	Deleterious	PM2, PS (segregation)	**Likely Pathogenic**

	III								
29	I	5.582	1	Disease causing	Damaging	Damaging	Deleterious	PM2, PP3, PP4	**Uncertain significance**
		1.414	1	Disease causing	Damaging	Damaging	Deleterious	PM2, PP3, PP2, PP4	**Uncertain significance**
30	I II	5.701	1	Disease causing	Damaging	Damaging	Deleterious	PP3, PP2, PP (segregation), PP4	**Uncertain significance**
31	I	2.413	1	Disease causing				PM2, PM4, PP4, PP3	**Likely Pathogenic**

Table 4.3.2 Pathogenicity prediction and variant classification of the variants displayed in Table 4.3.1. PhyloP scores vary between -14 and +6 and measure conservation at each individual base, sites predicted to be conserved are assigned a positive score, fast evolving sites are assigned a negative score. PhastCons values vary between 0 and 1 and reflect the probability that each nucleotide belongs to a conserved element. MutationTaster uses a Bayes classifier to predict the effect of a mutation from a feed of classifiers. SIFT damaging prediction score= <0.05. Provean deleterious score = <-2.5. PolyPhen-2 predictions are appraised qualitatively as benign or damaging. The ACMG consensus guidelines, including supporting evidence, are also shown.

4.3.5.4. Variant classification

Variant classification was performed as detailed in the methods section 2.14.1.3 and the results are shown in Table 4.3.2. Each criteria that each variant met is also shown which are then combined allowing for determination of the classification. In total, of the 36 variants occurring within the known IT-related genes; 8 variants were deemed “pathogenic” (22%), 12 were deemed “likely pathogenic” (33%) and 16 were of “unknown significance” (45%).

Of the eight variants deemed pathogenic; four variants in index cases 2.I, 5.I, 14.I, and 24.I were classified as pathogenic as they are previously identified variants known to be detrimental to the function of the protein and causative of disease. The remaining four variants deemed to be pathogenic residing in index cases; 6.I, 13.I (includes two variants, one in *ITGA2B* and one in *THPO*), and 17.I were classified due to the protein size altering effect of the nonsense and frameshift variants.

No variants were classified as benign or likely benign, however one variant, an in-frame insertion within *DIAPH1* (c.1820insCTCCTC), did show supporting evidence of benign classification in addition to supporting pathogenic evidence.

4.3.5.5. A high percentage of variants within RUNT1-related transcription factor

Of the 36 variants identified, nine variants were identified in 10 index cases within the RUNT-related transcription factor (*RUNX1*). The variants consisted of six missense variants, two splice site changes and one nonsense premature stop codon-generating variants. The variants represent the plausible candidate variants

causative of disease in 32% of all index cases (10/31), identified with variants in known IT causing genes. With the exception of a missense variant, c.16G>A, p.Asp6Asn, occurring in patients 16.I, 16.II and 16.III, all variants occur within the highly conserved genetic region encoding the runt homology domain of RUNX1 (Figure 4.3.2).

4.3.5.6. Sanger sequencing confirmation

All variants were confirmed by Sanger sequencing to remove potentially false-positive derived results. Additional family members, both affected and unaffected, that were also recruited to the study were available for families; 4, 26, 27, and 28. In all cases the candidate variants in known IT causing genes segregated with disease status and were not present within any unaffected family members. Example sequencing traces of variants are shown in

Figure 4.3.3.

4.3.6. Potentially damaging variants in novel candidate genes

After scrutinising individuals for variants within the panel of 358 platelet associated genes, focus was turned towards the novel variants especially in individuals without a variant in a gene within the gene panel. In three families, WES analysis revealed potentially damaging candidate variants in genes not known to cause IT at the time of analysis. Six variants were identified in the three families that may be causative of disease (Table 4.3.3). All variants were novel when scrutinised within the EVS, 1000G and our in house GAPP database of over 1200 exomes. Upon further analysis within the ExAC database, only the variants within *ANKRD18A* and *GNE* within family 37 were not previously identified.

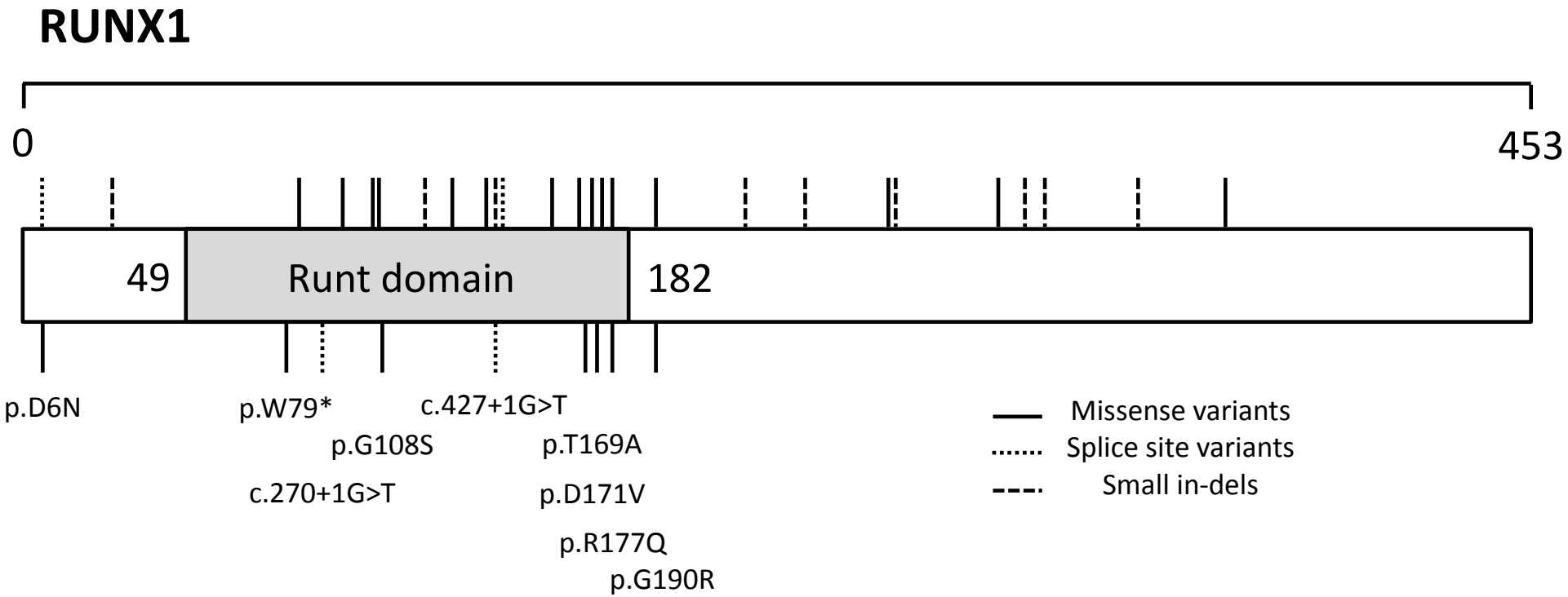
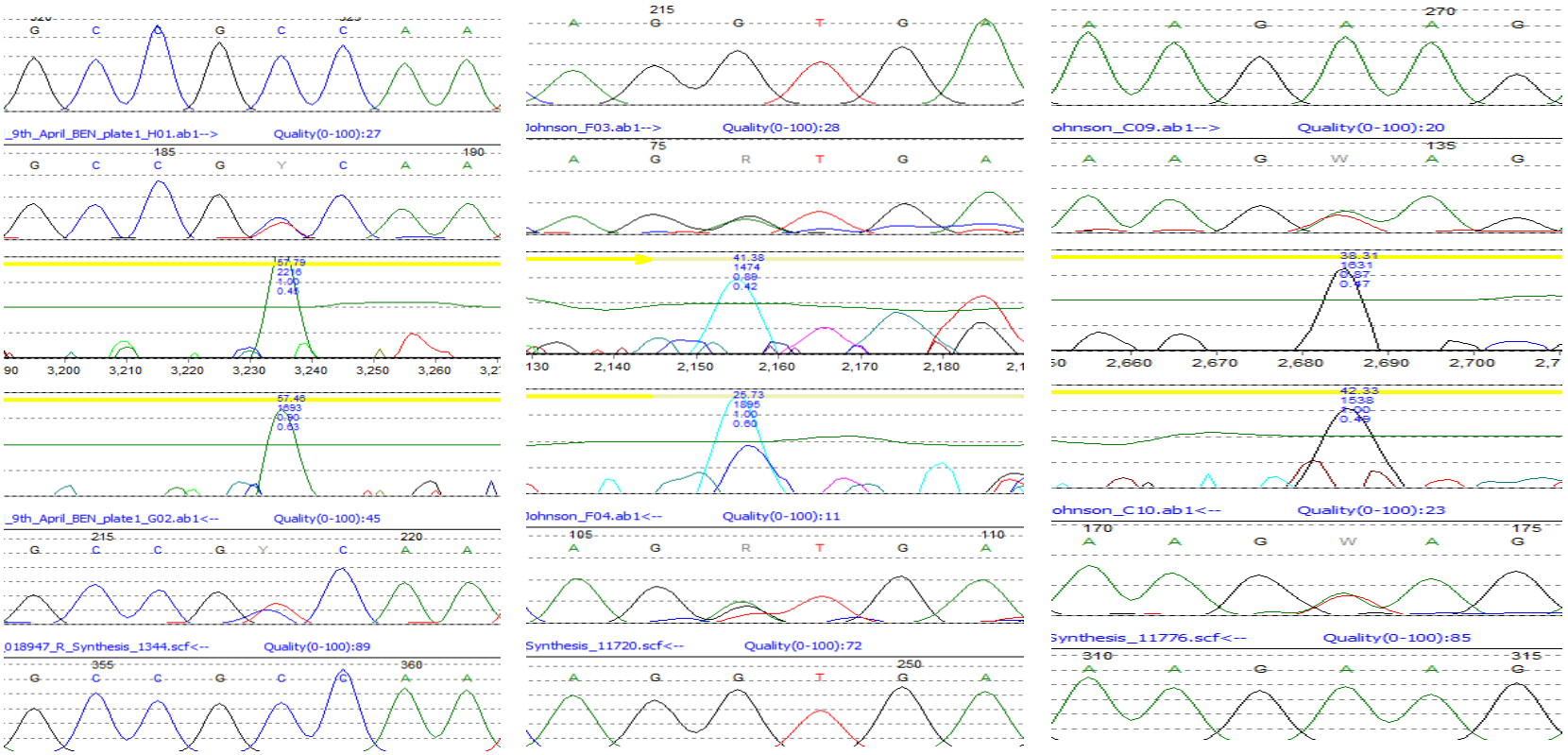


Figure 4.3.2 Schematic showing the protein location of all previously published variants within RUNX1 (Top) compared to all variants identified by WES analysis (Bottom). Numbers refer to amino acid locations and the Runt-homology domain spanning from amino acid 49-182 is also displayed. A high number of variants are located within the Runt-homology domain potentially implying a genotype-phenotype correlation.



Patient 4.I
CYCS
C,155C>T
p.Ala52Val

Patient 9.I
GFI1B
c.814+1G>A

Patient 13.I
ITGA2B
c.2176A>T
p.Lys726*

Figure 4.3.3 Representative sanger sequencing traces confirming three variants identified by WES analysis.

Family	Gene	Variant	Protein effect	Prevalence	PhyloP	PhastCons	Mutation Taster	SIFT	Provean	PolyPhen-2	ACMG criteria	Classification
36	PADI2	c.1496A>G	p.Lys499Arg	0.000008681	1.647	1	D	T	N	B	PM (segregation)	Uncertain Significance
	TTF2	c.3265C>G	p.His1089Asp	1.65E-05	5.131	1	D	D	D	D	PM (segregation), PP3	Uncertain Significance
41	ANKRD18A	c.2395_2397del[hom]	p.Glu799del[hom]	Novel	0.772	0.965	P		D		PM2, PP (segregation), PM6	Uncertain Significance
	GNE	c.1339G>A[hom]	p.Gly447Arg[hom]	Novel	5.343	1	D	D	N	D	PM2, PP (segregation), PM6	Uncertain Significance
	FRMPD1	c.1526C>T[hom]	p.Ala509Val[hom]	0.0003708	-1.459	0	P	T	N	B	PP (segregation), PM6	Uncertain Significance
43	MKL1	c.1723G>A	p.Val575Met	0.0007718	3.358	1	D	D	N	D		Uncertain Significance

Table 4.3.3 Potentially damaging variants in novel candidate genes. When a variant has been previously observed it was annotated in the prevalence column with the database it was observed in. PhyloP and Phastcons scores are shown following the cut offs displayed in Table 4.3.2. Variants are noted as D – Disease causing and P – Polymorphism in MutationTaster, D – Damaging and T – Tolerated in SIFT, D – Deleterious and N – Neutral in Provean, D – Damaging and B – Benign in PolyPhen-2 in silico pathogenicity prediction software.

4.3.6.1. Novel missense candidate variants in *PADI2* and *TTF2*

Two variants in genes not known previously to cause IT were identified in three individuals from family 36. The three patients were part of a larger kindred including four unaffected family members that were also recruited to the study. The two variants were part of 14 novel variants shared between the three family members as derived from WES analysis. Sanger sequencing of all 14 variants within the four unaffected family members narrowed down the candidate variants to only two missense changes; *PADI2*;c.1496A>G and *TTF2*;c.3265C>G. Both variants segregate with disease status, not being present within any of the unaffected related individuals. Further analysis showed that both variants were present within the ExAC database at frequencies of 0.000008681 and 0.00001648, respectively. Both variants are highly conserved at the point of variation, however only the variant within *TTF2* was predicted to be unanimously deleterious by all pathogenicity prediction software tools used. Both variants are classified as of “unknown significance”.

4.3.6.2. Variants within *ANKRD18A*, *GNE* and *FRMPD1* in two related individuals from consanguineous relationships

WES analysis of two related patients (41.I and 41.II) was adapted due to the specific inheritance patterns of the patients. Both patients were born from consanguineous relationships within a single large consanguineous kindred, therefore, analysis was focused on the identification of shared homozygous variants due to recessive segregation of disease. Two missense variants; *GNE*; c.1339G>A, p.Gly447Arg and *FRMPD1*; c.1526C>T, p.Ala509Val were identified in addition to an in-frame deletion of three base pairs; *ANKRD18A*; c.2395_2397del, p.Glu799del. At the time of

analysis, the link between *GNE* and thrombocytopenia was not yet published. Both patients show no signs of the accompanying muscular dystrophy that is attributed to variants within *GNE,* they are therefore considered in this section and not section 4.3.5 regarding variants in genes known to cause IT. Both patients were of South Asian ethnicity so prevalence analysis was adapted to analyse the variants within the database of the same ethnicity. Variants in *ANKRD18A* and *GNE* were absent from South Asian and all ethnic databases, the variant within *FRMPD1* was noted at a frequency of 0.0003708 in the South Asian population present within the ExAC database. The variant within *GNE* was highly conserved, however all variants showed variation in results among the pathogenicity prediction software used and as such all three variants are classified as of "unknown significance" using the ACMG consensus guidelines.

4.3.6.3. Missense variant in the recently proposed IT linked gene; *MKL1*

One individual, patient 43.I, was observed with a rare frequency variant (≤0.01) within *MKL1*. The variant was initially identified as novel and was the only gene of haemostatic relevance of 109 novel variants identified within the patient following WES analysis. The variant; *MKL1*; c.1723G>A, p.Val575Met, has been noted previously at a frequency of 0.0007718 (allele count of 6/7774 in the EXAC database). As *MKL1* is present within the panel of 358 platelet related genes the variant was also identified during the comparison to the panel. However, the variant was observed in a heterozygous state where only one homozygous variant has previously been identified as leading to IT secondary to immunodeficiency. The variant is highly conserved, but pathogenicity prediction is not in agreement and subsequent classification was of "unknown significance".

4.3.6.4. Sanger sequencing confirmation

All novel sequence variants have been confirmed by Sanger sequencing. This includes sequencing within unaffected related individuals for family 36 to help determine segregation of variants with disease status that will be discussed further in chapter 6.

4.4. Discussion

Presented within this chapter is the first large scale application of WES analysis to patients with inherited bleeding diathesis.

A novel pipeline for the analysis of variants called, annotated and aligned sequence data was developed to determine candidate variants. The pipeline was designed to be all-encompassing; allowing analysis of variants that may affect platelet formation, function and lifespan as well as assessing novel variants in patients where there are not likely to be candidate genes. The average incidence of IT has most recently been estimated to be approximately 270 cases per 1 million live births (Balduini 2014). The bioinformatics pipeline uses a traditional rare “cut off” value of 100 cases per 1 million live births or a prevalence of 1% (0.01) within the population databases tested. Although the incidence is believed to be increasing, potentially due to the rise of genetic diagnosis, the pipeline retained this value for two reasons:

1) That on average, the population databases used within post-annotated bioinformatics often understates the prevalence of rare variants. This is most evident in the number of variants that are deemed novel within dbSNP135, EVS, 1000G and our in-house database of over 1200 exomes

but have been previously identified in larger scale studies such as the ExAC database.

2) That variants with a MAF exceeding 0.01 that are in genes of the panel of 358 platelet related genes are not excluded but are considered less likely and analysed further if a potential candidate with a low frequency is not identified.

Splice site variants were excluded if they exceeded 4bp from the intron-exon boundary. The vast majority of splice site variants are believed to occur at positions ±1 and 2 from the intron-exon boundary (Krawczak, Reiss et al. 1992). At these positions, splice site variants are likely to be detrimental to the length of the protein sequence by exon skipping or intron inclusion (Richards, Aziz et al. 2015). By expanding our analysis to include splice site variants up to and including those that occur 5bp from the intron-exon boundary, we aim to include the vast majority that may impact protein function and filter out the variants more likely to not have a functional effect.

The panel of 358 platelet-related genes has been developed within the UK-GAPP study drawing on data from published literature, animal models of disease and expert knowledge. This panel of genes has grown throughout the lifespan of the UK-GAPP study in line with recent advances in the knowledge of IT. Although limiting the broad scope of WES, it creates a strong starting point for the determination of candidate variants. Overall average 20x coverage was relatively high for the genes in question but was impacted by a failure of the software to recognise five gene IDs effectively scoring each gene with coverage of 0%. 20x coverage in genes known to previously cause IT exceeded 75% with the exception of *GP1BB*. As variants in areas of low

coverage are often removed following QC, *GP1BB* was checked manually using integrated genomics viewer (http://software.broadinstitute.org/software/igv/ Last accessed October 2016) for all exomes and no variants were identified. One caveat of using a cut off of 20x coverage is the possibility that variants may be missed, particularly in those in difficult to amplify regions such as the first exon and intron/exon boundaries. When lowering the coverage to less than 20x a much higher percentage of each gene previously known to cause IT was identified reducing the possibility of missed variants.

Copy number variants occurring in genes previously known to cause IT were only found in two patients after analysis of CNV data using the ExomeDepth tool. The 18,455bp variant identified in *MKL1* (nsv829233), found in patient 22.I, was previously identified as part of a large scale study to determine Asian-specific CNVs (Park, Kim et al. 2010). Located on chromosome 22, the variant spans the last 11 coding exons of *MKL1,* which are predicted to be lost by deletion. This variant was first identified by oligo aCGH and is considered a common CNV among the population it was first identified in. No immunodeficiency has been noted within the patient, suggesting the presence of a *MKL1* CNV may not impact protein function causing disease.

Two CNVs were identified in patient 45.I. No observable SNVs, small scale insertions/deletions or splice site variants in genes from the panel of 358 platelet related genes were observed within this patient. WES analysis revealed a previously identified duplication within *FLNA* and an unidentified duplication within *ANKRD26.* The variant within *FLNA* has been previously identified in a large scale SNP array to determine human CNV variants (Shaikh, Gai et al. 2009). It occurs in a region with

complex rearrangements in addition to a previously identified CNV duplication. The variant in *ANKRD26* occurs within a region spanned by previously identified copy number deletions but a duplication has yet to be observed. The duplication is predicted to effect internal exons of *ANKRD26*, to date only variants within a small stretch of the upstream activating sequence of *ANKRD26* have been reported to cause IT. These variants elicit an effect by disrupting the normal silencing of *ANKRD26* expression during late megakaryopoiesis. Theoretically, to elicit an effect similar to variants within the 5'UTR, coding sequence variants would need to extend the relative lifespan of ANKRD26 so that a persistent effect on the TPO/MPL signalling pathway can lead to impaired proplatelet formation. Predicted duplications to internal segments of the gene would most likely lead to an alternately transcribed protein. This may or may not have an impact on *ANKRD26* based signalling but further work would be needed to determine this.

Overall, genetic variants within previously known IT causing genes were identified in 69% of index cases by WES analysis. The majority of variants were missense and most variants were novel and not previously identified in the databases scrutinised. Missense variants were expected to be most common due to relative impact on the expressed protein; missense variants are generally tolerated by retaining expression and partial function of the affected allele.

When considering pathogenicity, WES analysis revealed a positive prediction of pathogenicity (classified "pathogenic" or "likely pathogenic" in a gene consistent with the patients phenotype and zygosity consistent with expected inheritance) in 43% of index cases (20/47). Twenty-five percent of the index cases (12/47) were of uncertain/possible pathogenicity (results classified of "uncertain significance" in

known IT causing genes). The remaining 32% of index cases (15/47) had a negative prediction of pathogenicity (no convincing variants identified in known causative genes).

Positive variant discovery rates are comparable to or exceed previous large scale WES clinical multicentre studies of Mendelian disorders (Yang, Muzny et al. 2013; Chong, Buckingham et al. 2015). The number of patients with a proposed genetic aetiology of disease is reduced when compared to previous studies into IT however. This may be a result of using a strict classification system to predict the *in silico* capability of the variants to cause disease. If the analysis is allowed to encompass all patients with a variant in a gene previously linked to IT, our detection rate of 69% far exceeds that of any previous large scale study into IT.

In addition, by only recruiting patients with an unknown aetiology of disease we potentially exclude patients that may have been clinically diagnosed with a form of IT. In particular, this may correspond to patients with known BSS and MYH9-related disorders as these two forms of IT are routinely tested for in many haematological centres within the UK. Because of this, the spectrum of variants generated by WES analysis differs from previous large scale genetic studies into IT. If patients with predicted BSS or MYH9-related disorders by clinical test were included in our analysis, our positive detection rate would most likely be higher as well.

Even though patients with BSS and MYH9-related disorders should not be present within our cohort, three index cases, corresponding to four patients, were identified with variants within genes causative of the two disorders.

Three patients, 11.I, 11.II and 12.I, were identified with variants within *GP1BA* encoding the glycoprotein 1b platelet alpha subunit. The two variants observed were missense changes, one of which was novel. All three patients were analysed by flow cytometry and no reduction in the cell surface expression of CD42b was noted in any patient. Patient 12.I, was phenotypically analysed by both methodologies, however aggregation to ristocetin (RIPA) was not measured. Patient 11.I was the only patient to present with a macrothrombocytopenia, although this was not within the normal range of giant platelets observed in patients with BSS. The patients with variants in *GP1BA* therefore are most likely representative of an atypical presentation of BSS, consistent with the variance in clinical presentation of previously identified variants (Savoia, Kunishima et al. 2014).

One patient, 14.I, was identified with a variant within *MYH9*; c.3493C>T, p.Arg1165Cys. The variant was predicted to cause a missense change in the amino acid sequence and has been previously reported to cause Sebastian syndrome (Seri, Cusano et al. 2000). The patient was tested initially by microscopic assessment of a peripheral blood smear after conventional staining but no leukocyte inclusions were noted. The clinical symptoms of the patient do however correlate with disease with secondary associated congenital cataracts and hearing loss present from an early age. In addition, a marked increase in platelet volume was noted, indicating the presence of macrothrombocytopenia. Döhle-like bodies are only present within 42-84% of individuals with MYH9-related disease, therefore the variant within *MYH9* is likely a positive diagnosis of the genetic aetiology of disease in this patient but the patient presents with an atypical phenotype (Seri, Cusano et al. 2000; Kunishima, Matsushita et al. 2003).

Excluding patients with BSS and MYH9-related disease, two of the more prevalent disorders, has given rise to a relatively large percentage of individuals identified with variants within *RUNX1*. Variants were present in 10 index cases, corresponding to 15 patients. Two thirds of patients with a variant in *RUNX1* presented with a defect in granule secretion, a common attribute of patients with categorised mutations in *RUNX1* (Stockley, Morgan et al. 2013). All variants, with the exception of a missense variant in patients 16.I, 16.II and 16.III, occur within the runt homology domain. The runt homology domain mediates DNA binding and heterodimerization with CBFβ. Previous variants span the entire gene (Figure 4.3.2) but the majority of variants within the runt homology domain have a dominant-negative effect on protein function and are predicted to be associated with a higher propensity to develop leukaemia (Michaud, Wu et al. 2002). To date, haematological malignancies have not been reported in any patients; however, the brother of patient 21.I has a history of AML but was unavailable for testing. Apart from a nonsense causing SNV in patient 17.I, c.236G>A, and a known missense variant in patient 23.I and 23.II, c.530G>A, the remaining variants are classified as “likely pathogenic” or of “uncertain significance”. An increasing number of patients with IT are being reported with variants in *RUNX1* (Latger-Cannard, Philippe et al. 2016). Without functional confirmation to determine the effect of variation it is difficult to confirm causality; however this raises the question of whether it should be considered as clinically significant as BSS and MYH9-related disorders and be primarily screened by genetics upon initial diagnosis with IT.

Patient 22.I, is a unique case of a patient identified with a variant in *RUNX1.* The patient was observed following WES analysis with a missense variant c.505A>G

which leads to a threonine to alanine change at position 169 within the amino acid sequence. The patient was recruited to the study posthumously and as such platelet counts, morphology and platelet function testing were unavailable, however a reduced platelet count was noted in neonatal examination. Upon post-mortem examination excessive bruising and subcutaneous haemorrhaging was observed. The variant observed in *RUNX1* is therefore likely to be the causative variant of the haemostatic disorder observed, although uncertain significance of pathogenicity was predicted, however, it is unlikely to explain the cause of death.

WES analysis revealed two missense variants within patient 29.I. The variants resided within *TUBB1* and *TPM4. TPM4* is a novel gene not previously implicated within IT within the literature. Unpublished data from a murine ENU-based mutagenic screen as performed by Professor Ben Kile, elucidated the role of TPM4 in megakaryopoiesis. Mice containing a prematurely truncated protein Tpm4 display an increase in mean platelet volume and decrease in platelet number because of aberrant proplatelet formation characterised by impaired branching. An increased active dephosphorylated cofilin level is observed, suggesting enhanced actin turnover and altered Myh9 localisation is also observed. Functional confirmation is required to distinguish whether the two variants are additive to the reduction or platelet count or one is a tolerated variation.

A benefit of WES analysis is the lack of limitation to a subset of genes. This allows the possibility to find candidate variants in novel genes not previously known to cause IT. A combination of WES analysis with segregation analysis by Sanger sequencing has previously been adopted by the UK-GAPP study in the discovery of variants within *SLFN14* (Fletcher, Johnson et al. 2015). Three adjacent variants were

identified within three index cases, 26.I, 27.I and 28.I, who shared a similar phenotype of a secondary qualitative secretion defect due a reduction in the number of dense granules; in addition to a reduction in platelet count, and increase in MPV and IPF. Since the initial publication, additional patients sharing a similar phenotype have been identified with variants in *SLFN14* (Marconi, Di Buduo et al. 2016).

WES analysis combined with segregation analysis by Sanger sequencing has narrowed down 14 novel variants within patients 36.I, 36.II and 36.III to two candidate variants of disease. Variants were identified in *PADI2* and *TTF2.* Neither gene has previously been implicated in haematological abnormalities. Mutations in *PADI2* are associated with schizophrenia, breast cancer and rheumatoid arthritis (Watanabe, Nunokawa et al. 2009; McElwee, Mohanan et al. 2012; Chang, Xia et al. 2013), whereas mutations in *TTF2* are causative of Thyroid dysgenesis (Castanet and Polak 2010). Although both variants require further work to confirm the functional effect of the variation, WES has provided us with the first steps in determining the impact of these two genes within IT.

Family 41 was consciously approached in a different way and led to the identification of three novel recessive variants within *ANKRD18A*, *FRMPD1* and *GNE*, in affected cousins of a single consanguineous kindred. The molecular function of ANKRD18A is currently unknown and FRMPD1 functions to regulate the subcellular localisation of activation of G-protein signalling 3 (AGS3) (An, Blumer et al. 2008). GNE, an enzyme in the sialic acid biosynthesis pathway, is highly expressed in cells of the haematopoietic lineage when compared to ANKRD18A and FRMPD1. Currently mutations within *GNE* are known to be the genetic cause of sialuria (OMIM269921) and hereditary inclusion body myopathy (HIBM; OMIM600737) (Seppala, Lehto et al.

1999; Eisenberg, Avidan et al. 2001). Recently however, two separate groups have reported patients with compound heterozygous variations in *GNE*, causing GNE related myopathy with congenital thrombocytopenia (Izumi, Niihori et al. 2014; Zhen, Guo et al. 2014). Platelet counts in the individuals within the reported literature do not exceed $45x10^9$/L, a severe count which is shared in patients 41.I and 41.II. Thrombocytopenia though is observed secondary to degenerative muscle myopathy presenting from adolescence onwards. Patients 41.I and 41.II are currently aged 7 and 3, respectively and show no signs of myopathy. It is feasible though that the variants within *GNE* are causative of disease in addition to myopathy that has yet to clinically present due to the patients age. For this to be confirmed, patients would need to undergo regular observation and potentially platelet associated IgG examination as elevation is characteristic of the disorder. Without this confirmation, it is difficult to draw a conclusion regarding the genetic aetiology of these patients as currently, the patient's phenotype lacks consistency with the genetic defects observed.

Currently the link between *MKL1* and IT is a secondary intermittent symptom to a severe immunodeficiency. WES analysis revealed one patient, 43.I, with a novel missense variant, c.1723G>A, p.Val575Met, occurring within a highly conserved region. This variant represents the only variant to occur in a gene with previous haematological implications. The zygosity observed does not match that of published literature but the crucial role of MKL1 in megakaryocyte maturation through its binding partner SRF makes it an interesting candidate of disease. One further variant in *MKL1* was observed in addition to a "likely pathogenic" frameshift causing insertion within *TUBB1* in patient 31.I. Due to the predicted loss of function of the frameshift

causing *TUBB1* variant it is unlikely that the variant within *MKL1* is additive to the patient's phenotype. However the presence of two variants within two individual patients is interesting for further study.

4.5. Further work

The effectiveness of molecular next generation sequencing relies on subsequent functional confirmation of the effect of genetic variation. Further work would follow this mantra in trying to functionally confirm all genetic variants observed as disease causing. Variants can be classified dependant on a number of *in silico* prediction software tools but without *in vitro* or *in vivo* studies into the mechanism of mutation they are often ambiguous and uncertain.

Due to the high frequency of variants within *RUNX1,* with phenotypic similarities in a reduction in platelet secretion, further work will first initially aim to categorise these variants. Variants within *RUNX1* and *FLI1* additionally, are often associated with an elevated MYH10 expression in platelet lysates as a biomarker of genetic variance. Determining the expression of MYH10 will allow for initial categorisation of the effect of each variant. Due to associated predisposition to myeloid malignancy with variants in *RUNX1,* it is critical to establish diagnosis as early as possible to aid in patient management and guidance.

A number of other genes are often associated with biomarkers of variance. These include: A lack of an intact β-tubulin ring in resting fixed platelets when stained by immunohistochemistry in patients with variants in *TUBB1*. Disorganised actin bundling in megakaryocytes derived from patients harbouring *ACTN1* variants, as well as an increase in apoptotic activity in patients with variants within *CYCS*. All of

these can be tested with further work to be able to confirm or deny the genetic diagnosis.

Determining causality as a result of genetic variance in genes not previously known to cause IT is a more difficult task. As many of the genes mentioned have no previous haemostatic relevance, linking the genetic defect to a reduction in platelet count requires first establishing the role of the encoded protein within normal healthy haemostasis. Often the best method to achieve this is through the use of animal models of disease, but with interspecies differences this is often time consuming and costly. If effective though, it has the potential to expand our current knowledge of haemostasis and aid in further study and clinical categorisation.

4.6. Key findings

- WES analysis revealed 69% of patients with a genetic variant within a gene previously associated with IT.
- Of these variants 43% were deemed to have a positive prediction of pathogenicity and are likely the causative variant of disease within the individual patient.
- A high percentage of variants were observed in *RUNX1* suggesting the importance of the gene within the genetic landscape of IT.
- In addition to variants identified in known genes, variants in novel genes or genes in which variants were associated with a phenotype inconsistent with previously viewed phenotypes were identified in three families.
- Likely causative CNVs of disease were not found within in any of the patients analysed.

5. Thrombocytopenia gene specific panel sequencing

5.1. Summary of background to this research

Targeted panel gene sequencing using next generation sequencing techniques have recently been applied to investigate the genetic basis of IT. As new genes emerge following the continuing use of WES and WGS in small individual kindreds, a static panel of genes becomes a non-sustainable long-term option for clinical diagnosis. WES, without limitations is therefore the best methodology for the future. However, as a technically more challenging method it is not without caveats.

One of the main difficulties is the sheer volume of data that is created. As sequencing technology progresses, the rate-limiting step within the method becomes not the time taken to sequence the sample, but the time taken to carry out the bioinformatic analysis and confirm any plausible genetic variations of disease. Whole exome sequencing is a temporally condensed version of WGS. It also retains a novel capability to identify variants in genes not known previously to cause disease, which although powerful is also a first step to continued work to confirm or deny causality of identified genetic variants.

WES, when applied to a large cohort of patients with IT, is effective in determining the potential genetic aetiology of disease. However, in this thesis, the study of 69 individuals with IT of unknown aetiology at the time of diagnosis, 71% patients were identified with variants in genes previously known to cause IT. If we exclude a subset of patients with variants in *SLFN14,* which was identified as a novel candidate gene, the percentage hit rate is still 62%. Although providing insight into novel genes with a

potential role in haemostasis, the main finding was a higher percentage of patients with variants in known IT genes when compared to previous studies.

As new sequencing library capture methods are developed, the speed of sample preparation time is vastly reduced. This is mostly noted through the recently released capture methods; Illumina Nextera Rapid Custom Capture Enrichment and Agilent SureSelect^{QXT}. Both methods propose an improvement in sample preparation without limitations in sequence depth, coverage and accuracy (Garcia-Garcia, Baux et al. 2016). When applied to small-scale custom gene panels, the preparation time can be reduced to a day. In addition, DNA input is also reduced allowing for amplification from <50ng of DNA.

Due to the high percentage of variants within known IT genes as identified by WES in this study, and the increasing advances in custom panel next generation sequencing, an IT specific next generation sequencing panel was designed and included within the UK-GAPP patient workflow. Incorporating a small custom panel prior to WES has the potential to filter out patients with a genetic aetiology of disease within known IT causing genes. Coupled with the Agilent SureSelect^{QXT} transposon based system of sample preparation, an increase in the efficiency of genetic diagnosis, as well as a reduction in the overall cost, can potentially be achieved.

5.2. Aims of this chapter

This section aims to implement a next generation sequencing panel in the UK-GAPP patient workflow on patients recruited under the same UK-GAPP inclusion criteria. The panel will be designed to incorporate all genes known to be previously associated with IT, which effectively pre-screens patients before WES. The panel will

also take advantage of a rapid sample preparation technique allowing for quick genetic diagnosis following patient phenotyping and hypothetically improving the overall efficiency of the workflow.

5.3. Bioinformatics pipeline to determine candidate variants

In total 34 patients, eight of which used for validation, analysed by the IT specific next generation sequencing panel were analysed post sequencing using an adapted pipeline of that used for the analysis of WES data. Variants were initially filtered on frequency, excluding variants with a MAF ≤0.01 in the 1000G database. Synonymous variants not predicted to change the amino acid sequence in the protein coding transcripts were then excluded. As the panel was designed to include the 5' and 3' UTR, variants were additionally filtered dependant on their genomic location within the coding region ±10bp into the intron-exon boundary. The UTRs were including in bait design to try and identify variants within the 5' UTR of *ANKRD26* so that all variants occurring within the 5'UTR of genes were analysed individually. Candidate variants identified were scrutinised using the same *in silico* pathogenicity prediction software and variant classification system as candidates from WES analysis as mentioned in chapter 4.

5.4. Quality of sequencing data, average number of variants and coverage of sequencing data

All individual DNA samples were processed and passed QC at two points throughout sample preparation. Prior to sample pooling, an average calibrated DNA concentration of 2.957ng/µL and a molarity of 13.3nmol/L was observed across all samples. All sequencing runs passed internal QC that is used within the West

Midlands Regional Genetics Service at the Birmingham Women's Hospital and internal QC from the SureCall analysis software. All candidate variants identified were classified as high quality mapped variants with a quality score of 255.

On average, 326 variants were noted in each individual analysed by the IT-specific next generation sequencing panel. This ranged from 265 to 400 variants across all samples analysed. An average of 73 variants, excluding synonymous variants, were noted with a MAF ≤0.01 in the 1000G database per individual. When each of the variants occurring in non-UTR regions were analysed for presence in the ExAC database, only the variants displayed in Table 5.6.1 were of a rare frequency.

Average coverage across the targeted regions was in excess of 95% in all patients analysed. An average read depth of 380 was noted at the site of each variation. This read depth was not observed below 121 at each point of all candidate variants and reached a filtered read depth of 823.

5.5. Validation of IT specific next generation sequencing panel

Validation of the IT specific next generation sequencing panel was performed by analysing the panel's sensitivity in detecting 8 previously identified variants. Variants were identified previously in WES analysis and confirmed by Sanger sequencing. Variants, at the time of validation, were likely candidate variants and include variants in genes not known previously to cause IT, they are shown in Table 5.5.1. All variants, excluding a previously identified frameshift causing insertion in *TUBB1*; c.1080_1081insG, p.Leu361Alafs*19 previously identified in patient 31.I, were successfully identified. All known candidate variants tested were the only candidate variants following bioinformatics analysis of panel sequencing results in each patient.

Patient	Gene	Variation	Type
2.I	*ANKRD26*	c.-126T>G	5'-UTR
17.I	*RUNX1*	c.G236A, pTrp79*	Nonsense
20.I	*RUNX1*	c.G332A, p.Gly108Ser	Partial heterozygous missense
21.I	*RUNX1*	c.427+1G>T	Splice site variant
26.I	*SLFN14*	c.A652G, p.Lys218Glu	Missense
31.I	*TUBB1*	c.1080_1081insG, p.Leu361Alafs*19	Frameshift causing insertion
36.I	*WAS*	c.G1456A, p.Glu486Lys	X-linked Missense
41.I	*ANKRD18A*	c.2395_2397del, p.Glu799del	Non-frameshift causing deletion

Table 5.5.1 Table depicting the eight patients and the eight known candidate variants utilised for the validation of the IT-specific next generation sequencing panel. A heterozygous status with a low frequency mutant allele was observed in patient 20.I by Sanger sequencing as was deemed a partial heterozygote variation.

5.6. Candidate variants observed and variant prevalence in 26 new patients

In total 26 new patients were analysed by an IT-specific next generation sequencing panel. All patients, with the exception of 64.I, were single affected cases lack additional recruited affected family members. Following post-sequencing bioinformatics analysis as detailed in section 5.3, variants were observed in 77% of individuals (Table 5.6.1). This equates to 20 patients who were observed to harbour a genetic variant within a gene previously known or likely to be associated with IT.

In total, 31 variants were noted in the 20 patients observed with a genetic variant in a gene previously known to cause IT. Seven patients were observed with two variants in two alternate genes and two patients were noted with three variants. No patients were noted with two variants occurring within the same gene and all variants were observed in a heterozygous state. One variant; *GP5*; c.867G>C, p.Met289Ile, was noted in two unrelated patients, 57.I and 58.I.

The majority of variants identified were missense variants affecting a single amino acid. This equated to 87% of the variants observed. In addition; one 5'UTR start gain was noted in patient 48.I (*TUBB1*; c.-88G>C), one frameshift causing deletion was noted in patient 50.I (*GP1BB*; c.del120-142, p.Arg42Cysfs*14), one stop loss variant was noted in patient 54.I (*GATA1*; c.1240T>C, p.*414Arg+41), and one nonsense causing SNV was observed in patient 60.I (*ITGA2B*; c.2176A>T, p.Lys726*).

Family	Patient	Gene(s)	Genomic variation	Protein effect	Variation type	Prevalence	PhyloP	Phast cons	Mutation taster	Poly Phen -2	SIFT	Provean	ACMG criteria	Classification
48	I	ABCG5	c.293C>G	p.Ala98Gly	Missense	0.005 (rs145154937)	5.277	1	D	D	D	D	PP3	**Uncertain significance**
		ABCG8	c.1667T>C	p.Phe556Ser	Missense	0.0009 (rs548098742)	2.914	1	D	D	D	D	PP3	**Uncertain significance**
		TUBB1	c.-88G>C		5'UTR start gain	0.004 (1000G)(rs150 072434)								**Uncertain significance**
49	I	ABCG5	c.1864A>G	p.Met622Val	Missense	0.00541 (rs140374206)	-0.748	0	P	B	T	N	BP4	**Uncertain significance**
		NBEAL2	c.6631G>A	p.Asp2211Asn	Missense	Novel	5.515	0.997	D	D	D	D	PM2, PP3	**Uncertain significance**
50	I	CYCS	c.155C>T	p.Ala52Val	Missense	Novel	5.962	1	D	B	D	D	PM2	**Uncertain significance**
		GP1BB	c.del120-142	p.Arg42Cys fs*14	Frameshif t deletion	Novel			D				PM2, PP3, PM4, PP4	**Likely pathogenic**
51	I	FLI1	c.812G>A	p.Arg271Gln	Missense	Novel	5.983	1	D	D	D	D	PM2, PP3	**Uncertain significance**
		MYH9	c.2872G>A	p.Ala958Thr	Missense	0.0009 (rs151036570)	6.088	1	D	B	T	N		**Uncertain significance**
52	I	FLNA	c.5948C>T	p.Ser1983Leu	Missense	0.0026 (rs187029309)	5.952	1	P	D	D	D		**Uncertain significance**
53	I	FLNA	c.7583A>T	p.Asp2528Val	Missense	Novel	4.858	1	D	D	D	D	PM2, PP3,	**Uncertain significance**
		MYH9	c.7C>G	p.Gln3Glu	Missense	0.0015 (rs56200894)	4.643	0.998	D	B	D	N		**Uncertain significance**
		TUBB1	c.1199G>A	p.Ser400Asn	Missense	Novel	5.88	0.972	D	B	D	N	PM2	**Uncertain significance**
54	I	GATA1	c.1240T>C	p.*414Arg+41	Stop loss	Known	2.408	0.572	P				PM4, PS1, PP5	**Likely pathogenic**
55	I	GP1BA	c.206C>T	p.Pro69Leu	Missense	0.001872 (rs138825640)	-1.407	0	P	B	D	D		**Uncertain significance**
56	I	GP1BB	c.242T>G	p.Leu81Arg	Missense	Novel	-0.162	0.175	P	D	D	D	PM2, PP4	**Uncertain significance**
57	I	GP5	c.867G>C	p.Met289Ile	Missense	0.003101 (rs142440028)	2.516	0.55	P	B	D	D		**Uncertain significance**
58	I	GP5	c.867G>C	p.Met289Ile	Missense	0.003101 (rs142440028)	2.516	0.55	P	B	D	D		**Uncertain significance**

		STIM1	c.182A>G	p.Glu61Gly	Missense	0.00004941 (rs202160755)	2.851	1	D	B	D	D		**Uncertain significance**
59	I	ITGA2B	c.886G>A	p.Gly296Arg	Missense	Novel	3.205	0.999	D	B	D	D	PM2	**Uncertain significance**
		RUNX1	c.386C>A	p.Ala129Glu	Missense	Known (rs267607026)	6.077	1	D	D	D	D	PP3, PS1, PS3	**Pathogenic**
60	I	ITGA2B	c.2176A>T	p.Lys726*	Nonsense	Novel	1.419	0.957	D				PM2, PM4, PVS1, PP4	**Pathogenic**
61	I	ITGA2B	c.2417G>A	p.Ser806Asn	Missense	Novel	0.148	0.286	P	B	T	N	PM2	**Uncertain significance**
		WAS	c.995T>C	p.Val332Ala	Missense	0.0051 (rs2737799)	0.096	0	P	B	T	N	BP4	**Uncertain significance**
62	I	MKL1	c.569C>T	p.Pro190Leu	Missense	0.00016 (rs200309955)	3.693	0.987	D	B	T	D		**Uncertain significance**
63	I	MKL1	c.1492G>C	p.Val498Leu	Missense	0.000008638 (rs199750225)	2.138	0.998	D	B	T	N		**Uncertain significance**
64	I	MYH9	c.2152C>T	p.Arg718Trp	Missense	Known	2.044	1	D	D	D	D	PP3, PS1, PS3, PM (segregation), PP4	**Pathogenic**
65	I	MYH9	c.5074G>A	p.Ala1692Thr	Missense	Novel	4.087	1	D	B	T	N	PM2	**Uncertain significance**
66	I	MYH10	c.2987C>T	p.Ala965Val	Missense	0.0079	4.822	1	D	B	D	D		**Uncertain significance**
		NBEAL2	c.4361C>T	p.Thr1454Met	Missense	0.0001	3.227	0.999	D	B	T	N		**Uncertain significance**
67	I	TUBB1	c.421G>A	p.Gly141Arg	Missense	0.00003295 (rs778975827)	5.803	1	D	D	D	D	PP3	**Uncertain significance**
68	I	Unknown												
69	I	Unknown												
70	I	Unknown												
71	I	Unknown												
72	I	Unknown												
73	I	Unknown												

Table 5.6.1 Variants identified by analysis of the IT-specific next generation sequencing panel. Prevelance is shown in the ExAC consortium if not specified otherwise. PhyloP and Phastcons scores are shown following the cut offs displayed in Table 4.3.2. Variants are noted as D – Disease causing and P – Polymorphism in MutationTaster, D – Damaging and T – Tolerated in SIFT, D – Deleterious and N – Neutral in Provean, D – Damaging and B – Benign in PolyPhen-2 *in silico* pathogenicity prediction software.

Of the 31 variants, 11 (35%) were novel and not previously identified in any of the databases scrutinised. 17 variants have been observed previously and the prevalence of these variants in the ExAC database, unless otherwise stated, is displayed in Table 5.6.1. When comparing all previously observed variants, an average MAF of 0.004 is noted. All variants were observed at a frequency of less than 0.0051 and all previously identified variants, with the exception of two, were present within the latest build of dbSNP.

Three variants were identified that are previously known to cause IT. These were found in patients; 54.I (*GATA1*; c.1240T>C, p.*414Arg+41), 59.I (*RUNX1*; c.386C>A, p.Ala129Glu), and 64.I (*MYH9*; c.2152C>T, p.Arg718Trp).

Two variants previously identified by WES analysis of 69 patients were also identified in patients analysed by the IT specific panel sequencing. These variants; *CYCS*; c.155C>T, p.Ala52Val, and *ITGA2B*; c.2176A>T, p.Lys726* were identified in patients 50.I and 60.I, respectively. Previously, the associated variants had been identified in 4.I (and additional family members) and 13.I, respectively by WES analysis.

5.7. Conservation, pathogenicity prediction and variant classification

Conservation at the site of variation was determined by PhyloP and PhastCons *in silico* software. Conservation scores for all variants occurring within known IT causing genes in the 20 patients are shown in Table 5.6.1. Conservation scores were unavailable for the 5'UTR variant observed in *TUBB1* in patient 48.I and the frameshift causing deletion in *GP1BB* observed in patient 50.I. Average scores of

3.372 and 0.7945 were observed across all variants in PhyloP and PhastCon analysis, respectively. The majority of variants occurred at sites of high conservation and the two methodologies used were in agreement in all instances.

Pathogenicity was predicted using *in silico* prediction software as preformed in chapter 4 and the results of each test are displayed in Table 5.6.1. Classification often varied amongst the software used for each variant indicating the benign potential of the variants observed.

In total, of the 32 variants noted across all patients analysed, three variants were classified as “pathogenic” and two “likely pathogenic” when considering the ACMG consensus guidelines. Three of the five variants deemed to have a positive prediction of pathogenicity were variants previously identified in alternate studies as disease causing. The two remaining variants were those predicted to be detrimental to the length of the expressed protein. The remaining 27 variants without a positive prediction of pathogenicity were classified as of “unknown significance”. Only two variants displayed supporting evidence of a benign classification of the variant.

5.8. Discussion

An IT gene specific next generation sequencing panel was developed in order to pre-screen patients prior to WES. The aim was to efficiently filter out patients with variants in known IT causing genes allowing for a rapid genetic diagnosis and subsequent focus on WES for patients who may harbour variants in novel genes. In addition, the cost implications were an important consideration given that the WES was more than 4 times (£800 vs ~£200) as expensive compared to the targeted panel sequencing.

All sequencing passed QC at all points throughout sample preparation and QC, cluster density and overall sequencing data was sufficient when compared to routine sequencing using alternate capture methods performed within the laboratory of analysis at the West Midlands Regional DNA laboratory at Birmingham Women's Hospital. Although considered a rapid capture method, Agilent SureSelect^{QXT} sample preparation does not quite reach optimum depth of coverage, evenness and target enrichment when compared to alternate methods of capture including Agilent SureSelect^{XT} (Shigemizu, Momozawa et al. 2015; Garcia-Garcia, Baux et al. 2016). When applied to our custom designed panel; average coverage easily exceeded a universally accepted minimum 20x coverage for efficiently calling variants and an average read depth of 380 was identified at the points of variation (Lelieveld, Spielmann et al. 2015).

In comparative tests it has been noted that SureSelect^{QXT} performs well in GC rich areas. This may explain the increase in variants identified in *GP1BB* when compared to WES analysis. With a GC content of 73%, *GP1BB* often suffers from a reduction in coverage, which is why in WES analysis the gene was manually analysed. Utilising the next generation sequencing panel there was no drop in coverage within *GP1BB* for all patients analysed and two variants, in patient 50.I and 56.I, were identified which may be causative of disease which could be potentially missed by other sequencing methodologies.

One of the additional benefits of using next generation sequencing panels and a capture method such as Agilent SureSelect^{QXT} is the relatively low amount of starting material required. A starting quantity of 50ng of DNA from each individual was utilised for sample preparation. Previous studies have shown effective sample

preparation and subsequent sequencing to achieve a depth of coverage sufficient for variant calling is possible from as little as 6.25ng with an adapted technique (Chung, Son et al. 2016). When compared to a required quantity of ~1µg of DNA for WES, the application of an IT specific panel has the potential to aid in cases where little genomic content is available.

In total, candidate variants were identified in 77% of individuals analysed by the IT specific panel. This detection rate far exceeds previous large-scale targeted panel sequencing studies and the application of WES to patients with IT of unknown aetiology discussed in chapter 4. One possible explanation of this is due to the relative increase in average read coverage when compared to WES analysis, especially at the point of variation. When comparing prevalence, however, next generation panel sequencing identifies a large number of variants that have been previously identified with a low MAF. This may be an indication that the variants are tolerated within the population and are not causative of disease. The most comprehensive database of genetic variation is noted to be the ExAC database, which may include data from affected individuals as part of separate studies. It is plausible, therefore, that although the variants have previously been noted, they are causative of a mild reduction in platelet count that has, or has not, been previously diagnosed in all other patients with the shared variant. To determine the reality of this would require further conformational work.

Comparing pathogenicity prediction and variant classification of the variants determined by WES analysis, a larger percentage of variants were deemed to be of unknown significance. This may reflect on a reduced rate of sensitivity and a higher proportion of false negative variants identified. However, it is worth considering that

the majority of variants, 56% (18/32), displayed supporting evidence of pathogenicity but lacked sufficient evidence to be classified as such. This could potentially be an indication of the lack of related affected individuals recruited to the study, for these patients, negatively affecting classification because of a lack of segregation analysis.

Three patients, following analysis, were noted with the presence of a variant previously known to cause IT. Patient 54.I, a 12 year old male with a history of cutaneous bleeding and a mild reduction in platelet count ($101x10^9$/L) was noted with a stop loss variant in *GATA1*; c.1240T>C, p.*414Arg+41. The predicted effect of variation is a loss of the wild type stop codon and extension of the protein by 41 amino acids. Most reported variants within *GATA1* occur within the N-terminal zinc finger domain, leading to a disruption of the binding of GATA1 to FOG1. The stop loss variant noted in patient 54.I was first identified in a 67 year old male proband who suffers from easy bruising (Singleton, Roxby et al. 2013). The patient's platelet counts varied between 86-$94x10^9$/L at different times of testing and no other differences in haematological cell numbers were noted. The patient was initially sequenced due to the presence of a rare X-linked blood group Lu(a-b-) phenotype which results in the marked decrease in expression of Lutheran glycoprotein on the erythrocyte cell surface. To date, serological analysis using flow cytometry to analyse the presence of Lutheran on the erythrocyte cell surface has not been undertaken in patient 54.I. Also the presence of giant occasional macrothrombocytes, a marker of the published phenotype, have not been observed in patient 54.I.

A previously identified causative variant was noted in *RUNX1* in patient 59.I. The missense variant, c.386C>A, p.Ala129Glu, was found in addition to a missense variant in *ITGA2B*. The variant has previously been reported to be causative of

FPD/AML. in three patients from a single pedigree (Preudhomme, Renneville et al. 2009). All three patients were identified with the p.Ala129Glu germline mutation causative of FPD/AML. All patients developed AML as a result of a secondary somatic event occurring within *RUNX1* progressing to patient death in all cases. Patient 59.I is a male with a mild reduction in platelet count to $94x10^9$/L. Following platelet function testing, no reduction in platelet secretion (a hallmark of variants within *RUNX1*) was noted. However, it is highly likely that the variant observed in *RUNX1* is causative of the haemostatic phenotype observed. Whether the variant within *ITGA2B* is additive to the phenotype is unlikely as the platelet count is considered mild in severity but may warrant further investigation.

Patient 64.I, is the only patient analysed by the IT specific panel with affected family members also recruited to the study. The patient forms part of a pedigree of four affected family members with a shared phenotype and clinical symptoms. Following analysis of the IT specific panel sequencing, a missense variant was identified in *MYH9*; c.2152C>T, p.Arg718Trp. This variant has been noted once previously in a patient initially diagnosed with Sebastian syndrome (Pecci, Panza et al. 2008). The variant occurs within the motor domain of MYH9 and is associated with an increased risk of deafness and nephritis, however, no secondary symptoms have previously been reported in patient 64.I or any of the affected family members also recruited to the UK-GAPP study.

The presence of a variant in *MYH9* in patient 64.I is surprising due to the routine pre-screening for disorders such as BSS and MYH9-related disease in haemophilia care centres before recruitment to the UK-GAPP study. This abnormality of identifying patients with variants in genes negatively selected for inclusion is shared when

analysing patients by WES. Analysing patients using the IT-specific panel has elucidated variants in genes known to cause BSS and MYH9-related disease in seven patients. With the exception of patients 50.I and 65.I, who present with the characteristic increase in MPV to the magnitude of observable giant platelets, the remaining patients show an unaltered MPV. No Döhle like body leukocyte inclusions were noted on peripheral blood smears of patients 64.I and 65.I and no patients presented with secondary symptoms relating to specific IT disorders. It is most likely, therefore, that the variants in these genes are either causative of an atypical presentation of BSS and MYH9-related disease or are tolerated and not the genetic aetiology of disease in these patients.

A phenotype-genotype correlation is often utilised in aiding in the diagnosis of a patient's disease. Patient 61.I presented with a marked reduction in the cell surface levels of CD41, the integrin alpha IIb, to around 50% of the levels observed within the travel control tested simultaneously. When analysed by the IT-specific next generation sequencing panel, a missense variant was identified in *ITGA2B*. This variant, c.2417G>A, p.Ser806Asn, is novel within all databases but predicted benign and not well conserved at the site of variation. The variant occurs within the extracellular domain and the integrin alpha IIb heavy chain. Although not predicted to, the reduction in cell surface CD41 is indicative of the possibility that the variant in *ITGA2B* affects either protein levels or cellular localisation potentially leading to the observed platelet-based bleeding phenotype. This is the only occurrence of a genotype-phenotype correlation in all patients analysed by the IT-specific next generation sequencing panel. Although three variants were identified in *ITGA2B* and

one variant was identified in *GP1BA,* none of the patients, with the exception of 61.I, were observed with a reduction in the corresponding cell surface receptor levels.

Interestingly a reduction in CD42b, encoded by *GP1BA*, was noted in patient 50.I, who harbours a potentially deleterious large deletion of *GP1BB* that spans two previously reported disease-causing variants (Moran, Morateck et al. 2000; Kunishima, Naoe et al. 2001). Although not occurring in the encoded gene, the variant, due to the detrimental effect of a frameshift causing deletion, may have propensity to disrupt the stability of the receptor complex leading to a reduction in cell surface expression.

As mentioned previously in chapter 4, one patient, 47.I, was initially analysed by the IT specific next generation sequencing panel and subsequently analysed by WES. Following sequencing using the IT specific next generation sequencing panel, analysis determined that no plausible variants were identified in the known IT causing genes within the panel. It was therefore decided that the patient should undergo WES as the genetic aetiology of disease most likely occurs within a gene not known to previously cause thrombocytopenia. Currently, due to absence of related affected family members and the absence of variants in genes with a haemostatic relevance, no likely candidate variants have been identified that may be associated with the haemostatic phenotype observed within the patient.

One of the main drawbacks of panel sequencing is the static amount of genes that can be analysed. Without a change in the bait design or the inclusion of baits ‘spiked’ into the capture library, the panel can only sequence the genes that it was designed for. This is a caveat that is apparent within the IT-specific next generation sequencing

panel designed in this chapter. The panel was designed to include 27 genes known previously to cause IT, in addition to three likely candidates as a result of WES analysis prior to the formation of the panel as part of the UK-GAPP study and extensive literature searching. Currently the panel therefore absent of the capture baits for eight genes that are reported to cause inherited thrombocytopenia. Although advantageous in providing a rapid insight into the genetic aetiology of IT in patients with unknown cause prior to WES, it is incomplete unless continuously updated. This problem is faced by all current panel-based sequencing platforms including the large multi-disorder ThromboGenomics platform (Simeoni, Stephens et al. 2016). Although more encompassing in its design, it suffers from the same limitations. Spiking baits into a pre-existing capture library is possible, especially if the pre-designed capture library is small in size, however, this is not practical in a clinical setting, suggesting that WES may be a more well-rounded method in determining the genetic aetiology of disease.

5.9. Further work

As with variants determined by WES analysis, the variants observed following the application of the IT-specific next generation sequencing panel requires further confirmational work to be determined disease causing. Further work would focus around this point mainly, utilising many of the biomarkers of disease attributed to variants in certain genes.

A lack of genotype-phenotype correlation shown in patients harbouring variants in *ITGA2B, GP1BA* and *MYH9* in particular is an interesting observation. The possibility that these variants are disease-causing rests on the functional confirmation of the

effect of variation. However, if causative, the patients represent a unique subset of each individual disease that does not share the typical phenotypic presentation of previous cases. The likelihood that patients exist without the secondary symptoms and qualitative defects in platelet function attributed to these disorders is therefore relatively high. Further work would benefit by expanding the patient cohort to recruit patients that may potentially harbour variants in the genes associated with GT, BSS and MYH9-related disease.

Six patients in total were observed without any variants in genes of the IT specific panel. Following the UK-GAPP patient workflow that was applied to patient 47.I, further work would focus on analysing all patients by WES. Due to the absence of variants within the panel of 30 genes, there is a high chance that the genetic aetiology of disease is due to variants in novel genes not previously implicated in IT. Analysis of these patients in particular may progress our current knowledge of IT through the determination of novel causative genes.

5.10. Key findings

- Analysis of 26 patients using an IT specific next generation sequencing panel has identified candidate variants in 77% of individuals.
- Variants include three variations previously associated with IT.
- The IT-specific next generation sequencing panel is therefore an efficient method of pre-screening patients for variants in known IT causing genes prior to WES.

6. Determining the functional effects of sequencing variants

6.1. Summary of background to this research

With the progressive use of next generation sequencing and genotyping in research and clinical diagnosis, an increasing number of sequence variants are being identified. One of the main problems facing this field now is interrogating variants uncovered by sequencing to determine their physiological effects and disease outcomes.

In the two previous chapters, WES and an IT specific next generation sequencing panel have identified a number of variants that are likely pathogenic for the phenotypic presentation of IT. Thus far, classification of these variants has focused on a number of *in silico* prediction software. Indeed, prediction software can aid in the prioritisation of variants but the results should often be interpreted with caution. Although showing a relatively high level of sensitivity (the “true positive” rate) a low specificity (“true negative” rate) is often observed and previously known disease causing variants can be misclassified (Tchernitchko, Goossens et al. 2004; Flanagan, Patch et al. 2010).

Many of the variants identified within this work occur in genes known to previously cause IT and some have been previously identified as pathogenic in alternate studies. Often phenotypic variance is noted between individuals with the same disease and even in cases when the same genetic variant is observed. In addition, growing evidence suggests that many disorders are not simple in their presentation. Indeed, in this study alone a number of patients have been identified with potentially

causative variants in known IT causing genes but a lack of phenotypic similarity when compared to previously reported patients. Because of this heterogeneity confirmation of the functional effect of genetic variation is a necessity for a complete and accurate diagnosis.

Many disorders are accompanied by phenotypic biomarkers characteristic of genetic variation. Previously the presentation of a number of these biomarkers has been harnessed in order to determine the functional effect of variation. This is often an effective and efficient method to determine the relevance of genetic variation. However, in cases of variants in novel genes, extensive accompanying study is required firstly to link the gene to a role within haemostatic control before determining the functional relevance of genetic variation.

6.2. Aims of this chapter

This chapter will work on previous results from a combination of WES and an IT specific next generation sequencing panel to try and elucidate the functional effect of the genetic variations observed. The chapter will focus firstly on the large number of patients observed with a genetic variant in *RUNX1* and *FLI1* utilising the biomarker presence of MYH10 expression. Lastly, the chapter will aim to categorise a novel variant within *TTF2;* c.3265C>G, in a single kindred, focusing initially on the work prior to prioritisation of the variant in *TTF2* as the most likely candidate of disease.

6.3. MYH10 expression in patients with genetic variants in *RUNX1* and *FLI1*

Sixteen patients were noted with variants in *RUNX1* by the two next generation sequencing methods utilised. In addition, five patients also displayed variants in *FLI1.*

In total, stored platelet protein was available for 14 patients. Protein lysate from digested muscle tissue was available for one patient, 22.I. The patients analysed by WB included 11 patients with *RUNX1* variants and 3 with variants within *FLI1*. MYH10 expression was determined using a goat polyclonal IgG antibody mapping near the N-terminus of MYH10 (MYH10 antibody N-17) (SantaCruz, UK, #sc-47204). Primary antibody was used at a 1:500 concentration in 5% skimmed milk powder in TBS-T. Patient platelet protein and protein lysates were analysed in tandem with three cell line lysates; DAMI, HEK293T and HEL. GAPDH was used as a loading control. A range of exposure times were used to determine expression. Shown in Figure 6.3.1 is scanned picture of the WB of MYH10 expression.

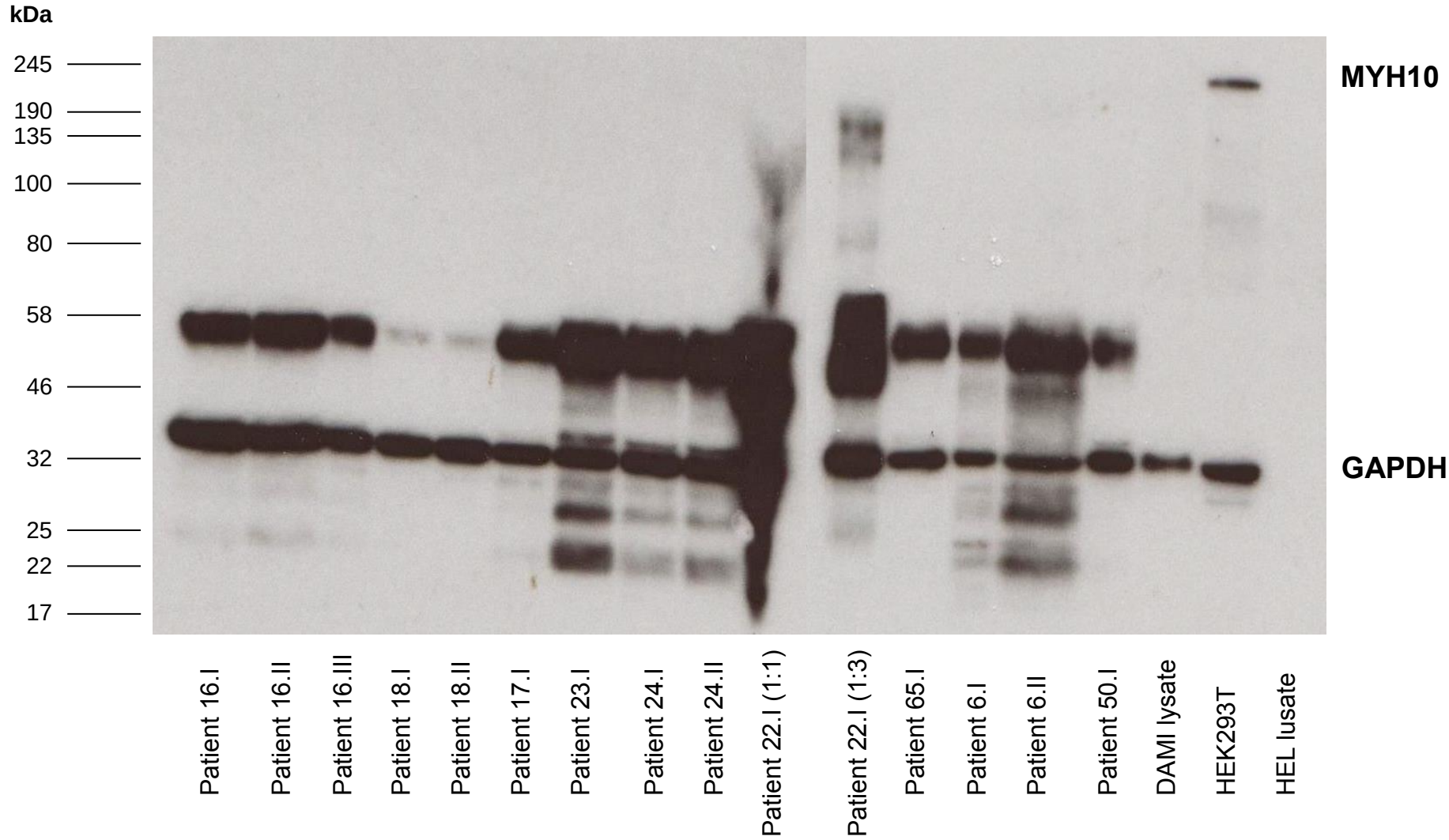


Figure 6.3.1 Western blot to determine MYH10 expression in patients with RUNX1 and FLI1 genetic variants as well as DAMI, HEK293T and HEL cell lysates. MYH10 expression compared to a loading control of GAPDH. No visible expression of MYH10 was noted in any patient and only within the HEL cell lysate.

6.4. Determining a likely candidate variant in a kindred with three affected family members

Three family members, patients 36.I, 36.II and 36.III were recruited to the UK-GAPP study with a shared phenotype of a mild reduction in platelet count. Clinical manifestation including easy cutaneous bruising and bleeding in non-traumatic instances was shared among all three family members. Following initial phenotypic and genotypic investigation four additional unaffected family members with no history of bleeding were also recruited to the study. The family pedigree for all members is shown in Figure 6.4.1. Included are the patients and unaffected individuals most recent platelet counts.

6.4.1. Identification of 14 novel variants shared among all three affected family members following WES

Whole exome sequencing was performed on the three affected family members prior to the recruitment of the additional family members. Analysis depicted 14 novel SNV and small insertion/deletion variants shared among the individuals which are shown in Table 6.4.1. With the exception of a SNV in *WAS*; c.1456G>A, p.Glu486Lys, none of the remaining variants were identified in genes previously associated with haemostasis. When assessing the pathogenicity, the variant in *WAS* was predicted to disrupt the splice acceptor motif of exon 12 of the coding region due to occurring 3 bases into the exon, c.1456G>A. The three affected individuals presented with no secondary symptoms of recurrent infection usually attributed to WAS. A likely diagnosis of XLT was therefore predicted and the variant in *WAS* prioritised for further investigation.

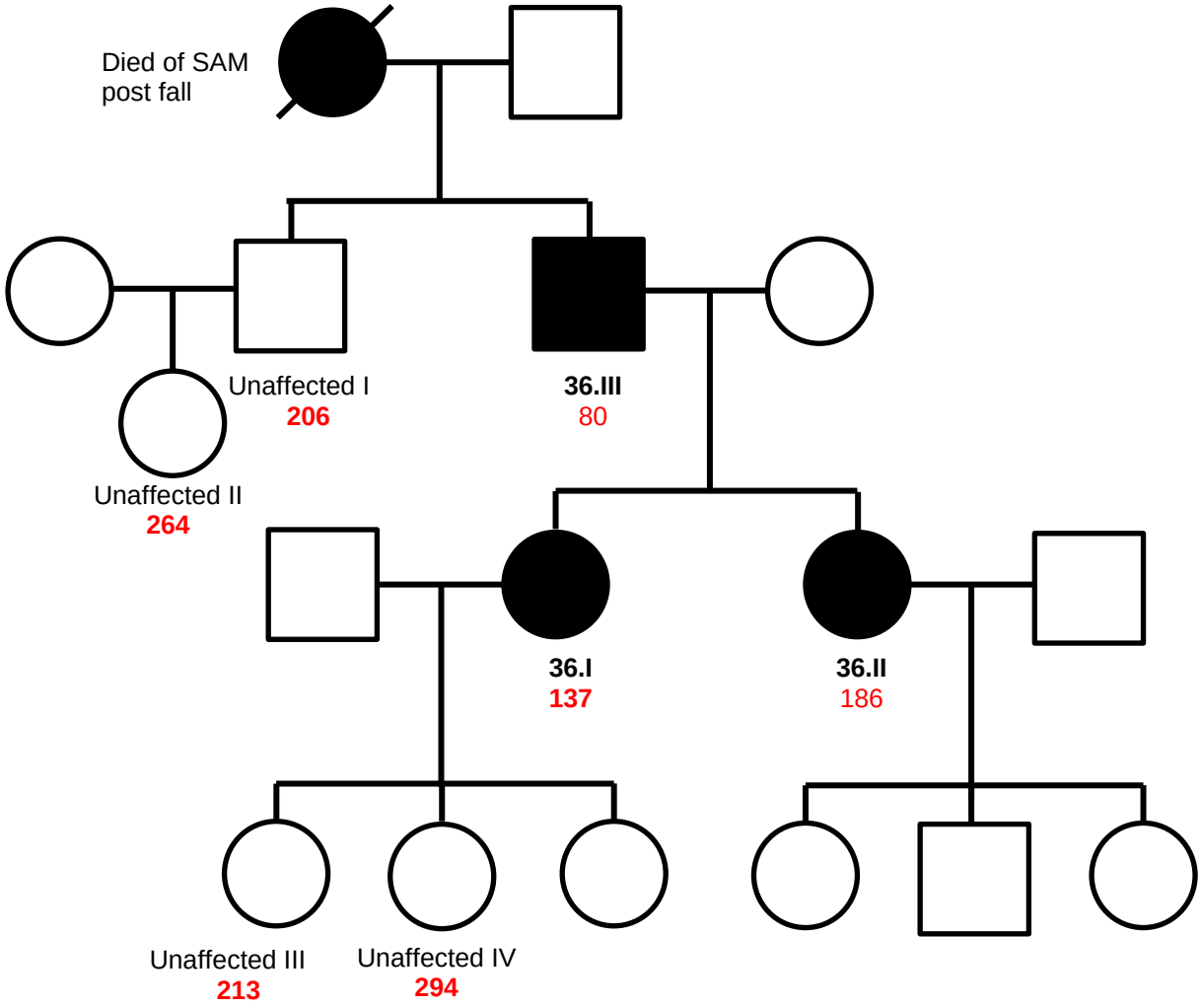


Figure 6.4.1 Family pedigree of the three affected family members of family 36 and the four recruited unaffected family member controls. Inluded are whole blood platelet counts (x10^9/L)

Gene	Genomic variation	Protein effect	Variation type
MARCH1	c.368G>A	p.Arg123Gln	Missense
ALDOB	c.964G>T	p.Glu322*	Nonsense
BAI2	c.4430C>G	p.Thr1477Ser	Missense
CAMKK2	c.1600_1601insAAA	p.534insLys	Non-frameshift insertion
CNGB1	c.2060A>G	p.Asn687Ser	Missense
GSC2	c.263C>T	p.Ala88Val	Missense
MOB3B	c.214A>C	p.Asn72His	Missense
MYBPHL	c.862C>T	p. Pro288Ser	Missense
MYCL1	c.672C>A	p.Asp224Glu	Missense
PADI2	c.1496A>G	p.Lys499Arg	Missense
PEX1	c.3623G>A	p.Arg1208Gln	Missense
RAD21	c.1433G>A	p.Arg478Gln	Missense
TTF2	c.3265C>G	p.His1089Asp	Missense
WAS	c.1456G>A	p.Glu486Lys	Missense

Table 6.4.1 Table showing Sanger sequencing results of all 14 shared novel variants among the three affected family members.

6.4.2. Skewed X-inactivation assay

X-inactivation was assessed to determine the likelihood that presentation of XLT is possible in the two heterozygous female patients due to skewed X-inactivation favouring retention of the mutant allele. When comparing the digested trace (unmethylated active X chromosome) to the undigested trace (methylated and inactive X chromosome) a normal random distribution between the frequency of maternal and paternal CAG lengths is noted in both patients (

Figure **6.4.2).** This indicates that X-inactivation is not skewed favouring retention of one allele over the other.

6.4.3. Assessing WASp expression by Western blot

WB analysis was used to determine whether the expression of WASp encoded by *WAS* was disrupted by the genetic variant in the last coding exon. Protein lysates prepared from washed platelets were available for all three patients. A total quantified amount of 7µg of protein was available for patient 36.III therefore 7µg was loaded for all three patients in tandem with the same quantity of protein lysate from an unaffected unrelated individual. In addition, 18µg of protein was also assessed in patients 36.I and 36.II in tandem with the same quantity of unaffected unrelated protein lysate. A mouse monoclonal primary antibody (B-9) targeted to amino acids 1-250 of the human protein was used to determine WASp expression (SantaCruz, UK, #sc-13139). The primary antibody was used at a concentration of 1:500 and a loading control of anti-GAPDH was also blotted for separately as detailed in the methods section 2.17. Western blots are shown in Figure 6.4.3 and quantification of

the anti-WASp WB using the LI-COR® Odyssey® Fc imaging system is shown in Figure 6.4.3.

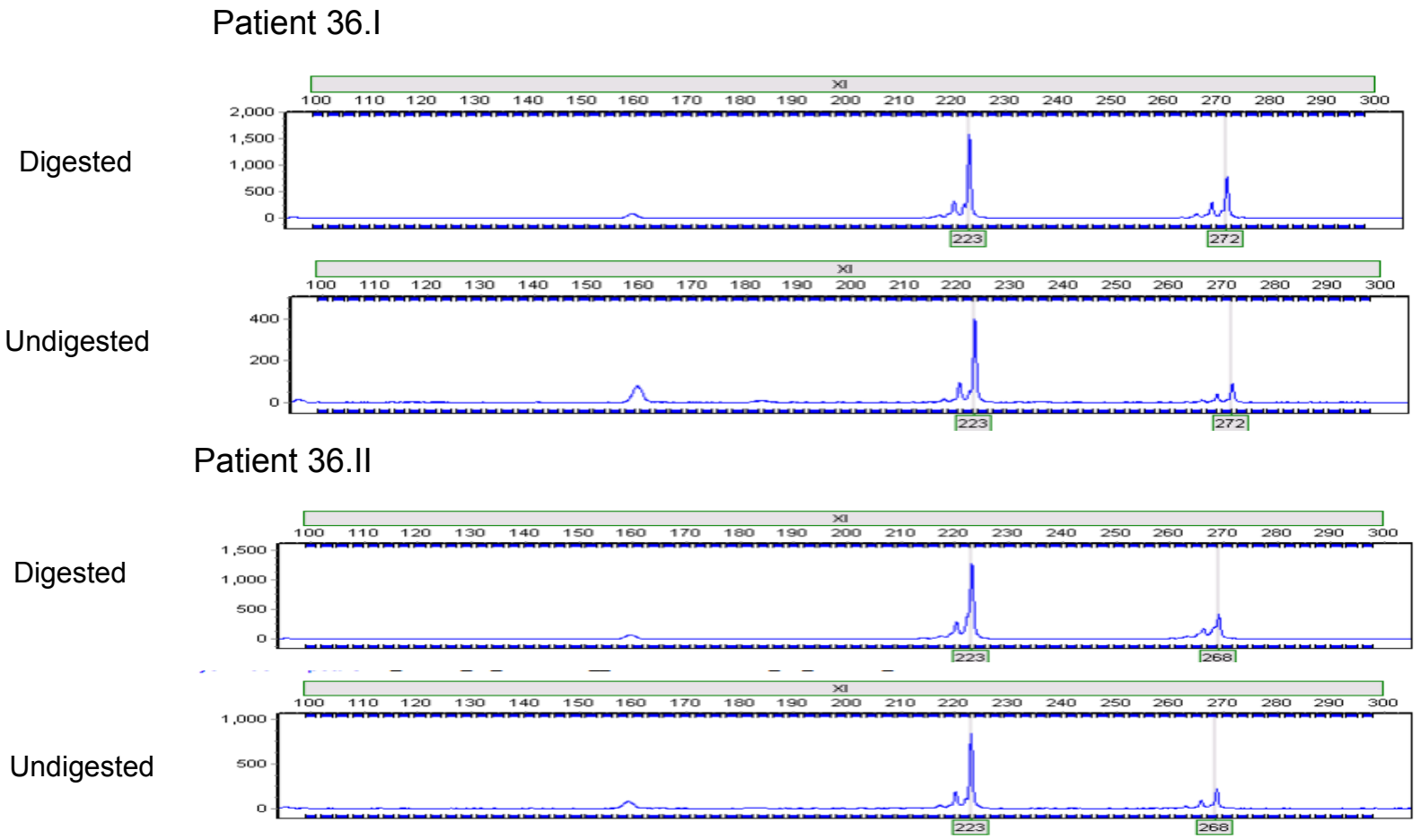


Figure 6.4.2 Skewed X-inactivation assay of patients 36.I and 36.II. Peaks at 223 and 272 CAG repeats in patient 36.I and 223 and 266 in patient 36.II are noted in both the digested and undigested traces indicating normal distribution of X-inactivation.

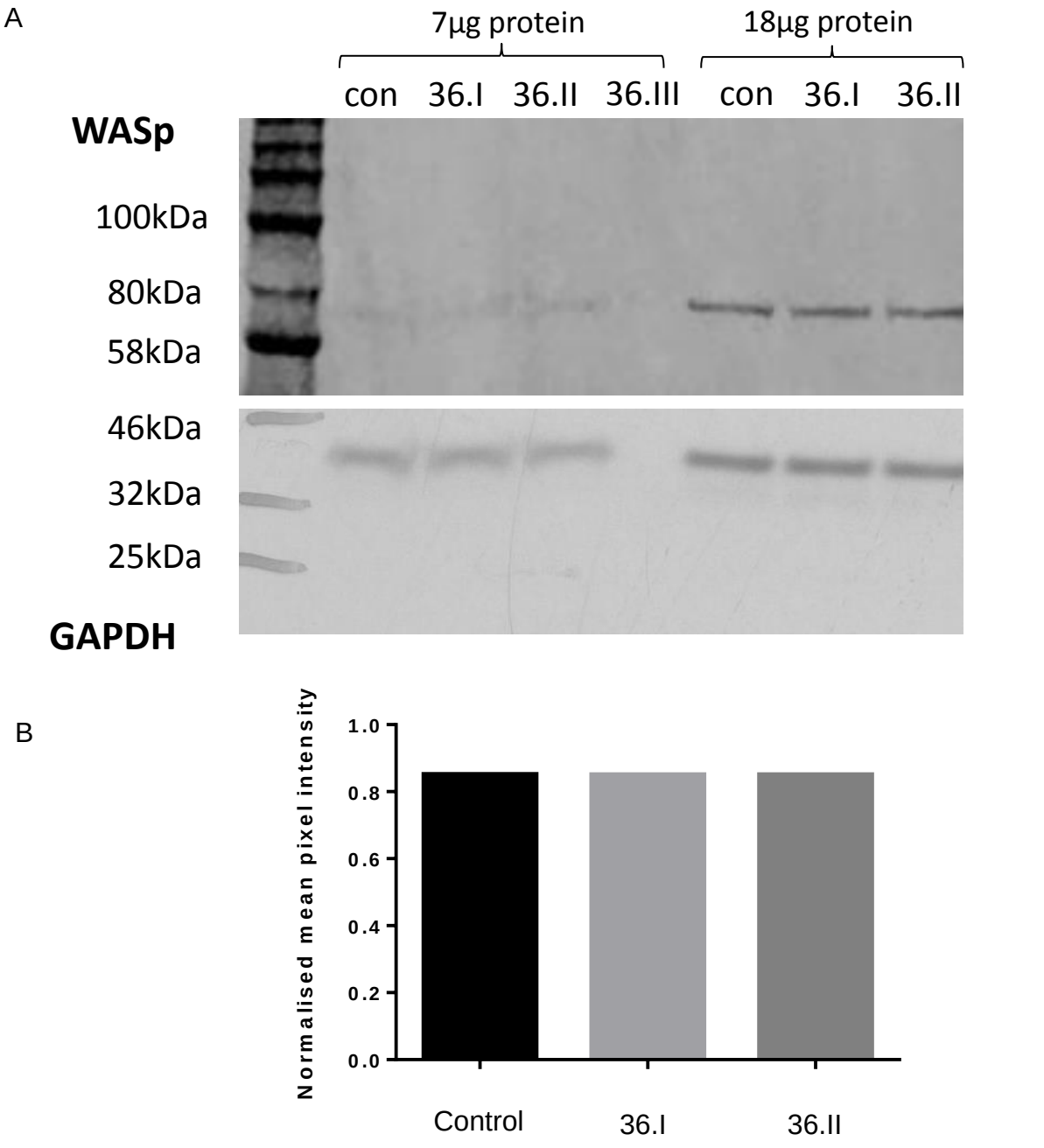


Figure 6.4.3 A. Western blot of WASp expression in patient platelet lysates. WASp expression is compared to a loading control of GAPDH. No difference in the expression of WASp is noted between affected individuals and controls. B. Quantification of expression normalised to the loading control. No difference in quantified expression is noted.

6.4.4. Platelet spreading to determine the effect of *WAS* variation on the actin cytoskeleton

Spreading of washed platelets from the three affected individuals was used to determine any effect on actin cytoskeleton remodelling upon activation. Platelets were spread as detailed in section 2.18. Patient platelets were spread in tandem with platelets from an unrelated unaffected control following the same procedure. Five images were taken per patient/control and 50 cells were counted per image in a predefined region to quantify the average platelet size when spread. Representative images of platelet spreading are shown in Figure 6.4.4, average platelet diameter over all cells counted is shown in Figure 6.4.4. No difference was noted in the number of platelets defined as fully spread, currently spreading or un-spread when comparing the three affected individuals to the unrelated control.

6.4.5. Sanger sequencing of the 14 shared novel variants in four unaffected related family members

Following functional investigation of the prioritised variant in *WAS,* four unaffected related family members were recruited to the UK-GAPP study to investigate segregation of all candidate variants. All four unaffected individuals reported no symptoms relating to haemostasis and platelet counts exceeded $200x10^9$/L. All variants were sequenced by Sanger sequencing as detailed in section 2.15 in all affected and unaffected individuals. Results of sanger sequencing are displayed in Table 6.4.2. Only two variants; *PADI2*; c.1496A>G, p.Lys499Arg and *TTF2*; c.3265C>G, p.His1089Asp, segregated with disease and were only present in the affected individuals of family 36.

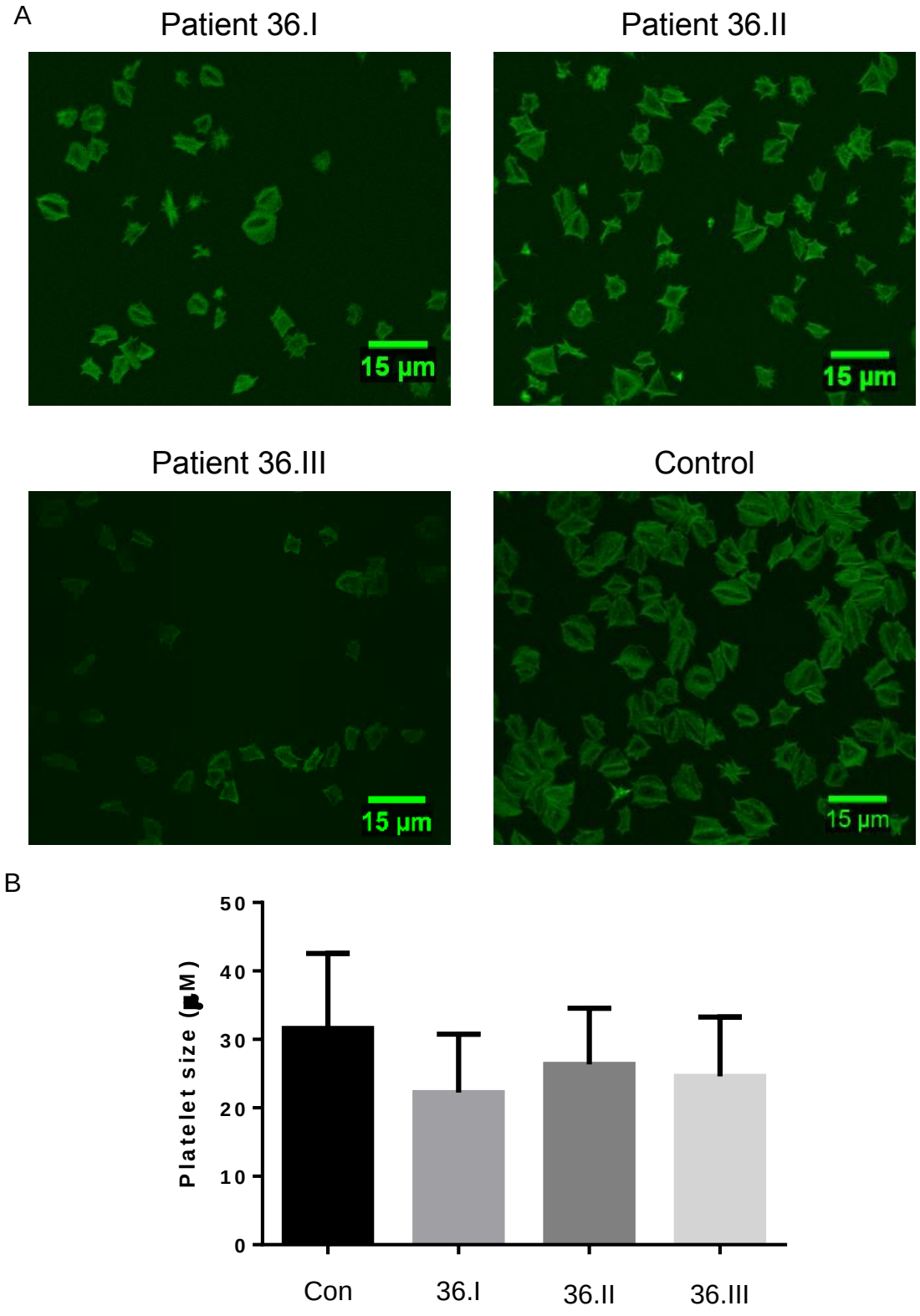


Figure 6.4.4 A. Platelet spreading on fibrinogen coated glass slides. No difference in observable levels of spreading is noted between all patients and the control. B. Average platelet diameter of all cells counted. No difference is noted in average platlet diameter between affected individuals and control.

Gene/Individual	36.I	36.II	36.III	Unaffected I	Unaffected II	Unaffected III	Unaffected IV
MARCH1	c.368G>A	c.368G>A	c.368G>A	c.368G>A	c.368G>A	WT	WT
ALDOB	c.964G>T	c.964G>T	c.964G>T	c.964G>T	c.964G>T	WT	WT
BAI2	c.4430C>G	c.4430C>G	c.4430C>G	c.4430C>G	WT	c.4430C>G	c.4430C>G
CAMKK2	c.1600_1601insAAA	c.1600_1601insAAA	c.1600_1601insAAA	c.1600_1601insAAA	c.1600_1601insAAA	c.1600_1601insAAA	WT
CNGB1	c.2060A>G	c.2060A>G	c.2060A>G	c.2060A>G	c.2060A>G	WT	WT
GSC2	c.263C>T	c.263C>T	c.263C>T	c.263C>T	c.263C>T	WT	WT
MOB3B	c.214A>C	c.214A>C	c.214A>C	WT	c.214A>C	c.214A>C	WT
MYBPHL	c.862C>T	c.862C>T	c.862C>T	WT	WT	c.862C>T	c.862C>T
MYCL1	c.672C>A	c.672C>A	c.672C>A	c.672C>A	WT	c.672C>A	c.672C>A
PADI2	c.1496A>G	c.1496A>G	c.1496A>G	WT	WT	WT	WT
PEX1	c.3623G>A	c.3623G>A	c.3623G>A	c.3623G>A	WT	c.G3623A	c.G3623A
RAD21	c.1433G>A	c.1433G>A	c.1433G>A	c.1433G>A	WT	WT	WT
TTF2	c.3265C>G	c.3265C>G	c.3265C>G	WT	WT	WT	WT
WAS	c.1456G>A	c.1456G>A	c.1456G>A	WT	c.1456G>A	WT	WT

Table 6.4.2 Sanger sequencing of all 14 novel shared variants in the three affected family members and four unaffected family members of family 36. Only variants in *PADI2*; c.1496A>G and *TTF2*; c.3265C>G are present only within the three affected family members and absent from all unaffected family members.

6.4.6. *In silico* expression of PADI2 and TTF2 in cells of the haematological lineage

Expression of PADI2 and TTF2 was determined using Blueprint epigenome data (Chen, Kostadima et al. 2014). RNAseq data showed weak to no expression of PADI2 throughout the progenitors leading to megakaryocytes where as TTF2 was expressed in all progenitor populations of the haematopoietic lineage (Figure 6.4.5). In particular, expression of the full length 1162 amino acid encoding RNA, transcript ID – ENST00000369466, is high among megakaryocyte and erythroblast cell populations.

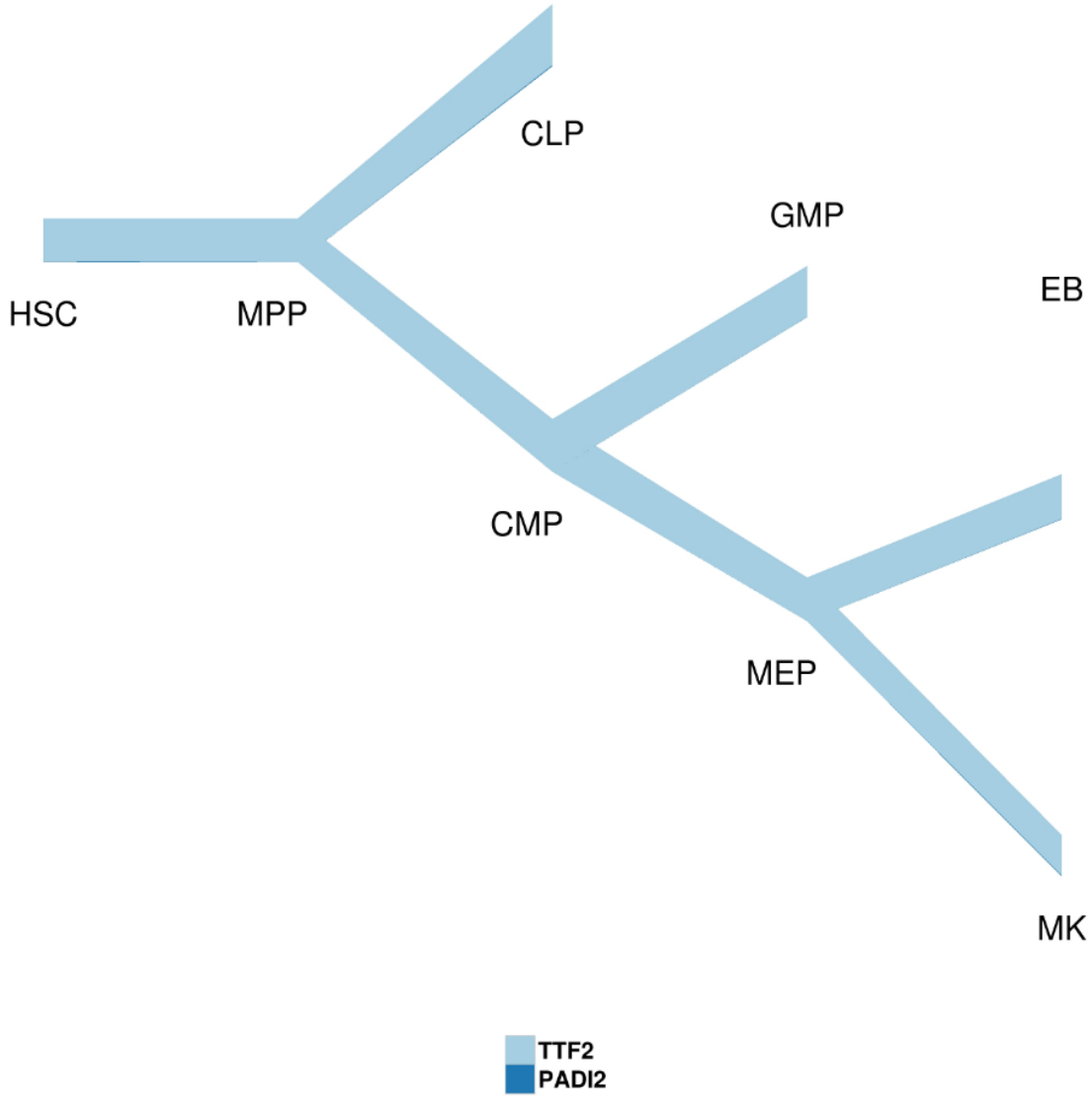


Figure 6.4.5 *In silico* expression analysis of TTF2 and PADI2 in cells of the haematopoietic cell lineage. Expression of TTF2 is noted in all cells of the haematopoietic lineage as opposed to no expression of PADI2. HSC – Haematopoietic stem cell, MPP – Mulitpotent progenitor, CMP – Common myeloid progenitor, MEP – Myelo-erythroid progenitor, EB – Erythroblast, and MK – Megakaryocyte.

6.4.7. Western blot to determine expression of TTF2 in platelets

Platelet protein prepared as detailed in section 2.9 and platelet lysates prepared as detailed in section 2.10 were available for patients 36.I and 36.II. Both preparations of protein in both patients were run in tandem with unaffected unrelated control samples prepared in the same way. 18µg of quantified protein in protein lysate samples was added for both patients and control. Samples were also run in tandem with cell lysates from HEL cells as a positive control. WB were probed with anti-TTF2 rabbit monoclonal antibody targeted to amino acids 50-150 of the human protein sequence [ERP12756] (Abcam, UK, #ab181116) and an anti-GAPDH as a loading control. Anti-TTF2 primary antibody was diluted 1:500 in 5% skimmed milk powder in TBS-T. WBs were visualised using ECL and an example is shown in Figure 6.4.6. WB analysis depicted that no observable expression of TTF2 was identified in either patient or control prepared platelet protein. Due to a limited preparation of both protein sources, no material was available for patient 36.III. A band was noted at 52kD in the anti-GAPDH expression blot. As the antibody is specific to a 36kD GAPDH this is predicted to be non-specific binding of the antibody.

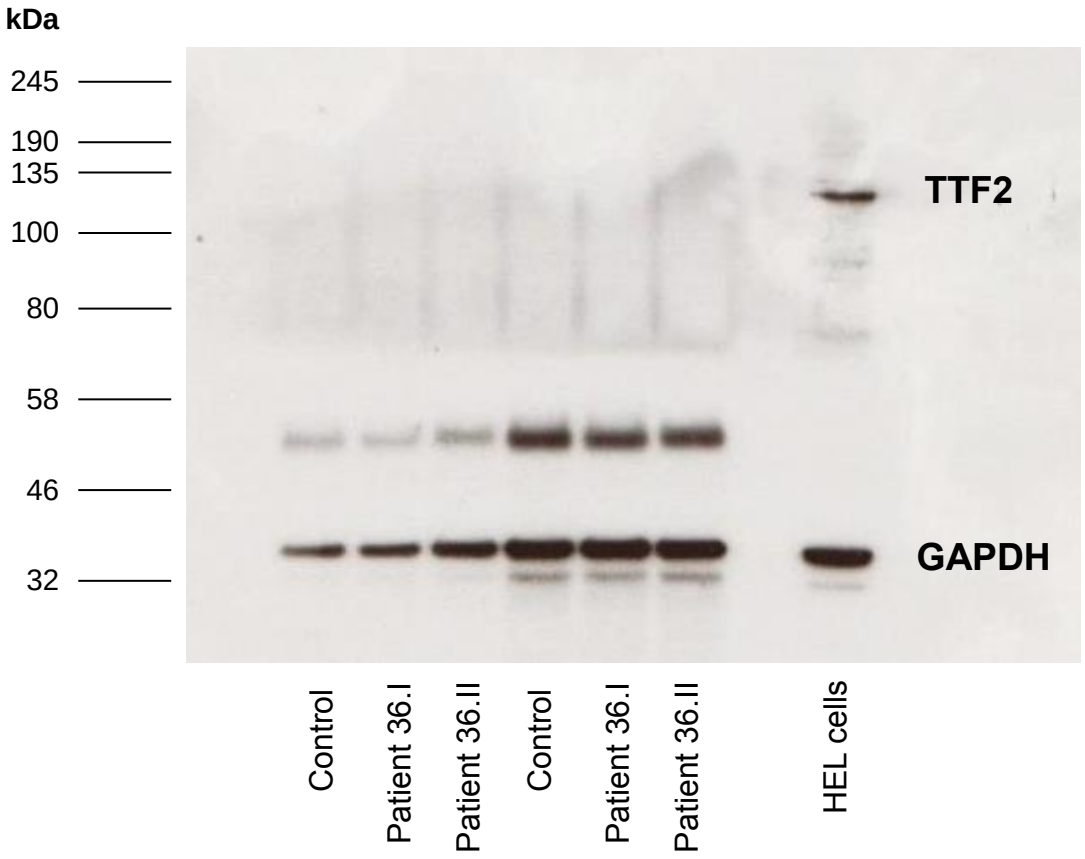


Figure 6.4.6 Western blot to determine the expression of TTF2 in patient cell lysates and HEL cell lysates as a positive control. TTF2 expression compared to a loading control of GAPDH expression. No observable expression of TTF2 was noted in either affected individuals or the control sample and was only present in the positive control.

6.4.8. Utilising the CRISPR-Cas9 system to analyse the *in vitro* effect of variations within TTF2

CRISPR-Cas9 allows for highly efficient and specific gene editing in a variety of cell types. To further investigate the effects of the TTF2 variants reported herein, a CRISPR-Cas9 workflow introducing a fluorescent tag into the C-terminal coding region of the protein was applied. Four guide sequences within 25 base pairs of the STOP codon were designed and cloned into the small guide RNA encoding vector BPK1520. A donor template carrying the mEos 3.2 fluorescent tag flanked by homology arms approximately 700 base pairs in length was designed and cloned into a pGem-T easy backbone. All four sequences exceeded a quality cut-off of 50 and had minimal off-target effects especially those confined to the target sequence. To date, all four guide sequences and mEos 3.2 donor clones have been successfully cloned, amplified and sequence verified. DNA has been prepared for transfection into cell lines to determine transfection efficiency as well as initial impact of genetic variation in *TTF2* once a stable cell line is established.

6.5. Discussion

Functional confirmation of sequencing variants focused on two main areas; determining the presence of MYH10 expression as a biomarker of *RUNX1* and *FLI1* mutations and elucidating the genetic cause of disease in a kindred of three affected individuals. Due to time constraints, many of the functional assays to be able to determine the effect of genetic variation were unfinished and in progress.

Western blot analysis of MYH10 expression in patients with *RUNX1* and *FLI1* variants identified a lack of expression in all patients analysed. This may indicate two

possibilities; that the genetic variations within the two aforementioned genes are benign and do not alter the function of their encoding proteins leading to an absence of aberrant MYH10 expression. Alternatively, a lack of MYH10 expression may be a result of an inaccurate assay. Two cell lysates were utilised as positive controls, variance was observed between the expression in both lysates with clear expression noted in lysates of HEL cells but a lack of expression in HEK293T cells. Both cell lines are noted to express MYH10 at significant levels for detection by WB within the Human Protein Atlas (http://www.proteinatlas.org/ Last accessed October 2016). In addition, nine of the patient derived platelet protein samples were prepared in excess of four years ago, at the time of patient recruitment. Samples have undergone multiple free-thaw cycles suggesting that the protein content of each sample may be degraded in quality leading to an inability to determine expression upon WB analysis.

The second part of this chapter focuses around a kindred with three affected individuals. All three individuals shared phenotypic symptoms relating to easy bruising, however, a slight variation in platelet count and secondary qualitative defects in platelet function was noted. Due to the similarities in clinical presentation this is most likely interpersonal variation.

Following WES, 14 novel shared variants were noted among the three affected individuals. A SNV in *WAS* was prioritised due to previous implications of the gene within IT. Due to the location of *WAS* on the X-chromosome, presentation of XLT and WAS in female patients is often rare and as a result of skewed X-inactivation (Andreu, Matamoros et al. 2007). X-inactivation was assessed in the two female patients, 36.I and 36.II, by methylation sensitive PCR of the CAG repeat within the androgen receptor. No skewing towards the retention of an active mutant allele was

noted in either patient indicating a normal heterozygous state of variation, different to the hemizygous state noted in the male patient 36.III.

Overall the functional categorisation of the variant within *WAS* suggests that the variant is benign and not causative of disease. As no difference in the expression of the encoded WASp was noted, it may be argued that a full length protein is retained but is impaired in its ability to impact the actin cytoskeleton. However, cytoskeleton remodelling upon activation leading to platelet spreading was unaffected suggesting its role within platelets was normal.

The attention was therefore shifted to the alternate 13 shared novel variants in genes of no haemostatic relevance. Sanger sequencing to determine segregation of variants with disease in the three affected family members and four unaffected related family members was undertaken to narrow down candidate variants. Two variants, occurring within *PADI2* and *TTF2,* were only present in the affected family members of family 36. When categorised as part of chapter 4 both variants were classified as of “uncertain significance”. However, expression of *TTF2* was far more significant in cells of the haemopoietic lineage, in particular within megakaryocyte cell populations.

TTF2 is predicted to be weakly expressed in platelets when scrutinised in expression databases, however no expression was noted in any of the patients or unrelated unaffected control samples analysed. This may indicate a lack of sensitivity of the assay or expression confined to progenitor cells of the megakaryocytic lineage. Work utilising the CRISPR-Cas9 gene editing system is ongoing, however sequence

verified guides and clones were identified and following subsequent transfection may provide insight into the role of TTF2 upstream of platelet production.

6.6. Further work

As the majority of the results within this chapter are ongoing, due to time constraints, further work will focus around completion of outstanding experiments. MYH10 expression in platelets is a well-established biomarker of genetic variance in *RUNX1* and *FLI1* (Antony-Debre, Bluteau et al. 2012; Lordier, Bluteau et al. 2012). Further work will initially aim to repeat this assay to determine whether the lack of expression is due to a benign effect of variation or a lack of sensitivity of the assay to determine expression, especially in samples prone to degradation. Variants in *RUNX1* are also often associated with a secondary qualitative defect in dense granule secretion. Indeed, a large proportion of our patients with variants within *RUNX1* phenocopy this trait in platelet function testing. Whole mount electron microscopy has previously been utilised to determine a reduction in dense granule number in patients with variants within *SLFN14* (Fletcher, Johnson et al. 2015). This method could be adopted for patients with variants in *RUNX1,* in particular to determine whether the reduction in dense granule secretion is due to a diminishment in number. However, although further categorising the phenotype of the affected patients this will not aid in tying the genetic variations within *RUNX1* to disease and further *in vitro* work would be required to do this.

Priority in family 36 lies with the variant within *TTF2* as the most likely candidate of disease due to its uniform expression throughout megakaryopoiesis. However, genes with little to no expression within platelets and cells of the megakaryocytic lineage

have been previously implicated in disease through variants within *ABCG5* and *ABCG8.*

Initial further work though would aim to continue to determine the *in vitro* effect of variants within *TTF2* before analysing the impact of variations within *PADI2.* Although a lack of expression was noted in patient-derived platelet protein, TTF2 is expressed throughout the megakaryocytic lineage in *in vitro* cell population RNAseq analysis. Determining the expression in CD34+ cells isolated from donated cord blood and cultured to allow megakaryocyte differentiation and maturation would be beneficial in determining whether genetic variation may impact thrombopoiesis (Robert, Cortin et al. 2012).

The next step utilising the CRISPR-Cas9 system would be to co transfect guide, donor and Cas9 expressing vectors into clonal cell lines prior to clonal expansion and determination of the effect of variation. Further work would then aim to apply this methodology to megakaryocytes derived from induced pluripotent stem cells (iPS cells) (Takayama and Eto 2012). This powerful approach is currently being applied to a number of IT causing genes and may provide novel insight into the functional role of *TTF2* in megakarypoiesis.

6.7. Key findings

- A lack of MYH10 expression in platelet protein in patients with variants in *RUNX1* and *FLI1.*
- Elimination of a variant within *WAS,* through functional characterisation of the effect of variation, as disease-causing in a kindred with three affected individuals.

- Determination of variants within *PADI2* and *TTF2* as potential candidate variants of disease through Sanger sequencing-derived segregation analysis.
- Lack of *TTF2* expression with prepared platelet protein.
- Initial steps to analyse the *in vitro* effect of variants within *TTF2* in cells of the haematopoietic lineage through the CRISPR-Cas9 system have been completed.

7. Overall Discussion and Conclusions

IT represents a heterogeneous group of disorders characterised by a reduction in the circulating platelet count. Indeed, when considering all patients entered into the UK-GAPP study with IT of presumed unknown genetic aetiology to date, a wide variability is observed among the physical characteristics measured. Age and sex distribution among the 95 patients was consistent with other large-scale studies into IT and bleeding diathesis, as previously mentioned. This can also be said for the platelet characteristics including platelet count and size, which vary from severe to borderline normal. One unique attribute of this study is the inclusion of patients with a platelet count exceeding the normal cut-off of $150x10^9$/L. These patients were included due to the likelihood that the effect of genetic variation has reduced their platelet count but not significantly below a standardly recognised threshold. Considering that many of these patients would have been overlooked if the usual cut-off was applied, they represent an interesting sub-set which may offer a unique insight into the genetic aetiology of disease. Interestingly, variants were noted in a wide variety of known IT-causing genes in this subset of patients indicating that the reduction in platelet count may be consistent with other patients but these patients have a higher overall circulating level of platelets.

Phenotypically analysing all patients recruited to the study gives the UK-GAPP study group a novel stance in the field of IT research. By understanding, in depth, the potential functional effect of disease it can help with narrowing candidate variants in subsequent genetic analysis. It also provides crucial information to support any candidate variants in genes previously known to cause disease. This is the case in

patients with a number of variants within *RUNX1,* who share the phenotypic attribute of a reduction in measurable platelet secretion, as well as patient 61.I who showed a reduction in surface CD41, which was supported by subsequent identification of a variant in *ITGA2B*.

Platelet function testing, however, is not without its limitations. As mentioned previously there is a lack of an industry gold standard for determining all ranges of platelet function defects. This may explain the lack of any observable defect in 29% of patients analysed. These patients could potentially harbour defects in platelet cytoskeleton remodelling, platelet adhesion or a number of platelet functions that are not covered by the platelet function tests utilised within the study. Alternatively the bleeding associated with these patients could be attributed directly and solely to the reduction in platelet count.

Utilising WES analysis is a novel approach to determining the causality of IT. In this study WES has also been applied to a different subset of patients than what was previously tested in alternative studies. As such, the genetic picture of disease portrayed in this work is not comparable to previous large-scale studies into the disease. However, detection rates, when considering those predicted to be likely disease causing, are comparable to both previous large-scale studies into IT and other WES based approaches to Mendelian disease proving the effectiveness of WES in this disease subtype.

When compared to previous studies utilising targeted sequencing, a superior overall detection rate for all variants was noted. However, no likely causative variants in genes previously known to cause IT or novel genes were found in 13 index cases.

With the exception of one index case, 37.I, all individuals were single affected cases. All individuals displayed a sustained reduction in platelet count or the presence of affected family members that were unavailable for recruitment to the study indicating the genetic origins of disease. There are therefore two plausible explanations for the lack of variants identified. One explanation is that the genetic variant(s) causative of disease occur in novel genes and the post-sequencing bioinformatics pipeline utilised is not sensitive enough to detect the variants. The other explanation is that variants occur in genomic regions beyond the coding exome. To date, excluding variants that occur within the 5'UTR of *ANKRD26* causative of THC2, only one variant has been previously identified in a regulatory region outside of the coding sequence. This variant is the low frequency SNP in *RBM8A* co-inherited with a null allele in patients with TAR (Albers, Paul et al. 2012). There is a possibility that variants in regulatory regions may explain the cause of disease in patients without a likely candidate variant of disease identified by WES.

Focusing on variants in genes previously known to cause IT initially showed, most interestingly, a large proportion of patients with variants in *RUNX1*. Causative of FPD/AML, variants were previously determined to be relatively rare in frequency with only a handful of variants known in published literature and mutation databases. However, WES analysis has revealed that its prevalence, when excluding patients with an already established causative variant of disease, is more inclined to more prevalent disorders such as BSS and MYH9-related disease. This information carries a strong clinical implication, suggesting that FPD/AML should be considered upon initial diagnosis of IT.

Focusing our initial post-sequencing analysis of variants within genes known to previously be associated with IT lead to design and application of a IT-specific next generation sequencing panel as a pre-screening tool prior to WES. An Agilent SureSelect^{QXT} platform was chosen for library preparation as an efficiently and accurately optimum method. In validation testing, a slight lack of sensitivity was noted, however, QC, coverage and the presence of false-positive results all exceeded quality cut-offs upon application of the sequencing panel to the subset of selected patients. As a pre-screening tool prior to WES, the IT-specific next generation sequencing panel provided a potential candidate variant of disease in 77% of patients analysed. Although requiring functional categorisation to determine the effect of variation it is an effective tool to minimise the time and costs taken to analyse the genetic aetiology of patients with IT.

The ability to determine the genetic aetiology of IT is often bound to be problematic, in part due to the heterogeneity of disease but also due to the likelihood that identified variants are benign. Functional characterisation of variants is therefore required to conclusively determine whether the variants are disease causing or not. Indeed a number of variants within the ExAC database are believed to be benign, even those previously thought to be associated with disease (Song, Gardner et al. 2016). In this study, follow-up functional work is ongoing but is fundamental to determine the mutation status of all variants identified, in particular those in genes not known previously to cause IT.

7.1. Future work arising from this study

There are several hypotheses arising from the work described in this thesis:

Hypothesis 1

Patients with a borderline platelet count between 150-200x10^9/L, that exceed previously recognised thresholds of thrombocytopenia, represent an interesting subset of IT previously unassessed.

Suggested work

Fully analyse all patients with a platelet count between 150-200x10^9/L using the UK-GAPP patient workflow to determine if these patients harbour previously unidentified genetic variants that are causative of a very mild reduction in platelet count.

Hypothesis 2

A high proportion of patients identified with variants in *RUNX1* indicate a clinical relevance of the gene in IT.

Suggested work

Extend the recruitment of patients to include those that are potentially pre-screened routinely for BSS and MYH9-related disease in haemophilia referral centres to determine the true prevalence of *RUNX1* variants in IT. This can be followed subsequently by the development of a genetic pre-screening protocol for clinical use that can be combined with genetic counselling following a positive result.

Hypothesis 3

Patients with no plausible candidate variants identified by WES analysis are likely to harbour variants in novel genes or outside of the coding exome.

Suggested work

Recruitment of additional affected and unaffected family members, where applicable, to help define candidate variants by determining genetic segregation of variants with disease. Application of WGS to patients with no likely candidate variants causative of IT to consider potential variants occurring in regulatory regions.

Hypothesis 4

Variants in *TTF2* are causative of IT in a kindred with three affected individuals.

Suggested work

Progress with functional characterisation of the effect of variation through the utilisation of the CRISPR-Cas9 genome editing technology. Null alleles will first be analysed in relevant cell lines to determine the effect of a complete lack of expression. Cell lines and iPSC derived megakaryocytes will be further used to determine the effect of the variant identified within family 36 on haemostasis and the production, function and lifespan of platelets.

Publications associated with this study

1. **Johnson B**, Fletcher SJ, Morgan NV. (2016) “Inherited thrombocytopenia: novel insights into megakaryocyte maturation, proplatelet formation and platelet lifespan” Platelets. 27(6)
2. Lozano ML, Cook A,… **Johnson B,** *et al.* (2016) “Novel mutations in RASGRP2, which encodes CalDAG-GEFI, abrogate Rap1 activation, causing platelet dysfunction”. Blood. 128(9
3. **Johnson, B**., Lowe, G.C., Futterer, J., *et al*., (2016) “Whole exome sequencing identifies genetic variants in inherited thrombocytopenia with secondary qualitative function defects.” Haematologica.
4. Fletcher, S.J., **Johnson, B.,** Lowe, G.C., *et al*., (2015) “SLFN14 mutations underlie thrombocytopenia with excessive bleeding and platelet defects.” J Clin Invest. 125(9)

International conference presentations associated with this study

1. Whole exome sequencing in patients with inherited thrombocytopenia and excessive bleeding is an efficient way to identify genetic variants in known and novel genes." **Johnson B.** Oral presentation – Highlights of ISTH

   25th International Society on Thrombosis and Haemostasis congress, 2015, Toronto, Canada.
2. "Whole exome sequencing as a tool for detection of the genetic basis of inherited thombocytopenias." **Johnson, B.,** Lowe, G.C., Drake, S., *et al.*

   47th European Human Genetics Conference, 2014, Milan, Italy.

References

Adzhubei, I. A., S. Schmidt, et al. (2010). "A method and server for predicting damaging missense mutations." Nat Methods **7**(4): 248-249.

Akashi, K., D. Traver, et al. (2000). "A clonogenic common myeloid progenitor that gives rise to all myeloid lineages." Nature **404**(6774): 193-197.

Al Daama, S. A., Y. H. Housawi, et al. (2013). "A missense mutation in ANKRD26 segregates with thrombocytopenia." Blood **122**(3): 461-462.

Albers, C. A., A. Cvejic, et al. (2011). "Exome sequencing identifies NBEAL2 as the causative gene for gray platelet syndrome." Nat Genet **43**(8): 735-737.

Albers, C. A., D. S. Paul, et al. (2012). "Compound inheritance of a low-frequency regulatory SNP and a rare null mutation in exon-junction complex subunit RBM8A causes TAR syndrome." Nat Genet **44**(4): 435-439, S431-432.

Aldrich, R. A., A. G. Steinberg, et al. (1954). "Pedigree demonstrating a sex-linked recessive condition characterized by draining ears, eczematoid dermatitis and bloody diarrhea." Pediatrics **13**(2): 133-139.

An, N., J. B. Blumer, et al. (2008). "The PDZ and band 4.1 containing protein Frmpd1 regulates the subcellular location of activator of G-protein signaling 3 and its interaction with G-proteins." J Biol Chem **283**(36): 24718-24728.

Andreu, N., N. Matamoros, et al. (2007). "Two novel mutations identified in the Wiskott-Aldrich syndrome protein gene cause Wiskott-Aldrich syndrome and thrombocytopenia." Int J Mol Med **19**(5): 777-782.

Andrews, R. K. and M. C. Berndt (2013). "Bernard-Soulier syndrome: an update." Semin Thromb Hemost **39**(6): 656-662.

Antony-Debre, I., D. Bluteau, et al. (2012). "MYH10 protein expression in platelets as a biomarker of RUNX1 and FLI1 alterations." Blood **120**(13): 2719-2722.

Arai, F., A. Hirao, et al. (2004). "Tie2/angiopoietin-1 signaling regulates hematopoietic stem cell quiescence in the bone marrow niche." Cell **118**(2): 149-161.

Arinobu, Y., S. Mizuno, et al. (2007). "Reciprocal activation of GATA-1 and PU.1 marks initial specification of hematopoietic stem cells into myeloerythroid and myelolymphoid lineages." Cell Stem Cell **1**(4): 416-427.

Avanzi, M. P., M. Izak, et al. (2015). "Actin inhibition increases megakaryocyte proplatelet formation through an apoptosis-dependent mechanism." PLoS One **10**(4): e0125057.

Avecilla, S. T., K. Hattori, et al. (2004). "Chemokine-mediated interaction of hematopoietic progenitors with the bone marrow vascular niche is required for thrombopoiesis." Nat Med **10**(1): 64-71.

Bader-Meunier, B., V. Proulle, et al. (2003). "Misdiagnosis of chronic thrombocytopenia in childhood." J Pediatr Hematol Oncol **25**(7): 548-552.

Balduini, C. (2014). "Diagnosis and management of inherited thrombocytopenias." European Human Genetics Conference 2014, Milan, Italy.

Balduini, C. L., A. Pecci, et al. (2012). "Inherited thrombocytopenias: the evolving spectrum." Hamostaseologie **32**(4): 259-270.

Ballmaier, M., H. Schulze, et al. (1998). "Defective c-Mpl signaling in the syndrome of thrombocytopenia with absent radii." Stem Cells **16 Suppl 2**: 177-184.

Bambace, N. M. and C. E. Holmes (2011). "The platelet contribution to cancer progression." J Thromb Haemost **9**(2): 237-249.

Basciano, P. A., J. Matakas, et al. (2015). "beta-1 tubulin R307H SNP alters microtubule dynamics and affects severity of a hereditary thrombocytopenia." J Thromb Haemost **13**(4): 651-659.

Behnke, O. (1968). "An electron microscope study of the megacaryocyte of the rat bone marrow. I. The development of the demarcation membrane system and the platelet surface coat." J Ultrastruct Res **24**(5): 412-433.

Bernard, J. a. J. P. S. (1948). "Su rune nouvelle variete de dystrophie thrombocytaire-hemorragipare congenital." Sem Hop Paris **24**: 3217–3223.

Berndt, M. C., Y. Shen, et al. (2001). "The vascular biology of the glycoprotein Ib-IX-V complex." Thromb Haemost **86**(1): 178-188.

Bertino, A. M., X. Q. Qi, et al. (2003). "Apoptotic markers are increased in platelets stored at 37 degrees C." Transfusion **43**(7): 857-866.

Biino, G., C. L. Balduini, et al. (2011). "Analysis of 12,517 inhabitants of a Sardinian geographic isolate reveals that predispositions to thrombocytopenia and thrombocytosis are inherited traits." Haematologica **96**(1): 96-101.

Biino, G., P. Gasparini, et al. (2012). "Influence of age, sex and ethnicity on platelet count in five Italian geographic isolates: mild thrombocytopenia may be physiological." Br J Haematol **157**(3): 384-387.

Bluteau, D., A. Balduini, et al. (2014). "Thrombocytopenia-associated mutations in the ANKRD26 regulatory region induce MAPK hyperactivation." J Clin Invest **124**(2): 580-591.

Bluteau, D., L. Lordier, et al. (2009). "Regulation of megakaryocyte maturation and platelet formation." J Thromb Haemost **7 Suppl 1**: 227-234.

Bolton-Maggs, P. H., E. A. Chalmers, et al. (2006). "A review of inherited platelet disorders with guidelines for their management on behalf of the UKHCDO." Br J Haematol **135**(5): 603-633.

Born, G. V. (1962). "Aggregation of blood platelets by adenosine diphosphate and its reversal." Nature **194**: 927-929.

Bottega, R., C. Marconi, et al. (2015). "ACTN1-related thrombocytopenia: identification of novel families for phenotypic characterization." Blood **125**(5): 869-872.

Boyles, J., J. E. Fox, et al. (1985). "Organization of the cytoskeleton in resting, discoid platelets: preservation of actin filaments by a modified fixation that prevents osmium damage." J Cell Biol **101**(4): 1463-1472.

Boztug, K., M. Schmidt, et al. (2010). "Stem-cell gene therapy for the Wiskott-Aldrich syndrome." N Engl J Med **363**(20): 1918-1927.

Breton-Gorius, J., R. Favier, et al. (1995). "A new congenital dysmegakaryopoietic thrombocytopenia (Paris-Trousseau) associated with giant platelet alpha-granules and chromosome 11 deletion at 11q23." Blood **85**(7): 1805-1814.

Brewer, D. B. (2006). "Max Schultze (1865), G. Bizzozero (1882) and the discovery of the platelet." Br J Haematol **133**(3): 251-258.

Burrows, R. F. and J. G. Kelton (1990). "Thrombocytopenia at delivery: a prospective survey of 6715 deliveries." Am J Obstet Gynecol **162**(3): 731-734.

Casari, C., V. Du, et al. (2013). "Accelerated uptake of VWF/platelet complexes in macrophages contributes to VWD type 2B-associated thrombocytopenia." Blood **122**(16): 2893-2902.

Castanet, M. and M. Polak (2010). "Spectrum of Human Foxe1/TTF2 Mutations." Horm Res Paediatr **73**(6): 423-429.

Castermans, D., K. Volders, et al. (2010). "SCAMP5, NBEA and AMISYN: three candidate genes for autism involved in secretion of large dense-core vesicles." Hum Mol Genet **19**(7): 1368-1378.

Cattaneo, M. (2002). "Desmopressin in the treatment of patients with defects of platelet function." Haematologica **87**(11): 1122-1124.

Chan, M. V., P. C. Armstrong, et al. (2011). "Optical multichannel (optimul) platelet aggregometry in 96-well plates as an additional method of platelet reactivity testing." Platelets **22**(7): 485-494.

Chang, X., Y. Xia, et al. (2013). "PADI2 is significantly associated with rheumatoid arthritis." PLoS One **8**(12): e81259.

Chen, L., M. Kostadima, et al. (2014). "Transcriptional diversity during lineage commitment of human blood progenitors." Science **345**(6204): 1251033.

Chen, Y., S. Boukour, et al. (2013). "The abnormal proplatelet formation in MYH9-related macrothrombocytopenia results from an increased actomyosin contractility and is rescued by myosin IIA inhibition." J Thromb Haemost **11**(12): 2163-2175.

Chen, Z., O. Naveiras, et al. (2007). "The May-Hegglin anomaly gene MYH9 is a negative regulator of platelet biogenesis modulated by the Rho-ROCK pathway." Blood **110**(1): 171-179.

Chesarone, M. A., A. G. DuPage, et al. (2010). "Unleashing formins to remodel the actin and microtubule cytoskeletons." Nat Rev Mol Cell Biol **11**(1): 62-74.

Choi, Y., G. E. Sims, et al. (2012). "Predicting the functional effect of amino acid substitutions and indels." PLoS One **7**(10): e46688.

Chong, J. X., K. J. Buckingham, et al. (2015). "The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities." Am J Hum Genet **97**(2): 199-215.

Chung, J., D. S. Son, et al. (2016). "The minimal amount of starting DNA for Agilent's hybrid capture-based targeted massively parallel sequencing." Sci Rep **6**: 26732.

Cines, D. B., J. B. Bussel, et al. (2004). "Congenital and acquired thrombocytopenia." Hematology Am Soc Hematol Educ Program: 390-406.

Ciovacco, W. A., W. H. Raskind, et al. (2008). "Human phenotypes associated with GATA-1 mutations." Gene **427**(1-2): 1-6.

Clemetson, K. J. and J. M. Clemetson (1995). "Platelet GPIb-V-IX complex. Structure, function, physiology, and pathology." Semin Thromb Hemost **21**(2): 130-136.

Consortium, I. H. G. S. (2004). "Finishing the euchromatic sequence of the human genome." Nature **431**(7011): 931-945.

Dale, G. L., P. Friese, et al. (2002). "Stimulated platelets use serotonin to enhance their retention of procoagulant proteins on the cell surface." Nature **415**(6868): 175-179.

Dasouki, M. J., S. K. Rafi, et al. (2013). "Exome sequencing reveals a thrombopoietin ligand mutation in a Micronesian family with autosomal recessive aplastic anemia." Blood **122**(20): 3440-3449.

Dawood, B. B., G. C. Lowe, et al. (2012). "Evaluation of participants with suspected heritable platelet function disorders including recommendation and validation of a streamlined agonist panel." Blood **120**(25): 5041-5049.

Dawood, B. B., J. Wilde, et al. (2007). "Reference curves for aggregation and ATP secretion to aid diagnose of platelet-based bleeding disorders: effect of inhibition of ADP and thromboxane A(2) pathways." Platelets **18**(5): 329-345.

De Rocco, D., C. Cerqua, et al. (2014). "Mutations of cytochrome c identified in patients with thrombocytopenia THC4 affect both apoptosis and cellular bioenergetics." Biochim Biophys Acta **1842**(2): 269-274.

Debili, N., L. Coulombel, et al. (1996). "Characterization of a bipotent erythro-megakaryocytic progenitor in human bone marrow." Blood **88**(4): 1284-1296.
Desmet, F. O., D. Hamroun, et al. (2009). "Human Splicing Finder: an online bioinformatics tool to predict splicing signals." Nucleic Acids Res **37**(9): e67.
Dong, F., S. Li, et al. (2005). "Genotype-phenotype correlation in MYH9-related thrombocytopenia." Br J Haematol **130**(4): 620-627.
Dormann, D., J. Kardoeus, et al. (1998). "Flow cytometric analysis of agonist-induced annexin V, factor Va and factor Xa binding to human platelets." Platelets **9**(3-4): 171-177.
Dowdy, C. R., R. Xie, et al. (2010). "Definitive hematopoiesis requires Runx1 C-terminal-mediated subnuclear targeting and transactivation." Hum Mol Genet **19**(6): 1048-1057.
Drachman, J. G. (2004). "Inherited thrombocytopenia: when a low platelet count does not mean ITP." Blood **103**(2): 390-398.
Duke, W. W. (1983). "The relation of blood platelets to hemorrhagic disease. By W.W. Duke." JAMA **250**(9): 1201-1209.
Eisenberg, I., N. Avidan, et al. (2001). "The UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase gene is mutated in recessive hereditary inclusion body myopathy." Nat Genet **29**(1): 83-87.
Fabre, J. E., M. Nguyen, et al. (1999). "Decreased platelet aggregation, increased bleeding time and resistance to thromboembolism in P2Y1-deficient mice." Nat Med **5**(10): 1199-1202.
Falet, H. (2013). "New insights into the versatile roles of platelet FlnA." Platelets **24**(1): 1-5.
Fiore, M., N. Firah, et al. (2012). "Natural history of platelet antibody formation against alphaIIbbeta3 in a French cohort of Glanzmann thrombasthenia patients." Haemophilia **18**(3): e201-209.
Flanagan, S. E., A. M. Patch, et al. (2010). "Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations." Genet Test Mol Biomarkers **14**(4): 533-537.
Fletcher, S. J., B. Johnson, et al. (2015). "SLFN14 mutations underlie thrombocytopenia with excessive bleeding and platelet secretion defects." J Clin Invest **125**(9): 3600-3605.
Force., B. C. f. S. i. H. G. H. T. (2003). "Guidelines for the investigation and management of idiopathic thrombocytopenic purpura in adults, children and in pregnancy." Br J Haematol **120**(4): 574-596.
Foudi, A., D. J. Kramer, et al. (2014). "Distinct, strict requirements for Gfi-1b in adult bone marrow red cell and platelet generation." J Exp Med **211**(5): 909-927.
Fox, N., G. Priestley, et al. (2002). "Thrombopoietin expands hematopoietic stem cells after transplantation." J Clin Invest **110**(3): 389-394.
Frangoul, H., J. Keates-Baleeiro, et al. (2010). "Unrelated bone marrow transplant for congenital amegakaryocytic thrombocytopenia: report of two cases and review of the literature." Pediatr Transplant **14**(4): E42-45.
Frenette, P. S., C. V. Denis, et al. (2000). "P-Selectin glycoprotein ligand 1 (PSGL-1) is expressed on platelets and can mediate platelet-endothelial interactions in vivo." J Exp Med **191**(8): 1413-1422.
Freson, K., R. De Vos, et al. (2005). "The TUBB1 Q43P functional polymorphism reduces the risk of cardiovascular disease in men by modulating platelet function and structure." Blood **106**(7): 2356-2362.
Furie, B., B. C. Furie, et al. (2001). "A journey with platelet P-selectin: the molecular basis of granule secretion, signalling and cell adhesion." Thromb Haemost **86**(1): 214-221.

Gachet, C. (2001). "ADP receptors of platelets and their inhibition." Thromb Haemost **86**(1): 222-232.
Garcia-Garcia, G., D. Baux, et al. (2016). "Assessment of the latest NGS enrichment capture methods in clinical context." Sci Rep **6**: 20948.
Geddis, A. E. and K. Kaushansky (2006). "Endomitotic megakaryocytes form a midzone in anaphase but have a deficiency in cleavage furrow formation." Cell Cycle **5**(5): 538-545.
George, J. N., J. P. Caen, et al. (1990). "Glanzmann's thrombasthenia: the spectrum of clinical disease." Blood **75**(7): 1383-1395.
Germeshausen, M., M. Ballmaier, et al. (2006). "MPL mutations in 23 patients suffering from congenital amegakaryocytic thrombocytopenia: the type of mutation predicts the course of the disease." Hum Mutat **27**(3): 296.
Ghevaert, C., A. Salsmann, et al. (2008). "A nonsynonymous SNP in the ITGB3 gene disrupts the conserved membrane-proximal cytoplasmic salt bridge in the alphaIIbbeta3 integrin and cosegregates dominantly with abnormal proplatelet formation and macrothrombocytopenia." Blood **111**(7): 3407-3414.
Gieger, C., A. Radhakrishnan, et al. (2011). "New gene functions in megakaryopoiesis and platelet formation." Nature **480**(7376): 201-208.
Golden, A., S. P. Nemeth, et al. (1986). "Blood platelets express high levels of the pp60c-src-specific tyrosine kinase activity." Proc Natl Acad Sci U S A **83**(4): 852-856.
Goode, B. L. and M. J. Eck (2007). "Mechanism and function of formins in the control of actin assembly." Annu Rev Biochem **76**: 593-627.
Gordge, M. P. (2005). "Megakaryocyte apoptosis: sorting out the signals." Br J Pharmacol **145**(3): 271-273.
Gueguen, P., K. Rouault, et al. (2013). "A missense mutation in the alpha-actinin 1 gene (ACTN1) is the cause of autosomal dominant macrothrombocytopenia in a large French family." PLoS One **8**(9): e74728.
Halene, S., Y. Gao, et al. (2010). "Serum response factor is an essential transcription factor in megakaryocytic maturation." Blood **116**(11): 1942-1950.
Hall, J. G., J. Levin, et al. (1969). "Thrombocytopenia with absent radius (TAR)." Medicine (Baltimore) **48**(6): 411-439.
Hamamy, H., P. Makrythanasis, et al. (2014). "Recessive thrombocytopenia likely due to a homozygous pathogenic variant in the FYB gene: case report." BMC Med Genet **15**: 135.
Harker, L. A., U. M. Marzec, et al. (1996). "Dose-response effects of pegylated human megakaryocyte growth and development factor on platelet production and function in nonhuman primates." Blood **88**(2): 511-521.
Harrison, C. J., A. V. Moorman, et al. (2005). "Interphase molecular cytogenetic screening for chromosomal abnormalities of prognostic significance in childhood acute lymphoblastic leukaemia: a UK Cancer Cytogenetics Group Study." Br J Haematol **129**(4): 520-530.
Harrison, P. (2005). "The role of PFA-100 testing in the investigation and management of haemostatic defects in children and adults." Br J Haematol **130**(1): 3-10.
Harrison, P. and E. M. Cramer (1993). "Platelet alpha-granules." Blood Rev **7**(1): 52-62.
Hart, S. M. and L. Foroni (2002). "Core binding factor genes and human leukemia." Haematologica **87**(12): 1307-1323.

Hartwig, J. H. and M. DeSisto (1991). "The cytoskeleton of the resting human blood platelet: structure of the membrane skeleton and its attachment to actin filaments." J Cell Biol **112**(3): 407-425.

Haudek, V. J., A. Slany, et al. (2009). "Proteome maps of the main human peripheral blood constituents." J Proteome Res **8**(8): 3834-3843.

Hayward, C. P. (2005). "Diagnosis and management of mild bleeding disorders." Hematology Am Soc Hematol Educ Program: 423-428.

Holmsen, H., I. Holmsen, et al. (1966). "Microdetermination of adenosine diphosphate and adenosine triphosphate in plasma with firefly luciferase system." Anal Biochem **17**(3): 456-473.

Horvat-Switzer, R. D. and A. A. Thompson (2006). "HOXA11 mutation in amegakaryocytic thrombocytopenia with radio-ulnar synostosis syndrome inhibits megakaryocytic differentiation in vitro." Blood Cells Mol Dis **37**(1): 55-63.

Hosokawa, K., T. Ohnishi, et al. (2011). "A novel automated microchip flow-chamber system to quantitatively evaluate thrombus formation and antithrombotic agents under blood flow conditions." J Thromb Haemost **9**(10): 2029-2037.

Ihara, K., E. Ishii, et al. (1999). "Identification of mutations in the c-mpl gene in congenital amegakaryocytic thrombocytopenia." Proc Natl Acad Sci U S A **96**(6): 3132-3136.

Imperato, M. R., P. Cauchy, et al. (2015). "The RUNX1-PU.1 axis in the control of hematopoiesis." Int J Hematol **101**(4): 319-329.

Irby, R. B., W. Mao, et al. (1999). "Activating SRC mutation in a subset of advanced human colon cancers." Nat Genet **21**(2): 187-190.

Italiano, J. E., Jr., P. Lecine, et al. (1999). "Blood platelets are assembled principally at the ends of proplatelet processes produced by differentiated megakaryocytes." J Cell Biol **147**(6): 1299-1312.

Izumi, R., T. Niihori, et al. (2014). "GNE myopathy associated with congenital thrombocytopenia: a report of two siblings." Neuromuscul Disord **24**(12): 1068-1072.

Jalagadugula, G., G. Mao, et al. (2010). "Regulation of platelet myosin light chain (MYL9) by RUNX1: implications for thrombocytopenia and platelet dysfunction in RUNX1 haplodeficiency." Blood **116**(26): 6037-6045.

Jan, A. (2012). "Deficiency of the Y14 protein is a critical factor underlying the etiology of thrombocytopenia with absent radii syndrome." Clin Genet **82**(1): 29-30.

Jarvis, G. E., D. Bihan, et al. (2012). "A role for adhesion and degranulation-promoting adapter protein in collagen-induced platelet activation mediated via integrin alpha(2) beta(1)." J Thromb Haemost **10**(2): 268-277.

Jay, D., E. J. Garcia, et al. (2000). "Determination of a cAMP-dependent protein kinase phosphorylation site in the C-terminal region of human endothelial actin-binding protein." Arch Biochem Biophys **377**(1): 80-84.

Jelkmann, W. (2001). "The role of the liver in the production of thrombopoietin compared with erythropoietin." Eur J Gastroenterol Hepatol **13**(7): 791-801.

Jiang, X. and X. Wang (2004). "Cytochrome C-mediated apoptosis." Annu Rev Biochem **73**: 87-106.

Johnson, B., G. C. Lowe, et al. (2016). "Whole exome sequencing identifies genetic variants in inherited thrombocytopenia with secondary qualitative function defects." Haematologica.

Josefsson, E. C., D. L. Burnett, et al. (2014). "Platelet production proceeds independently of the intrinsic and extrinsic apoptosis pathways." Nat Commun **5**: 3455.

Junt, T., H. Schulze, et al. (2007). "Dynamic visualization of thrombopoiesis within bone marrow." Science **317**(5845): 1767-1770.
Kaleelrahman, M., A. Minford, et al. (2004). "Use of recombinant factor VIIa in inherited platelet disorders." Br J Haematol **125**(1): 95-96.
Kasirer-Friede, A., B. Moran, et al. (2007). "ADAP is required for normal alphaIIbbeta3 activation by VWF/GP Ib-IX-V and other agonists." Blood **109**(3): 1018-1025.
Kaushansky, K. (2008). "Historical review: megakaryopoiesis and thrombopoiesis." Blood **111**(3): 981-986.
Kaushansky, K., S. Lok, et al. (1994). "Promotion of megakaryocyte progenitor expansion and differentiation by the c-Mpl ligand thrombopoietin." Nature **369**(6481): 568-571.
Kehrel, E. and F. Brodde (2013). "State of the art in platelet function testing." Transfus Med Hemother **40**(2): 73-86.
King, S., M. Germeshausen, et al. (2005). "Congenital amegakaryocytic thrombocytopenia: a retrospective clinical analysis of 20 patients." Br J Haematol **131**(5): 636-644.
Klopocki, E., H. Schulze, et al. (2007). "Complex inheritance pattern resembling autosomal recessive inheritance involving a microdeletion in thrombocytopenia-absent radius syndrome." Am J Hum Genet **80**(2): 232-240.
Kopp, H. G., S. T. Avecilla, et al. (2005). "The bone marrow vascular niche: home of HSC differentiation and mobilization." Physiology (Bethesda) **20**: 349-356.
Kottke-Marchant, K. and G. Corcoran (2002). "The laboratory diagnosis of platelet disorders." Arch Pathol Lab Med **126**(2): 133-146.
Krawczak, M., J. Reiss, et al. (1992). "The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences." Hum Genet **90**(1-2): 41-54.
Kumar, P., S. Henikoff, et al. (2009). "Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm." Nat Protoc **4**(7): 1073-1081.
Kundu, S. K., E. J. Heilmann, et al. (1995). "Description of an in vitro platelet function analyzer--PFA-100." Semin Thromb Hemost **21 Suppl 2**: 106-112.
Kunishima, S., H. Kashiwagi, et al. (2011). "Heterozygous ITGA2B R995W mutation inducing constitutive activation of the alphaIIbbeta3 receptor affects proplatelet formation and causes congenital macrothrombocytopenia." Blood **117**(20): 5479-5484.
Kunishima, S., R. Kobayashi, et al. (2009). "Mutation of the beta1-tubulin gene associated with congenital macrothrombocytopenia affecting microtubule assembly." Blood **113**(2): 458-461.
Kunishima, S., T. Matsushita, et al. (2003). "Immunofluorescence analysis of neutrophil nonmuscle myosin heavy chain-A in MYH9 disorders: association of subcellular localization with MYH9 mutations." Lab Invest **83**(1): 115-122.
Kunishima, S., T. Naoe, et al. (2001). "Novel heterozygous missense mutation in the platelet glycoprotein Ib beta gene associated with isolated giant platelet disorder." Am J Hematol **68**(4): 249-255.
Kunishima, S., S. Nishimura, et al. (2014). "TUBB1 mutation disrupting microtubule assembly impairs proplatelet formation and results in congenital macrothrombocytopenia." Eur J Haematol **92**(4): 276-282.
Kunishima, S., Y. Okuno, et al. (2013). "ACTN1 mutations cause congenital macrothrombocytopenia." Am J Hum Genet **92**(3): 431-438.

Laricchia-Robbio, L., R. Fazzina, et al. (2006). "Point mutations in two EVI1 Zn fingers abolish EVI1-GATA1 interaction and allow erythroid differentiation of murine bone marrow cells." Mol Cell Biol **26**(20): 7658-7666.

Laricchia-Robbio, L., K. Premanand, et al. (2009). "EVI1 Impairs myelopoiesis by deregulation of PU.1 function." Cancer Res **69**(4): 1633-1642.

Latger-Cannard, V., C. Philippe, et al. (2016). "Haematological spectrum and genotype-phenotype correlations in nine unrelated families with RUNX1 mutations from the French network on inherited platelet disorders." Orphanet J Rare Dis **11**: 49.

Lelieveld, S. H., M. Spielmann, et al. (2015). "Comparison of Exome and Genome Sequencing Technologies for the Complete Capture of Protein-Coding Regions." Hum Mutat **36**(8): 815-822.

Levin, C., A. Koren, et al. (2015). "Deleterious mutation in the FYB gene is associated with congenital autosomal recessive small-platelet thrombocytopenia." J Thromb Haemost **13**(7): 1285-1292.

Levy, G. G., W. C. Nichols, et al. (2001). "Mutations in a member of the ADAMTS gene family cause thrombotic thrombocytopenic purpura." Nature **413**(6855): 488-494.

Li, H., B. Handsaker, et al. (2009). "The Sequence Alignment/Map format and SAMtools." Bioinformatics **25**(16): 2078-2079.

Lim, R. P., A. Misra, et al. (2007). "Analysis of conformational changes in WASP using a split YFP." Biochem Biophys Res Commun **362**(4): 1085-1089.

Locatelli, F., G. Rossi, et al. (2003). "Hematopoietic stem-cell transplantation for the Bernard-Soulier syndrome." Ann Intern Med **138**(1): 79.

Lopez, J. A., R. K. Andrews, et al. (1998). "Bernard-Soulier syndrome." Blood **91**(12): 4397-4418.

Lordier, L., D. Bluteau, et al. (2012). "RUNX1-induced silencing of non-muscle myosin heavy chain IIB contributes to megakaryocyte polyploidization." Nat Commun **3**: 717.

Lordier, L., A. Jalil, et al. (2008). "Megakaryocyte endomitosis is a failure of late cytokinesis related to defects in the contractile ring and Rho/Rock signaling." Blood **112**(8): 3164-3174.

Lutskiy, M. I., F. S. Rosen, et al. (2005). "Genotype-proteotype linkage in the Wiskott-Aldrich syndrome." J Immunol **175**(2): 1329-1336.

Maclachlan, A., S. P. Watson, et al. (2016). "Inherited platelet disorders: Insight from platelet genomics using next-generation sequencing." Platelets: 1-6.

Manchev, V. T., M. Hilpert, et al. (2014). "A new form of macrothrombocytopenia induced by a germ-line mutation in the PRKACG gene." Blood **124**(16): 2554-2563.

Marconi, C., C. A. Di Buduo, et al. (2016). "SLFN14-related thrombocytopenia: identification within a large series of patients with inherited thrombocytopenia." Thromb Haemost **115**(5): 1076-1079.

Markello, T., D. Chen, et al. (2015). "York platelet syndrome is a CRAC channelopathy due to gain-of-function mutations in STIM1." Mol Genet Metab **114**(3): 474-482.

Mason, K. D., M. R. Carpinelli, et al. (2007). "Programmed anuclear cell death delimits platelet life span." Cell **128**(6): 1173-1186.

McElwee, J. L., S. Mohanan, et al. (2012). "Identification of PADI2 as a potential breast cancer biomarker and therapeutic target." BMC Cancer **12**: 500.

Melazzini, F., F. Palombo, et al. (2016). "Clinical and pathogenetic features of ETV6 related thrombocytopenia with predisposition to acute lymphoblastic leukemia." Haematologica.

Michaud, J., F. Wu, et al. (2002). "In vitro analyses of known and novel RUNX1/AML1 mutations in dominant familial platelet disorder with predisposition to acute myelogenous leukemia: implications for mechanisms of pathogenesis." Blood **99**(4): 1364-1372.

Misceo, D., A. Holmgren, et al. (2014). "A dominant STIM1 mutation causes Stormorken syndrome." Hum Mutat **35**(5): 556-564.

Mizutani, H., T. Furubayashi, et al. (1992). "Mechanisms of corticosteroid action in immune thrombocytopenic purpura (ITP): experimental studies using ITP-prone mice, (NZW x BXSB) F1." Blood **79**(4): 942-947.

Moran, N., P. A. Morateck, et al. (2000). "Surface expression of glycoprotein ib alpha is dependent on glycoprotein ib beta: evidence from a novel mutation causing Bernard-Soulier syndrome." Blood **96**(2): 532-539.

Morison, I. M., E. M. Cramer Borde, et al. (2008). "A mutation of human cytochrome c enhances the intrinsic apoptotic pathway but causes only thrombocytopenia." Nat Genet **40**(4): 387-389.

Moroy, T., L. Vassen, et al. (2015). "From cytopenia to leukemia: the role of Gfi1 and Gfi1b in blood formation." Blood **126**(24): 2561-2569.

Mullis, K., F. Faloona, et al. (1986). "Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction." Cold Spring Harb Symp Quant Biol **51 Pt 1**: 263-273.

Muraoka, K., E. Ishii, et al. (1997). "Defective response to thrombopoietin and impaired expression of c-mpl mRNA of bone marrow cells in congenital amegakaryocytic thrombocytopenia." Br J Haematol **96**(2): 287-292.

Murphy, A. C., A. J. Lindsay, et al. (2016). "Congenital macrothrombocytopenia-linked mutations in the actin-binding domain of alpha-actinin-1 enhance F-actin association." FEBS Lett **590**(6): 685-695.

Nagle, D. L., M. A. Karim, et al. (1996). "Identification and mutation analysis of the complete gene for Chediak-Higashi syndrome." Nat Genet **14**(3): 307-311.

Ng, S. B., E. H. Turner, et al. (2009). "Targeted capture and massively parallel sequencing of 12 human exomes." Nature **461**(7261): 272-276.

Nichols, K. E., J. D. Crispino, et al. (2000). "Familial dyserythropoietic anaemia and thrombocytopenia due to an inherited mutation in GATA1." Nat Genet **24**(3): 266-270.

Nieswandt, B. and S. P. Watson (2003). "Platelet-collagen interaction: is GPVI the central receptor?" Blood **102**(2): 449-461.

Niihori, T., M. Ouchi-Uchiyama, et al. (2015). "Mutations in MECOM, Encoding Oncoprotein EVI1, Cause Radioulnar Synostosis with Amegakaryocytic Thrombocytopenia." Am J Hum Genet **97**(6): 848-854.

Noris, P., G. Biino, et al. (2014). "Platelet diameters in inherited thrombocytopenias: analysis of 376 patients with all known disorders." Blood **124**(6): e4-e10.

Noris, P., R. Favier, et al. (2013). "ANKRD26-related thrombocytopenia and myeloid malignancies." Blood **122**(11): 1987-1989.

Noris, P., S. Perrotta, et al. (2011). "Mutations in ANKRD26 are responsible for a frequent form of inherited thrombocytopenia: analysis of 78 patients from 21 families." Blood **117**(24): 6673-6680.

North, T. E., T. Stacy, et al. (2004). "Runx1 is expressed in adult mouse hematopoietic stem cells and differentiating myeloid and lymphoid cells, but not in maturing erythroid cells." Stem Cells **22**(2): 158-168.

Nurden, A. T., M. Fiore, et al. (2011). "Glanzmann thrombasthenia: a review of ITGA2B and ITGB3 defects with emphasis on variants, phenotypic variability, and mouse models." Blood **118**(23): 5996-6005.
Nurden, P., N. Debili, et al. (2011). "Thrombocytopenia resulting from mutations in filamin A can be expressed as an isolated syndrome." Blood **118**(22): 5928-5937.
Nurden, P., N. Debili, et al. (2006). "Impaired megakaryocytopoiesis in type 2B von Willebrand disease with severe thrombocytopenia." Blood **108**(8): 2587-2595.
Othman, M. (2011). "Platelet-type von Willebrand disease: a rare, often misdiagnosed and underdiagnosed bleeding disorder." Semin Thromb Hemost **37**(5): 464-469.
Paravicini, T. M., V. Chubanov, et al. (2012). "TRPM7: a unique channel involved in magnesium homeostasis." Int J Biochem Cell Biol **44**(8): 1381-1384.
Park, H., J. I. Kim, et al. (2010). "Discovery of common Asian copy number variants using integrated high-resolution array CGH and massively parallel DNA sequencing." Nat Genet **42**(5): 400-405.
Passegue, E., A. J. Wagers, et al. (2005). "Global analysis of proliferation and cell cycle gene expression in the regulation of hematopoietic stem and progenitor cell fates." J Exp Med **202**(11): 1599-1611.
Patel, S. R., J. L. Richardson, et al. (2005). "Differential roles of microtubule assembly and sliding in proplatelet formation by megakaryocytes." Blood **106**(13): 4076-4085.
Pecci, A., P. Gresele, et al. (2010). "Eltrombopag for the treatment of the inherited thrombocytopenia deriving from MYH9 mutations." Blood **116**(26): 5832-5837.
Pecci, A., E. Panza, et al. (2008). "Position of nonmuscle myosin heavy chain IIA (NMMHC-IIA) mutations predicts the natural history of MYH9-related disease." Hum Mutat **29**(3): 409-417.
Perez Botero, J., R. K. Pruthi, et al. (2016). "Practice patterns in the diagnosis of inherited platelet disorders within a single institution." Blood Coagul Fibrinolysis.
Pertuy, F., A. Eckly, et al. (2014). "Myosin IIA is critical for organelle distribution and F-actin organization in megakaryocytes and platelets." Blood **123**(8): 1261-1269.
Peyruchaud, O., A. T. Nurden, et al. (1998). "R to Q amino acid substitution in the GFFKR sequence of the cytoplasmic domain of the integrin IIb subunit in a patient with a Glanzmann's thrombasthenia-like syndrome." Blood **92**(11): 4178-4187.
Pippucci, T., A. Savoia, et al. (2011). "Mutations in the 5' UTR of ANKRD26, the ankirin repeat domain 26 gene, cause an autosomal-dominant form of inherited thrombocytopenia, THC2." Am J Hum Genet **88**(1): 115-120.
Plagnol, V., J. Curtis, et al. (2012). "A robust model for read count data in exome sequencing experiments and implications for copy number variant calling." Bioinformatics **28**(21): 2747-2754.
Preudhomme, C., A. Renneville, et al. (2009). "High frequency of RUNX1 biallelic alteration in acute myeloid leukemia secondary to familial platelet disorder." Blood **113**(22): 5583-5587.
Provan, D., R. Stasi, et al. (2010). "International consensus report on the investigation and management of primary immune thrombocytopenia." Blood **115**(2): 168-186.
Raccuglia, G. (1971). "Gray platelet syndrome. A variety of qualitative platelet disorder." Am J Med **51**(6): 818-828.
Radley, J. M. and C. J. Haller (1982). "The demarcation membrane system of the megakaryocyte: a misnomer?" Blood **60**(1): 213-219.

Ramesh, N., I. M. Anton, et al. (1997). "WIP, a protein associated with wiskott-aldrich syndrome protein, induces actin polymerization and redistribution in lymphoid cells." Proc Natl Acad Sci U S A **94**(26): 14671-14676.

Rao, A. K., S. Ghosh, et al. (1995). "Mechanisms of platelet dysfunction and response to DDAVP in patients with congenital platelet function defects. A double-blind placebo-controlled trial." Thromb Haemost **74**(4): 1071-1078.

Raslova, H., E. Komura, et al. (2004). "FLI1 monoallelic expression combined with its hemizygous loss underlies Paris-Trousseau/Jacobsen thrombopenia." J Clin Invest **114**(1): 77-84.

Raslova, H., L. Roy, et al. (2003). "Megakaryocyte polyploidization is associated with a functional gene amplification." Blood **101**(2): 541-544.

Ravn, H. B., S. D. Kristensen, et al. (1996). "Magnesium inhibits human platelets." Blood Coagul Fibrinolysis **7**(2): 241-244.

Record, J., D. Malinova, et al. (2015). "Immunodeficiency and severe susceptibility to bacterial infection associated with a loss-of-function homozygous mutation of MKL1." Blood **126**(13): 1527-1535.

Reed, G. L., M. L. Fitzgerald, et al. (2000). "Molecular mechanisms of platelet exocytosis: insights into the "secrete" life of thrombocytes." Blood **96**(10): 3334-3342.

Remijn, J. A., Y. P. Wu, et al. (2002). "Role of ADP receptor P2Y(12) in platelet adhesion and thrombus formation in flowing blood." Arterioscler Thromb Vasc Biol **22**(4): 686-691.

Richards, S., N. Aziz, et al. (2015). "Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology." Genet Med **17**(5): 405-424.

Richardson, J. L., R. A. Shivdasani, et al. (2005). "Mechanisms of organelle transport and capture along proplatelets during platelet production." Blood **106**(13): 4066-4075.

Robert, A., V. Cortin, et al. (2012). "Megakaryocyte and platelet production from human cord blood stem cells." Methods Mol Biol **788**: 219-247.

Robinson, J. T., H. Thorvaldsdottir, et al. (2011). "Integrative genomics viewer." Nat Biotechnol **29**(1): 24-26.

Robinson, P. N., S. Kohler, et al. (2008). "The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease." Am J Hum Genet **83**(5): 610-615.

Rodeghiero, F., A. Tosetto, et al. (2010). "ISTH/SSC Bleeding Assessment Tool: A Standardized Questionnaire and a Proposal for a New Bleeding Score for Inherited Bleeding Disorders." J Thromb Haemost: in press.

Romo, G. M., J. F. Dong, et al. (1999). "The glycoprotein Ib-IX-V complex is a platelet counterreceptor for P-selectin." J Exp Med **190**(6): 803-814.

Rossetti, S. and N. Sacchi (2013). "RUNX1: A microRNA hub in normal and malignant hematopoiesis." Int J Mol Sci **14**(1): 1566-1588.

Ruggeri, Z. M. (1993). "Mechanisms of shear-induced platelet adhesion and aggregation." Thromb Haemost **70**(1): 119-123.

Ruggeri, Z. M. and G. L. Mendolicchio (2007). "Adhesion mechanisms in platelet function." Circ Res **100**(12): 1673-1685.

Saleque, S., S. Cameron, et al. (2002). "The zinc-finger proto-oncogene Gfi-1b is essential for development of the erythroid and megakaryocytic lineages." Genes Dev **16**(3): 301-306.

Sanger, F., S. Nicklen, et al. (1977). "DNA sequencing with chain-terminating inhibitors." Proc Natl Acad Sci U S A **74**(12): 5463-5467.
Savoia, A., S. Kunishima, et al. (2014). "Spectrum of the mutations in Bernard-Soulier syndrome." Hum Mutat **35**(9): 1033-1045.
Schachtner, H., S. D. Calaminus, et al. (2013). "Megakaryocytes assemble podosomes that degrade matrix and protrude through basement membrane." Blood **121**(13): 2542-2552.
Schiff, D. E., W. D. Roberts, et al. (2004). "Thrombocytopenia and severe hyperbilirubinemia in the neonatal period secondary to congenital thrombotic thrombocytopenic purpura and ADAMTS13 deficiency." J Pediatr Hematol Oncol **26**(8): 535-538.
Schulze, H., M. Korpal, et al. (2006). "Characterization of the megakaryocyte demarcation membrane system and its role in thrombopoiesis." Blood **107**(10): 3868-3875.
Schwarz, J. M., D. N. Cooper, et al. (2014). "MutationTaster2: mutation prediction for the deep-sequencing age." Nat Methods **11**(4): 361-362.
Schwer, H. D., P. Lecine, et al. (2001). "A lineage-restricted and divergent beta-tubulin isoform is essential for the biogenesis, structure and function of blood platelets." Curr Biol **11**(8): 579-586.
Schwertz, H., S. Koster, et al. (2010). "Anucleate platelets generate progeny." Blood **115**(18): 3801-3809.
Senis, Y. A., A. Mazharian, et al. (2014). "Src family kinases: at the forefront of platelet activation." Blood **124**(13): 2013-2024.
Senyuk, V., K. K. Sinha, et al. (2007). "Repression of RUNX1 activity by EVI1: a new role of EVI1 in leukemogenesis." Cancer Res **67**(12): 5658-5666.
Seppala, R., V. P. Lehto, et al. (1999). "Mutations in the human UDP-N-acetylglucosamine 2-epimerase gene define the disease sialuria and the allosteric site of the enzyme." Am J Hum Genet **64**(6): 1563-1569.
Seri, M., R. Cusano, et al. (2000). "Mutations in MYH9 result in the May-Hegglin anomaly, and Fechtner and Sebastian syndromes. The May-Hegglin/Fechtner Syndrome Consortium." Nat Genet **26**(1): 103-105.
Seri, M., A. Pecci, et al. (2003). "MYH9-related disease: May-Hegglin anomaly, Sebastian syndrome, Fechtner syndrome, and Epstein syndrome are not distinct entities but represent a variable expression of a single illness." Medicine (Baltimore) **82**(3): 203-215.
Shaikh, T. H., X. Gai, et al. (2009). "High-resolution mapping and analysis of copy number variations in the human genome: a data resource for clinical and research applications." Genome Res **19**(9): 1682-1690.
Sharrocks, A. D. (2001). "The ETS-domain transcription factor family." Nat Rev Mol Cell Biol **2**(11): 827-837.
Shechter, M. (2003). "Does magnesium have a role in the treatment of patients with coronary artery disease?" Am J Cardiovasc Drugs **3**(4): 231-239.
Shigemizu, D., Y. Momozawa, et al. (2015). "Performance comparison of four commercial human whole-exome capture platforms." Sci Rep **5**: 12742.
Simeoni, I., J. C. Stephens, et al. (2016). "A high-throughput sequencing test for diagnosing inherited bleeding, thrombotic, and platelet disorders." Blood **127**(23): 2791-2803.
Sims, P. J., T. Wiedmer, et al. (1989). "Assembly of the platelet prothrombinase complex is linked to vesiculation of the platelet plasma membrane. Studies in Scott syndrome: an isolated defect in platelet procoagulant activity." J Biol Chem **264**(29): 17049-17057.

Singbartl, K., S. B. Forlow, et al. (2001). "Platelet, but not endothelial, P-selectin is critical for neutrophil-mediated acute postischemic renal failure." FASEB J **15**(13): 2337-2344.

Singleton, B. K., D. J. Roxby, et al. (2013). "A novel GATA1 mutation (Stop414Arg) in a family with the rare X-linked blood group Lu(a-b-) phenotype and mild macrothrombocytic thrombocytopenia." Br J Haematol **161**(1): 139-142.

Smith, E. C., J. N. Thon, et al. (2012). "MKL1 and MKL2 play redundant and crucial roles in megakaryocyte maturation and platelet formation." Blood **120**(11): 2317-2329.

Smyth, S. S., R. P. McEver, et al. (2009). "Platelet functions beyond hemostasis." J Thromb Haemost **7**(11): 1759-1766.

Solar, G. P., W. G. Kerr, et al. (1998). "Role of c-mpl in early hematopoiesis." Blood **92**(1): 4-10.

Sole, G., I. Coupry, et al. (2009). "Bilateral periventricular nodular heterotopia in France: frequency of mutations in FLNA, phenotypic heterogeneity and spectrum of mutations." J Neurol Neurosurg Psychiatry **80**(12): 1394-1398.

Song, W., S. A. Gardner, et al. (2016). "Exploring the landscape of pathogenic genetic variation in the ExAC population database: insights of relevance to variant classification." Genet Med **18**(8): 850-854.

Song, W. J., M. G. Sullivan, et al. (1999). "Haploinsufficiency of CBFA2 causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia." Nat Genet **23**(2): 166-175.

Spinler, K. R., J. W. Shin, et al. (2015). "Myosin-II repression favors pre/proplatelets but shear activation generates platelets and fails in macrothrombocytopenia." Blood **125**(3): 525-533.

Starck, J., N. Cohet, et al. (2003). "Functional cross-antagonism between transcription factors FLI-1 and EKLF." Mol Cell Biol **23**(4): 1390-1402.

Stevenson, W. S., M. C. Morel-Kopp, et al. (2013). "GFI1B mutation causes a bleeding disorder with abnormal platelet function." J Thromb Haemost **11**(11): 2039-2047.

Stockley, J., N. V. Morgan, et al. (2013). "Enrichment of FLI1 and RUNX1 mutations in families with excessive bleeding and platelet dense granule secretion defects." Blood **122**(25): 4090-4093.

Stoddart, M. T., P. Connor, et al. (2013). "Congenital amegakaryocytic thrombocytopenia (CAMT) presenting as severe pancytopenia in the first month of life." Pediatr Blood Cancer **60**(9): E94-96.

Stritt, S., P. Nurden, et al. (2016). "Defects in TRPM7 channel function deregulate thrombopoiesis through altered cellular Mg(2+) homeostasis and cytoskeletal architecture." Nat Commun **7**: 11097.

Stritt, S., P. Nurden, et al. (2016). "A gain-of-function variant in DIAPH1 causes dominant macrothrombocytopenia and hearing loss." Blood **127**(23): 2903-2914.

Tablin, F., M. J. Reeber, et al. (1988). "Platelets contain a 210K microtubule-associated protein related to a similar protein in HeLa cells." J Cell Sci **90 (Pt 2)**: 317-324.

Taichman, R. S. (2005). "Blood and bone: two tissues whose fates are intertwined to create the hematopoietic stem-cell niche." Blood **105**(7): 2631-2639.

Takayama, N. and K. Eto (2012). "In vitro generation of megakaryocytes and platelets from human embryonic stem cells and induced pluripotent stem cells." Methods Mol Biol **788**: 205-217.

Tchernitchko, D., M. Goossens, et al. (2004). "In silico prediction of the deleterious effect of a mutation: proceed with caution in clinical genetics." Clin Chem **50**(11): 1974-1978.

Thompson, A. A. and L. T. Nguyen (2000). "Amegakaryocytic thrombocytopenia and radio-ulnar synostosis are associated with HOXA11 mutation." Nat Genet **26**(4): 397-398.
Thompson, A. A., K. Woodruff, et al. (2001). "Congenital thrombocytopenia and radio-ulnar synostosis: a new familial syndrome." Br J Haematol **113**(4): 866-870.
Thon, J. N., M. T. Devine, et al. (2012). "High-content live-cell imaging assay used to establish mechanism of trastuzumab emtansine (T-DM1)--mediated inhibition of platelet production." Blood **120**(10): 1975-1984.
Thon, J. N., H. Macleod, et al. (2012). "Microtubule and cortical forces determine platelet size during vascular platelet production." Nat Commun **3**: 852.
Thon, J. N., A. Montalvo, et al. (2010). "Cytoskeletal mechanics of proplatelet maturation and platelet release." J Cell Biol **191**(4): 861-874.
Tijssen, M. R. and C. Ghevaert (2013). "Transcription factors in late megakaryopoiesis and related platelet disorders." J Thromb Haemost **11**(4): 593-604.
Topka, S., J. Vijai, et al. (2015). "Germline ETV6 Mutations Confer Susceptibility to Acute Lymphoblastic Leukemia and Thrombocytopenia." PLoS Genet **11**(6): e1005262.
Toth, O., A. Calatzis, et al. (2006). "Multiple electrode aggregometry: a new device to measure platelet aggregation in whole blood." Thromb Haemost **96**(6): 781-788.
Turro, E., D. Greene, et al. (2016). "A dominant gain-of-function mutation in universal tyrosine kinase SRC causes thrombocytopenia, myelofibrosis, bleeding, and bone pathologies." Sci Transl Med **8**(328): 328ra330.
van den Oudenrijn, S., M. Bruin, et al. (2000). "Mutations in the thrombopoietin receptor, Mpl, in children with congenital amegakaryocytic thrombocytopenia." Br J Haematol **110**(2): 441-448.
Varon, D., R. Dardik, et al. (1997). "A new method for quantitative analysis of whole blood platelet interaction with extracellular matrix under flow conditions." Thromb Res **85**(4): 283-294.
Venter, J. C., M. D. Adams, et al. (2001). "The sequence of the human genome." Science **291**(5507): 1304-1351.
Villa, A., L. Notarangelo, et al. (1995). "X-linked thrombocytopenia and Wiskott-Aldrich syndrome are allelic diseases with mutations in the WASP gene." Nat Genet **9**(4): 414-417.
Wagner, K. U., E. Claudio, et al. (2000). "Conditional deletion of the Bcl-x gene from erythroid cells results in hemolytic anemia and profound splenomegaly." Development **127**(22): 4949-4958.
Walker, L. C., J. Stevens, et al. (2002). "A novel inherited mutation of the transcription factor RUNX1 causes thrombocytopenia and may predispose to acute myeloid leukaemia." Br J Haematol **117**(4): 878-881.
Wall, M., K. C. Rayeroux, et al. (2012). "ETV6 deletion is a common additional abnormality in patients with myelodysplastic syndromes or acute myeloid leukemia and monosomy 7." Haematologica **97**(12): 1933-1936.
Ware, J., S. R. Russell, et al. (1990). "Nonsense mutation in the glycoprotein Ib alpha coding sequence associated with Bernard-Soulier syndrome." Proc Natl Acad Sci U S A **87**(5): 2026-2030.
Warr, A., C. Robert, et al. (2015). "Exome Sequencing: Current and Future Perspectives." G3 (Bethesda) **5**(8): 1543-1550.
Watanabe, Y., A. Nunokawa, et al. (2009). "A two-stage case-control association study of PADI2 with schizophrenia." J Hum Genet **54**(7): 430-432.

Wengler, G. S., L. D. Notarangelo, et al. (1995). "High prevalence of nonsense, frame shift, and splice-site mutations in 16 patients with full-blown Wiskott-Aldrich syndrome." Blood **86**(10): 3648-3654.
Whitfield, J. B. and N. G. Martin (1985). "Genetic and environmental influences on the size and number of cells in the blood." Genet Epidemiol **2**(2): 133-144.
Wiedmer, T., S. J. Shattil, et al. (1990). "Role of calcium and calpain in complement-induced vesiculation of the platelet plasma membrane and in the exposure of the platelet factor Va receptor." Biochemistry **29**(3): 623-632.
Willoughby, S., A. Holmes, et al. (2002). "Platelets and cardiovascular disease." Eur J Cardiovasc Nurs **1**(4): 273-288.
Wiskott, A. (1937). "Familiarer, angeborener Morbus Werlhofii?" Monatsschr. Kinderheilkd. **68**(212-216).
Yang, Y., D. M. Muzny, et al. (2013). "Clinical whole-exome sequencing for the diagnosis of mendelian disorders." N Engl J Med **369**(16): 1502-1511.
Youssefian, T., J. M. Masse, et al. (1997). "Platelet and megakaryocyte dense granules contain glycoproteins Ib and IIb-IIIa." Blood **89**(11): 4047-4057.
Yu, C., K. K. Niakan, et al. (2002). "X-linked thrombocytopenia with thalassemia from a mutation in the amino finger of GATA-1 affecting DNA binding rather than FOG-1 interaction." Blood **100**(6): 2040-2045.
Yu, M. and A. B. Cantor (2012). "Megakaryopoiesis and thrombopoiesis: an update on cytokines and lineage surface markers." Methods Mol Biol **788**: 291-303.
Zhang, L., M. Orban, et al. (2012). "A novel role of sphingosine 1-phosphate receptor S1pr1 in mouse thrombopoiesis." J Exp Med **209**(12): 2165-2181.
Zhang, M. Y., J. E. Churpek, et al. (2015). "Germline ETV6 mutations in familial thrombocytopenia and hematologic malignancy." Nat Genet **47**(2): 180-185.
Zhen, C., F. Guo, et al. (2014). "A family with distal myopathy with rimmed vacuoles associated with thrombocytopenia." Neurol Sci **35**(9): 1479-1481.
Zhu, Q., C. Watanabe, et al. (1997). "Wiskott-Aldrich syndrome/X-linked thrombocytopenia: WASP gene mutations, protein expression, and phenotype." Blood **90**(7): 2680-2689.

Appendix

Table A. Database of 358 platelet related disease and the percentage coverage 20x

Gene	Percentage covered 20x
ABCA12	95.78903846
ABCB4	91.87596154
ABCC4	94.33057692
ABCG5	94.33923077
ABCG8	94.58923077
ACSL4	89.83423077
ACTN1	95.97653846
ACVRL1	86.78846154
ADAMTS13	82.55192308

ADCY3	94.18192308
ADCY6	96.89115385
ADCY7	93.01884615
ADORA2B	99.71942308
ADRA2A	88.06634615
ADRA2B	99.77769231
ADRBK1	96.19711538
AK3	55.78346154
AKT1	96.86942308
AKT2	95.20153846
ALOX12	93.64807692
ANKRD12	93.75903846
ANKRD18A	33.42519231
ANKRD18B	33.72769231
ANKRD26	88.99711538
ANKRD33	99.03942308
AP3B1	94.50730769
AP3D1	91.79173077
AP3M1	96.0675
AP3S1	52.58692308
APC	98.35365385
ARHGAP1	92.67423077
ARHGAP17	90.04076923
ARHGAP32	97.34884615
ARHGAP6	90.92846154
ARHGDIA	99.87653846
ARHGDIB	94.91942308
ARHGEF12	94.94096154
ARHGEF3	92.61442308
ARRB1	96.75096154
ASPN	95.82096154
BAK1	84.71307692
BCL2L1	95.73134615
BCOR	95.31596154
BET1L	88.50326923
BLOC1S1	87.50596154
BLOC1S2	84.00711538
BLOC1S3	64.64711538
BLOC1S4	0
BLOC1S5	0
BLOC1S6	0
BMP4	99.61769231
BTBD9	88.92961538
BTK	90.95365385

C14orf133	93.49211538
C19orf55	95.45673077
C20orf42	98.17076923
C6orf25	94.98673077
CD226	96.20423077
CD36	95.57326923
CHD3	87.89211538
CLEC1B	84.17692308
CLEC4F	96.99903846
CNO	68.97596154
CSK	91.53903846
CTTN	92.78730769
CYCS	76.61653846
DAAM1	91.85519231
DIAPH1	90.83076923
DIAPH2	88.93423077
DIAPH3	90.49826923
DNAH11	94.54923077
DNM1L	90.95442308
DNM2	90.57903846
DNM3	93.96403846
DTNBP1	97.78865385
EFNB1	94.40038462
EPHA4	96.95403846
EPHB1	92.37596154
ERG	96.33038462
ETS1	96.40730769
ETV6	95.02826923
EXOC1	93.74442308
F2R	98.79557692
F2RL3	94.01692308
FARP2	95.73269231
FCER1G	83.29288462
FCGR2A	67.57211538
FERMT1	91.73211538
FERMT3	95.51115385
FGD3	94.28884615
FGR	98.005
FHOD1	96.28326923
FLI1	94.73076923
FLII	93.37019231
FLNA	97.00788462
FMNL1	81.43192308
FMNL3	92.46903846

FYN	95.37461538
GATA1	92.99692308
GDI2	85.11634615
GFI1	95.03865385
GFI1B	94.83923077
GNA12	91.09903846
GNA13	99.22730769
GNAI1	83.15596154
GNAI2	97.37134615
GNAQ	82.19326923
GNAZ	99.58403846
GNB2	95.55461538
GNB3	98.4325
GNG11	95.51576923
GNG12	99.955
GNG13	82.95346154
GNG5	57.10211538
GP1BA	98.47846154
GP1BB	48.24865385
GP5	94.20826923
GP6	87.45961538
GP9	85.64980769
GRAP2	92.87307692
GRB2	96.925
GRK5	93.96980769
GRK6	93.00576923
GUCY1A3	99.19346154
GUCY1B3	94.70807692
HBB	100
HOOK3	90.58788462
HOXA11	98.39903846
HPS1	90.14211538
HPS4	97.02096154
HTR2A	98.28846154
INPP5D	96.43903846
ITGA2	94.87115385
ITGA2B	90.47326923
ITGA5	92.47230769
ITGB1	91.83923077
ITGB3	91.50692308
ITPR1	94.55153846
JAK2	91.67096154
JMJD1C	96.59538462
KIAA1109	96.005

KIAA2018	99.5275
LAIR1	92.38615385
LAT	96.34288462
LCP2	90.20923077
LPAR1	98.79923077
LTBP1	93.08288462
LY6G6F	98.07923077
LYN	90.19884615
LYST	95.22557692
MAP2K2	94.87903846
MAP2K4	71.90673077
MAP3K9	93.925
MAPK1	96.88
MAPK13	98.06653846
MAPK14	98.94038462
MAPK8	96.26557692
MDS1	0
MECOM	97.42153846
MKL1	92.5525
MLK1	0
MLPH	83.94865385
MMP17	83.83634615
MNX1	64.92788462
MPL	95.89403846
MRPS34	96.49923077
MUC16	98.12096154
MUC2	92.51884615
MUTED	95.05711538
MYB	92.59384615
MYH10	95.62423077
MYH13	88.64865385
MYH9	96.80057692
MYL9	90.89884615
MYLK	95.07519231
MYLK2	93.62173077
MYO18B	89.13384615
MYO3A	93.55019231
MYO5A	94.82423077
MYO5B	87.30557692
NAPA	95.39730769
NAPG	92.46403846
NBEA	86.9125
NBEAL2	97.56807692
NFE2	99.05153846

NIPSNAP3A	90.91269231
NOTCH1	89.38865385
NOX1	88.43692308
NRG3	96.13076923
NSF	37.36192308
NXF1	94.20769231
ORAI1	92.00634615
P2RX1	96.11173077
P2RY1	99.26365385
P2RY12	100
P2RY13	100
PDE2A	90.25076923
PDE3A	90.60538462
PDE4D	93.30769231
PDE5A	95.16384615
PDPK1	36.81807692
PDZD3	95.2925
PDZK1	15.58576923
PEAR1	96.45153846
PECAM1	100
PGM3	94.78634615
PHOX2A	54.54903846
PIK3CA	90.36384615
PIK3CB	94.71692308
PIK3CD	87.63288462
PIK3CG	98.25461538
PIK3R1	95.99730769
PIK3R3	96.83307692
PIK3R5	96.96980769
PLA2G4A	96.91615385
PLA2G4C	93.49153846
PLCB2	97.0875
PLCB3	92.18576923
PLCG2	94.48692308
PLDN	75.15980769
PPP1CA	99.00038462
PPP1CB	97.86730769
PPP1CC	98.79423077
PPP1R12A	91.18615385
PPP1R12C	78.43576923
PPP1R14A	85.47346154
PPP1R2	45.66384615
PRKACA	96.39538462
PRKACB	92.21788462

PRKACG	100
PRKAR1A	77.49673077
PRKAR2A	94.83557692
PRKCA	92.09307692
PRKCB	94.39903846
PRKCD	96.76076923
PRKCQ	91.7475
PRKD1	98.66307692
PRKG1	91.3625
PRKG2	95.72442308
PTEN	83.83673077
PTGIR	92.86269231
PTGS1	99.03711538
PTK2	93.98653846
PTPN1	93.99519231
PTPN11	55.94115385
PTPN12	95.05057692
PTPN18	91.53596154
PTPN2	70.88884615
PTPN6	94.96384615
PTPN7	88.78
PTPN9	92.46019231
PTPRA	97.68
PTPRC	91.14269231
PTPRJ	92.07307692
RAB27A	99.31019231
RAB27B	97.62057692
RAB38	99.27326923
RAB4A	90.98846154
RABGGTA	90.39365385
RAC1	56.83538462
RAF1	93.3425
RAI1	98.35096154
RAP1B	58.62884615
RAP1GAP	92.23673077
RAP1GAP2	94.07826923
RAP1GDS1	85.94403846
RASGRP2	90.82
RBM8A	85.60326923
RGS10	90.32192308
RGS18	94.02134615
RGS19	96.77903846
RGS20	91.42634615
RGS9	92.94538462

RHOA	95.28730769
RHOC	96.63519231
RHOF	92.28615385
ROCK1	86.35980769
ROCK2	94.94538462
RUNX1	89.41538462
SCAMP2	87.93230769
SCAMP5	85.06346154
SCFD1	94.55307692
SELP	94.06769231
SERPINE2	93.29230769
SH2B3	87.26384615
SIRPA	74.93711538
SLC35D3	90.26326923
SLC9A3R1	94.555
SLC9A3R2	75.83384615
SLFN14	95.47711538
SMAD1	97.22134615
SMAD6	79.03769231
SNAP23	89.95788462
SNAP25	89.25403846
SNAP29	99.19
SNAPIN	93.63538462
SNX1	88.6075
SRA1	94.33307692
SRC	88.58730769
SRF	91.81730769
STIM1	90.21519231
STOM	91.51615385
STX11	99.55711538
STX12	96.40942308
STX2	90.69057692
STX4	93.61519231
STX6	96.60596154
STX7	98.15153846
STXBP1	95.39442308
STXBP2	96.60769231
STXBP3	93.66923077
STXBP4	93.48961538
STXBP5L	95.54057692
STXBP6	88.68615385
SUZ12	70.34346154
SYK	97.63346154
SYTL3	91.77326923

SYTL4	93.77846154
TAL1	76.37923077
TAOK1	93.57307692
TBXA2R	82.61076923
TEC	91.73115385
TGFBR3	94.92346154
THPO	90.60980769
TLN1	97.19653846
TLR2	98.30980769
TMCC2	90.78711538
TPM1	96.63076923
TPM4	77.58057692
TRAF4	95.36480769
TREML1	98.26788462
TRPM7	94.52211538
TTC37	97.67211538
TUBA3C	76.80365385
TUBB1	99.98019231
UNC13A	94.45807692
UNC13B	97.44692308
VAMP2	99.25480769
VAMP3	98.30519231
VAMP7	16.34615385
VAMP8	84.83596154
VAV1	91.29730769
VAV2	95.41961538
VAV3	94.33903846
VPS11	98.45057692
VPS16	94.97403846
VPS18	98.01461538
VPS33A	95.89807692
VPS33B	95.43365385
VPS39	94.17423077
VPS41	91.65442308
VPS4B	97.05673077
VPS52	92.45634615
VPS8	90.68557692
VWF	75.24192308
WAS	82.84884615
WDR66	92.87173077
ZFPMI	66.54634615

UNIVERSITY OF BIRMINGHAM

Birmingham Platelet Group

Participant Information Sheet on ‘Mild Platelet Disorders’

PART 1

1. Study title: Mild bleeding disorders caused by platelet defects

2. Invitation paragraph:

You are being invited to take part in a study. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with friends, relatives, your GP and your Consultant if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

The UK Clinical Research Collaboration (UK CRC) publishes a leaflet entitled "Understanding Clinical Trials”. This leaflet gives more information about medical research and looks at some questions you may wish to ask. A copy may be obtained

from the internet at http://www.ukcrc.org/publications/informationbooklets. If you do not have access to the Internet, please let us know and we will print a copy off for you.

3. What is the purpose of the study?

This study aims to understand why some patients bleed excessively. We are going to look at the following groups of patients:

1. Patients with mild bleeding who have been seen by a blood specialist who feels their symptoms could be due to a problem in the platelets ("clinically diagnosed platelet defects")
2. Patients who have unexpectedly heavy periods with no obvious cause
3. Patients with lifelong low platelet counts in the blood with no known cause

In addition, we will also include some "control" individuals who do not suffer from the above problems and are important to allow us to compare responses between those who bleed excessively and those who do not. If we only examined affected patients it would be difficult to tell how their response compared to that expected in a normal individual.

You will be included in the highlighted group above, if you choose to participate in this study.

For all populations except the control group, we suspect that you have a problem in small cells in the blood, known as platelets, but we are not certain that this is the case. This study will use new and more specialist tests in an attempt to confirm that there is a problem in your platelets and to use this information to help identify the gene or genes that cause this problem. More information on platelets and what the study aims to achieve is given below:

What are platelets?

Platelets are small cells within the blood that are involved in the clotting mechanism. Following injury, platelets 'stick' to each other and block (or plug) the site of injury. Patients with too few platelets or with platelets that are less sticky may bleed more than we would expect, and in combination with other factors such as medications or injury may cause the loss of a great deal of blood. By carrying out certain tests on platelets, it is possible to find out whether somebody has a problem with their platelets and whether this is likely to explain their bleeding problem.

The testing of platelet function

Platelet tests are routinely performed in the Haematology Clinic to investigate whether a patient has a platelet problem. In many cases, however, the tests do not give a clear answer on whether the patient has a platelet problem or on the cause of the problem. The present study will use a range of more specialist platelet tests to establish whether the patient has a platelet problem and to gain further information on the nature of the defect. This will require analysis of platelet function in freshly

donated blood. Work will be carried out by the Birmingham Platelet Group who have a large amount of experience in this field.

The search for defects in genes that could cause the platelet problem

The most likely explanation for a platelet problem is a defect (or problem) in one or more of the genes that you have been passed down (or inherited) from your mother and/or father. Genes can be described as the set of “instructions” that tell the body how to develop and work. A problem in one or more genes could cause the bleeding problem.

The results from the study of platelet function will have provided important clues on genes that may have caused the bleeding problem. The present study will look for errors in these genes using techniques that have been developed to identify gene defects. These studies require a small amount of blood that can be stored frozen for many years.

Is it possible that the gene defect will not be identified?

Yes. There is no guarantee that we will detect a change in either the platelet tests or in your gene. If we are unable to find a problem, we will tell you. If we fail to identify a problem, it does not mean you do not have a platelet or gene disorder – it is just that we have not been able to identify a problem with the techniques that we have used.

Are my family members likely to have the same gene defects and will they participate in the study?

Your mother, father and other family members may have a bleeding problem caused by a defect in the same gene or genes. Alternatively, your mother, father and other family members may have (or 'carry') one or more of the genes that is responsible for the bleeding problem, but may not have a bleeding problem.

If we are able to identify one or more gene defects that cause your platelet problem, we may wish to investigate, with their consent, other members of your family to see if they have a problem with their platelets and whether they have a defect in the same gene or genes. A separate information sheet is available for members of your family explaining these studies.

Will my blood be analysed for gene defects other than those which could cause the bleeding problem?

No. We are only interested in identifying defects that could cause your bleeding problem and we will target our search for abnormal genes to those involved in the way that platelets work under normal circumstances.

In patients with low platelet counts, we will need your consent to also look for mutations in the RUNX1 gene. Mutations in this gene have been shown to cause low

platelet numbers and may also predispose to several types of leukemia. You are free to decline your permission for this testing without compromising your participation in this study. This testing will not be carried out in the other participants in this study.

4. Why have I been chosen?

You have been chosen due to one of the following factors:

1. Your hospital doctor suspects that you have a problem with your platelets, but does not know the cause of the problem.

2. You have been identified as a suitable “control” or healthy volunteer.

5. Do I have to take part?

It is up to you to decide whether or not to take part. If you decide to take part you will be given this information sheet to keep and be asked to sign an agreement (or consent) form. By signing the consent form you are saying that you are happy to take part in the study and that you understand what has been explained to you.

If you choose not to take part, the clinical care you receive will not be affected in any way.

6. What will happen to me if I take part?

If you agree to take part you will be asked some questions about your medical history including questions to complete a bleeding questionnaire.

You will also be asked to give a sample of blood (between 10–50 ml, which is about the same volume as that of 2–10 teaspoonfuls). The blood will usually be taken at a local Haematology Clinic at a time that is suitable for your yourself and the Haematology Clinic, ideally early in the day so that the analyses can be carried out on the same day.

Depending on the results of the tests, we may ask you to give blood on up to three further occasions. This is to enable us to carry out different types of tests on your platelets and therefore find out more information on the problem. We will again ask for between 10–50 ml of blood, and this again can be taken either during a 'regular' visit to the Haematology Clinic / surgery or through a specific appointment. There will be a minimum of one month between the dates that the blood is taken and no upper time limit on the gap between the donations.

What will happen to my blood?

Your blood samples will be sent to research laboratories in Birmingham, Bristol and/or Sheffield for analysis of your platelets. We will also keep a frozen sample for a maximum of 10 years, which may be used to look for gene defects that could cause

the platelet problem or changes in blood proteins that may increase your chance of bleeding. The sample will be destroyed at this time, or earlier if you request.

We may wish to keep the sample for this length of time in the event that we are unable to find a problem with your genes in our first set of investigations. It is possible that, as we gain more knowledge, that we will wish to investigate the sample at later times to see if we can find a gene defect that may be the cause of the bleeding problem.

Please note that a frozen sample will not be stored for "control" individuals.

7. What are the possible disadvantages and risks of taking part?

You may experience a little pain and bruising around the vein where the blood sample is taken from. The needle used to take the blood will be identical to what has been used on you previously to take blood samples.

8. What are the possible benefits of taking part?

The study will generate important information on possible defects in your platelets and may also identify the genes that are responsible for the defects. The identification of a defect in your platelets may influence your clinical treatment.

If a gene defect is found, it will be possible to investigate whether one or more members of your family have the same gene defect, should they wish.

You will not receive any money for taking part in this study.

9. When will the study be completed?

Recruitment of patients will end in February 2015. Because we anticipate receiving many samples from throughout the UK, it may be many months before we can let you know if we have found a gene defect in your blood.

The nature of the gene defect will not be identified in all patients. However, developments may take place over the next few years that will provide important, new information on the likely cause of platelet bleeding disorders. We will therefore continue to follow developments in the field and, if appropriate, use this information to direct our efforts to identify the defective gene(s) in your frozen blood sample, which will be kept frozen for up to ten years, unless you ask for it to be destroyed. We will contact you as soon as we have identified a genetic defect in the sample.

10. Will my taking part in this study be kept confidential?

Yes. All information which is collected about you during the course of the research will be kept strictly confidential.

11. What will happen if I don't want to carry on with the study?

If you decide that you do not want to be in the study at any time after you have had your blood sample taken, you can ask for us to destroy your stored sample.

12. Further information on genetic testing

If we find a genetic defect related to bleeding, we will let you know of this. You will have the right to choose not to access this information. We may refer you for retesting by genetic services outside of this study for confirmation of the gene analysis.

We will discuss the full implications of the genetic studies with you with regard to your own children, having children in the future, insurance status and available counselling, as appropriate. Stored samples will not be used for other genetic studies without additional consent and presentation to an ethics committee for consideration.

13. Further tests that may be undertaken on your blood sample

In some circumstances, we will pursue more detailed testing of your blood samples. The types of detailed tests include, for example, using the stem cells from your blood sample to look at platelet production or examining the proteins in your platelets. Only tests for which permission has been formally granted to the Birmingham Platelet Group from the Research Ethics Committee (REC) will be performed. The purpose of the REC is to protect the rights, safety, dignity and well-being of research

participants, and to facilitate and promote ethical research that is of potential benefit to participants, science and society.

14. Contact for further information

Further information about this study can be obtained by contacting one of our group on telephone number 0121 415 8680. You will be able to leave a message if there is no-one in the office and we will get back to you. The members of the group are Mrs Gayle Halford (general enquires), Dr Marie Lordkipanidzé, Dr Gillian Lowe and Ms Rachael Taylor.

The study co-ordinator is a British Heart Foundation Professor, Stephen Watson (Centre for Cardiovascular Sciences, School of Clinical and Experimental Medicine, Institute of Biomedical Research, College of Medical and Dental Sciences, Wolfson Drive, University of Birmingham, Edgbaston, Birmingham, B15 2TT).

This completes Part 1 of the information sheet. If the information in Part 1 has interested you and you are considering participation, please continue to read the additional information in Part 2 before making any decisions.

PART 2

15. Who is funding this study?

The work is supported by the British Heart Foundation Research Charity.

16. Will I get paid for taking part?

You will not be paid for taking part.

17. Who has reviewed this study?

This study has been reviewed and approved by the West Midlands Multi-Centre Research Ethics Committee and by the Local Research Ethics Committee for your hospital.

18. What if something goes wrong?

If you are hurt by taking part in this research project, there are no special compensation arrangements.

If you are harmed due to someone's negligence, then you may have grounds for a legal action but you may have to pay for it. Regardless of this, if you wish to complain, or have any concerns about any aspect of the way you have been approached or treated during the course of this study, the normal National Health Service complaints mechanisms should be available to you. At present this would be via PALS (Patient Advice and Liaison Service) at your local hospital, or www.pals.nhs.uk

19. What will happen to the results of this study?

Your results will be passed on to your haematology doctor who will be asked to give them to you and to answer any questions you may have.

The collective results from all of the patients in the study may be presented in a scientific report to help other doctors who are treating patients with these bleeding disorders. No patient names will be used and it will not be possible to identify you from any report of the results of the study.

20. What if I wish to complain about the way in which the study has been conducted?

If you have any cause to complain about any aspect of the way in which you have been approached or treated during the course of the study, the normal National Health Service complaints mechanisms are available to you and are not compromised in any way because you have taken part in a research study. Complaints should be dealt with by an independent body such as PALS. To use the normal hospital complaints procedure, you should contact the PALS officer at your hospital or get more information at www.pals.nhs.uk

Thank you very much for reading through this information sheet.

UNIVERSITY OF BIRMINGHAM

Birmingham Platelet Group

INFORMED CONSENT FORM

7.2. PATIENT / FAMILY MEMBER

Thank you for reading the information about our research project. If you would like to take part, please read and sign this form.

Patient Identification Code for this study:

Title of Project: Mild bleeding disorders caused by platelet defects

Contact details for research team:

Your referring doctor / research nurse ……..

or

The study team: Mrs Gayle Halford (general enquiries), Dr Marie Lordkipanidzé and Dr Gillian Lowe

Birmingham Platelet Group, Institute for Biomedical Research, School of Clinical and Experimental Medicine, College of Medical and Dental Sciences, University of Birmingham, Birmingham B15 2TT, ☎ 0121 415 8680. Please leave a message if no-one is in the office and we will get back to you.

Please initial all boxes

1. I have read the attached information sheet on this project, and have been given a copy to keep. I have been able to ask questions about the project and I understand why the research is being done and any risks involved. ☐

2. I agree to give a sample of blood for research in this project. I understand how the sample will be collected, that giving a sample for this research is voluntary and that I am free to withdraw my approval for the use of the sample at any time without giving a reason and without my medical treatment or legal rights being affected. ☐

3. I give permission for someone from the local research team to look at my medical records to get information on my bleeding history. I understand that the information will be kept confident ☐

4. I agree to answering some questions about my medical history including those needed to complete a bleeding assessment by questionnaire. ☐

5. I understand that my referring doctor and I will be informed if any of the results of the medical tests done as a part of the research are important for my health. ☐

6. I understand that I will not benefit financially if this research leads to the development of a new treatment or medical test. ☐

7. I know how to contact the research team if I need to, and how to get information about the results of the research. ☐

8. Consent for storage of sample

I agree that the sample I have given and the information gathered about me can be stored in the Medical Schools of the Universities of Birmingham, Bristol and Sheffield for the purpose of analysing platelets, bleeding tendency and detection of the gene change responsible for the platelet disorder in myself or a family member, for a maximum of 10 years, after which it will be destroyed. ☐

If I wish I may request the return of the sample after analysis has been completed.

9. Consent for genetic research.

I give permission for genetic analysis to be carried out on the sample I give, as part of this project. I have received written information about this test and I understand what the result could mean to me and/or members of my family. ☐

10. Consent for stem cell studies

I give permission for the stem cells in my blood sample to be used to understand how my platelets are produced. ☐

YES

NO ☐

11. Consent for RUNX1 mutation testing **(ONLY FOR PARTICIPANTS WITH LOW PLATELET COUNTS).**

I give permission for genetic analysis to be carried out on the sample I give to specifically look at the RUNX1 gene. I understand that this mutation is associated with low platelet counts and can in some cases predispose to certain types of leukemia. ☐

YES

NO ☐

12. I want to be told the results of the tests undertaken in this study. I understand I can change my mind about this later. ☐

YES

NO ☐

…… …… ……

Name of patient Date Signature

(BLOCK CAPITALS)

…… …… ……

Name of person taking consent Date Signature

(if different from researcher)

…… …… ……

Name of researcher Date Signature

Thank you for agreeing to participate in this research

1 copy for patient, 1 for hospital notes, 1 for researcher

8. 11. Annexe 1: ISTH/SSC Bleeding Assessment Tool

The clinical appreciation of the presence and severity of bleeding symptoms is a fundamental step in the evaluation of patients referred for a possible bleeding disorder. In an attempt to improve the collection and reproducibility of the bleeding history, several Bleeding Assessment Tools (BAT) have been proposed and used. Currently available BAT have some limitations, particularly regarding the lack of pediatric-specific symptoms in some of them and the predominance of the severity of bleeding symptoms over other potentially clinically important features, such as the frequency of symptoms.

To overcome the above-mentioned limitations and to promote the standardization of the available BATs, a Working Group was established within the framework of the ISTH/SSC Subcommittees on VWF and on Perinatal/Pediatric Hemostasis (ISTH/SSC-BAT) during the 53rd SSC Annual Meeting held in Geneva in 2007. Members of the group first met in Toronto on January 2008 and then regularly at each subsequent SSC meeting. This paper presents a structured questionnaire and its clinical use agreed on by the ISTH/SSC-BAT together with a proposal for a new BS system to undergo validity and reliability testing in future studies.This new BAT is intended for inherited bleeding disorders in children and adults. The questionnaire should be collected by a physician or another adequately trained health-professional. Only symptoms and related treatments, if any, before and/or at diagnosis should be reported. Refer to the full text for additional instructions (Rodeghiero, Tosetto et al. 2010).

Minimal criteria defining a significant bleeding

For each specific bleeding symptom, the ISTH/SSC joint working group proposed minimal criteria in order to classify a symptom as significant and thus receive a score of 1 or more (see also Table 1):

1. Epistaxis: Any nosebleed, especially occuring after puberty, that causes patient concern (e.g., interference or distress with daily or social activities) is considered significant. In general, epistaxis should not be considered significant when it lasts less than 10 minutes, has a frequency of < 5 episodes/year, has a seasonal occurence, or is associated with infections of the upper respiratory tract or other identifiable cause (e.g., dusty dry air).

2. Cutaneous bleeding: Bruises are considered significant when 5 or more (> 1cm) in exposed areas; petechiae when adequately described by the patient or relatives; or hematomas when occurring without trauma.

3. Minor cutaneous wound: Any bleeding episode caused by superficial cuts (e.g., by shaving razor, knife, or scissors) or that requires frequent bandage changes is considered significant. Insignificant bleeding from wounds includes those of duration < 10 minutes and lesions that usually require stitches in normal subjects (e.g., under

the chin). Symptoms should also be manifest on more than one occasion to be considered significant.

4. Oral cavity bleeding: Gum bleeding should be considered significant when it causes frankly bloody sputum and lasts for 10 minutes or longer on more than one occasion. Tooth eruption or spontaneous tooth loss bleeding should be considered significant when it requires assistance or supervision by a physician, or lasts at least 10 minutes (bleeding associated with tooth extraction is considered separately). Bleeding occurring after bites to lips, cheek, and tongue should be considered significant when it lasts at least 10 minutes or causes a swollen tongue or mouth.

5. Hematemesis, melena, and hematochezia: Any gastrointestinal bleeding that is not explained by the presence of a specific disease should be considered significant.

6. Hematuria: Only macroscopic hematuria (from red to pale-pink urine) that is not explained by the presence of a specific urologic disease should be considered significant.

7. Tooth extraction: Any bleeding occurring after leaving the dentist's office and requiring a new, unscheduled visit or prolonged bleeding at the dentist's office causing a delay in the procedure or discharge should be considered significant.

8. Surgical bleeding: Any bleeding judged by the surgeon to be abnormally prolonged, that causes a delay in discharge, or requires some supportive treatment is considered significant.

9. Menorrhagia: Any bleeding that interferes with daily activities such as work, housework, exercise or social activities during most menstrual periods should be considered significant. Criteria for significant bleeding may include any of the following: changing pads more frequently than every 2 hours; menstrual bleeding lasting 7 or more days; and the presence of clots > 1 cm combined with a history of flooding. If a patient has previously made a record of her menstrual loss using a pictorial blood loss assessment chart (PBAC), a PBAC score higher than 100 also qualifies for a score of 1.

10. Post-partum bleeding. Vaginal bleeding or uterine discharge (lochia) that lasts for more than 6 weeks. Any bleeding of lesser duration that is judged by the obstetrician as abnormally heavy or prolonged, that causes a delay in discharge, requires some supportive treatment, requires changing pads or tampons more frequently than every 2 hours, or causes progressive anemia is also considered significant.

11. Muscle hematomas or hemarthrosis. Any spontaneous joint / muscle bleeding (not related to traumatic injuries) is considered significant.

12. CNS bleeding. Any subdural or intracerebral hemorrhage requiring diagnostic or therapeutic intervention is scored 3 or 4, respectively.

13. Other bleeding symptoms. When these bleeding symptoms occur during infancy, they are scored 1 or more. Their presence when reported by either the patient or a family member should always prompt detailed laboratory investigation.

Only symptoms and treatment BEFORE and AT diagnosis should be considered

1.	**Epistaxis**

1.1	Have you ever had spontaneous epistaxis?	□ Yes	□ No or trivial (skip to 2)
1.2	Have the symptom ever required medical attention ?	□ Yes	□ No (resolve spontaneously; skip to 1.6)
1.3	If answer to 1.2 is yes, please specify	□ Consultation only □ Cauterization/ Packing □ Treatment with desmopressin / antifibrinolytics/ iron therapy □ Treatment with plasma, platelet or factor concentrates □ Blood (RBC) transfusion	
1.4	How many times in your life did you receive any of the above treatments (# 1.3)?	□ 1 - 2 □ 3 to 5 □ 6 to 10 □ more than 10	
1.5	At what age did you first have symptoms?	□ Before 1 year □ Between 1-5 years of age □ Between 6-12 years of age □ Between 13-25 years of age □ After 25 years of age	
1.6	Approximate number of episodes NOT requiring medical attention	□ less than 1 per year □ 1 per year □ 1-5 every six month □ 1-3 every month □ 1 every week	
1.7	Duration of average single episode (min.) NOT requiring medical attention	□ 1 minute or less □ 1 - 10 minutes □ more than 10 minutes	

2.	Cutaneous bleeding (Bruising, ecchymoses, purpura, subcutanueos hematomas)

2.1	Have you ever had any of the above cutaneous bleeding?	□ Yes	□ No or trivial skip to 3
2.2	Have the symptom ever required medical attention?	□ Yes	No □ skip to 2.6
2.3	If answer to 2.2 is yes, please specify	□ Consultation only □ Treatment with desmopressin □ Treatment with plasma, platelets or factor concentrates □ Blood (RBC) transfusion	
2.4	How many times in your life did you receive any of the above treatments (# 2.3)?	□ 1 - 2 □ 3 to 5 □ 6 to 10 □ more than 10	
2.5	At what age did you first have symptoms?	□ Before 1 year □ Between 1-5 years of age □ Between 6-12 years of age □ Between 13-25 years of age □ After 25 years of age	
2.6	Approximate number of episodes NOT requiring medical attention	□ less than 1 per year □ 1 per year □ 1-5 every six month □ 1-3 every month □ 1 every week	
2.7	Type of bleeding	□ Petechiae □ Bruises □ Hematomas	
2.8	Location	□ Exposed sites □ Unexposed sites □ Both	
2.9	Common size	□ ≤ 1 cm □ >1 cm □ Extensive (palm sized or larger)	
2.10	How many bruises >1 cm in exposed areas in the most severe manifestation?	□ ≤ 5 □ > 5	
2.11	Location of petechiae	□ Limited to lower limbs □ Diffuse	

3.	Bleeding from minor wounds (not requiring stitches in the average patient)

3.1	Have you ever had prolonged bleeding from minor wounds?	□ Yes	□ No or trivial skip to 4
3.2	Have the symptom ever required medical attention ?	□ Yes	□ No skip to 3.6

3.3	If answer to 3.2 is yes, please specify	□ Consultation only □ Surgical hemostasis □ Treatment with desmopressin □ Treatment with plasma, platelet or factor concentrates □ Blood (RBC) transfusion
3.4	How many times in your life did you received any of the above treatments (# 3.3)?	□ 1 - 2 □ 3 to 5 □ 6 to 10 □ more than 10
3.5	At what age did you first have symptoms?	□ Before 1 year □ Between 1-5 years of age □ Between 6-12 years of age □ Between 13-25 years of age □ After 25 years of age
3.6	Approximate number of episodes NOT requiring medical attention	□ less than 1 per year □ 1 per year □ 1-5 every six month □ 1-3 every month □ 1 every week
3.7	Duration of average single episode (min.)	□ 1 to 10 minutes □ more than 10 minutes

4.	**Hematuria**

4.1	Have you ever had hematuria ?	□ Yes	□ No skip to 5
4.2	If answer to 4.1 is yes, please specify		
	Presence of associated urologic disease	Yes □ (skip to 5) Specify: □ Stones □ Infection □ Kidney/ bladder disease	No □

Please answer the following questions only for SPONTANEOUS symptoms (answer No to 4.1)

4.3	Have the symptom ever required medical attention ?	Yes □	No □ skip to 4.7
4.4	If answer to 4.3 is yes, please specify	□ Consultation only □ Surgery □ Treatment with desmopressin □ Treatment with plasma, platelet or factor concentrates □ Blood (RBC) transfusion	
4.5	How many times in your life did you received any of the above treatments (# 4.4)?	□ 1 - 2 □ 3 to 5 □ 6 to 10 □ more than 10	
4.6	At what age did you first have symptoms?	□ Before 1 year □ Between 1-5 years of age □ Between 6-12 years of age □ Between 13-25 years of age □ After 25 years of age	
4.7	Approximate number of episodes NOT requiring medical attention	□ less than 1 per year □ 1 per year □ 1-5 every six month □ 1-3 every month □ 1 every week	

5. Gastrointestinal bleeding (Hematemesis, Melena, Hematochezia)

5.1 Have you ever had gastrointestinal bleeding ? □ Yes □ No skip to 6

5.2 If answer to 5.1 is yes, please specify

Type of bleeding
□ Hematemesis
□ Melena
□ Hematochezia

Presence of associated GI disease
Yes □ No □
Specify:
□ Ulcer
□ Portal hypertension
□ Angiodysplasia

Please answer to the following questions only for SPONTANEOUS symptoms

5.3 Have the symptom ever required medical attention ? Yes □ No □ skip to 5.7

5.4 If answer to 5.3 is yes, please specify
□ Consultation only
□ Surgical haemostasis
□ Treatment with desmopressin
□ Treatment with plasma, platelet or factor concentrates
□ Blood (RBC) transfusion

5.5 How many times in your life did you received any of the above treatments (# 5.4)?
□ 1 - 2
□ 3 to 5
□ 6 to 10
□ more than 10

5.6 At what age did you first have symptoms?
□ Before 1 year
□ Between 1-5 years of age
□ Between 6-12 years of age
□ Between 13-25 years of age
□ After 25 years of age

5.7 Approximate number of episodes NOT requiring medical attention
□ less than 1 per year
□ 1 per year
□ 1-5 every six month
□ 1-3 every month
□ 1 every week

6. Oral cavity bleeding (Tooth eruption, spontaneous or after brushing/flossing, gum bleeding, bleeding after bites to lip & tongue)

6.1	Have you ever had oral cavity bleeding ?	□ Yes	□ No or trivial skip to 7
6.2	Have the symptom ever required medical attention ?	Yes □	No □ skip to 6.6
6.3	If answer to 6.2 is yes, please specify	□ Consultation only	
		□ Surgery (dental packing, suture, cauterization)	
		□ Treatment with desmopressin / iron therapy	
		□ Treatment with plasma, platelet or factor concentrates	
		□ Blood (RBC) transfusion	
6.4	How many times in your life did you received any of the above treatments (# 6.3)?	□ 1 - 2 □ 3 to 5 □ 6 to 10 □ more than 10	
6.5	At what age did you first have symptoms?	□ Before 1 year □ Between 1-5 years of age □ Between 6-12 years of age □ Between 13-25 years of age □ After 25 years of age	
6.6	Approximate number of episodes NOT requiring medical attention	□ less than 1 per year □ 1 per year □ 1-5 every six month □ 1-3 every month □ 1 every week	
6.7	Duration of average single episode (min.)	□ 1 to 10 minutes □ more than 10 minutes	

7. Bleeding after Tooth/ Teeth extraction

7.1 Have you ever had bleeding after tooth (teeth) extraction ? □ Yes □ No skip to 8

7.2 If answer to 7.1 is yes, please specify

Number of extractions □□

*Please fill in one of the following forms for **each** tooth extraction*

Age at extraction	□□	Type of extraction	□ Deciduous □ Permanent □ Molar
Actions taken to prevent bleeding		□ None □ Antifibrinolytics □ Desmopressin □ Plasma or clotting factor concentrates □ Platelet infusion	
Bleeding after extraction?		Yes □	No □
Actions taken to control bleeding		□ Resuturing □ Packing □ Antifibrinolytics □ Desmopressin □ Plasma or clotting factor concentrates □ Platelet infusion □ Blood (RBC) transfusion	

8. Bleeding after Surgery or Major Trauma

8.1 Have you ever had bleeding after surgery or major trauma ? □ Yes □ No, skip to 9

8.2 If answer to 8.1 is yes, please specify

Number of interventions □□

*Please fill in one of the following forms for **each** surgery or major trauma episode*

Age at intervention/trauma □□

Type of surgery

□ Tonsillectomy/Adenoids
□ Pharynx/Nose
□ Major-abdominal
□ Major-thoracic
□ Major-gynecology
□ Other

Actions taken to prevent bleeding

□ None
□ Antifibrinolytics
□ Desmopressin
□ Plasma or clotting factor concentrates
□ Platelet infusion

Bleeding after intervention? Yes □ No □

Actions taken to control bleeding

□ None
□ Resuturing
□ Packing
□ Antifibrinolytics
□ Desmopressin
□ Plasma or clotting factor concentrates
□ Platelet infusion
□ Blood (RBC) transfusion

9. Menorrhagia

9.1	Have you ever had very heavy menstrual bleeding (menorrhagia)?	□ Yes	□ No or trivial skip to 10
	If answer to 9.1 is yes, please specify	□ Changing pads/tampons more frequently than every 2 hours □ Bleeding more than 7 days □ Clot and flooding	
		Impairment of daily activities (work, housework, exercise, social activities):	□ Never or rarely □ Most menses
9.2	Have the symptom ever required medical attention ?	□ Yes	□ No skip to 9.6
9.3	If answer to 9.2 is yes, please specify	a □ Consultation only	
		b □ Pictorial Bleeding Assessment	Score _____
		c □ Antifibrinolytic therapy	
		d □ Iron therapy	
		e □ Hormonal therapy	
		f □ Combined antifibrinolytics & Hormonal therapy	
		g □ Hysterectomy / endometrial ablation / D & C	
		h □ Treatment with desmopressin, plasma or factor concentrates, platelet transfusion	
		i □ Blood (RBC) transfusion	
		l □ Hospital admission and emergency treatment	
9.4	How many times in your life did you received any of the above treatments (# 9.3 a-l)?	□ 1 - 2 □ 3 to 5 □ 6 to 10 □ more than 10	
9.5	At what age did you first have symptoms?	□ At menarche □ Between 14-25 years of age □ After 25 years of age	
9.6	Have you had time off work/school for menorrhagia?	□ < twice a year □ > twice a year	
9.7	Duration of menorrhagia	□ Since menarche □ > 12 months □ < 12 months	
9.8	Have you had acute menorrhagia requiring emergency treatment/hospital admission	□ Yes How many times: ____	□ No

10.	**Post-partum hemorrhage**

10.1	Number of successful pregnancies (live births)	□□
10.2	Have you ever had post-partum haemorrhage?	□ Yes □ No or trivial skip to 11
10.3	Did it occur	□ In the first 24 hours after delivery (Primary) □ Between 24 hours and 6 weeks postpartum (Secondary) □ Both Primary and Secondary
10.4	How long did vaginal discharge (lochia) last?	□ < 6 weeks □ > 6 weeks
10.5	Did it require changing pads/tampons more frequently than every 2 hours?	□ Yes □ No
10.6	Did this bleeding cause delay of hospital discharge/ readmission to hospital?	□ Yes □ No
10.7	Have the symptom ever required medical treatment?	□ Yes □ No
10.8	If answer to 10.7 is yes, please specify	□ Consultation only /oxytocin i.v. infusion □ Additional uterotonic medications □ Iron therapy □ Antifibrinolytic therapy □ Treatment with plasma or factor concentrates, platelet transfusion □ Blood (RBC) transfusion □ Any procedure requiring examination under anaesthesia □ Uterine balloon/package to tamponade the uterus □ Any procedure requiring critical care or surgical intervention (includes: hysterectomy, internal iliac artery legation, uterine artery embolization, uterine brace sutures)
10.9	Number of deliveries that required any of the above treatments (# 10.8)?	□□

11.	**Muscle hematomas or hemarthrosis (spontaneous)**

11.1	Have you ever had muscle hematomas or hemarthrosis ?	□ Yes	□ No or trivial skip to 12
11.2	Have the symptom ever required medical attention ?	□ Yes	□ No skip to 11.6
11.3	If answer to 11.2 is yes, please specify	□ Consultation only □ Surgical draining □ Treatment with desmopressin □ Treatment with plasma, platelet or factor concentrates □ Blood transfusion	
11.4	How many times in your life did you receive any of the above treatments (# 11.3)?	□ 1 - 2 □ 3 to 5 □ 6 to 10 □ more than 10	
11.5	At what age did you first have symptoms?	□ Before 1 year □ Between 1-5 years of age □ Between 6-12 years of age □ Between 13-25 years of age □ After 25 years of age	
11.6	Approximate number of episodes NOT requiring medical attention	□ less than 1 per year □ 1 per year □ 1-5 every six month □ 1-3 every month □ 1 every week	

12	Other bleedings

12.1 Have you ever had one of the following?

Excessive umbilical stump bleeding	□ Yes	□ No
Cephalohematoma	□ Yes	□ No
Bleeding at circumcision	□ Yes	□ No
Venipuncture bleeding	□ Yes	□ No
Suction Bleeding	□ Yes	□ No
Ovulation bleeding(in women)	□ Yes	□ No

12.2 Have one of these symptoms ever required medical attention? □ Yes □ No

12.3 If answer to 12.2 is yes, please specify

- □ Consultation only
- □ Surgery
- □ Treatment with desmopressin
- □ Treatment with plasma, platelet or factor concentrates
- □ Blood (RBC) transfusion

12.4 How many times in your life did you receive any of the above treatments (# 12.3) for this symptom?

- □ 1 - 2
- □ 3 to 5
- □ 6 to 10
- □ more than 10

Table 1. Bleeding score

SYMPTOMS (up to the time of diagnosis)			**8.1.1.1.1.** SCORE		
8.1.1.2.	0[§]	1[§]	2	3	4
EPISTAXIS	No/trivial	- > 5/year or - more than 10 minutes	**CONSULTATION ONLY***	Packing or cauterization or antifibrinolytic	Blood transfusion or replacement therapy (use of hemostatic blood components and rFVIIa) or desmopressin
Cutaneous	No/trivial	For bruises 5 or more (> 1cm) in exposed areas	Consultation only*	Extensive	Spontaneous hematoma requiring blood transfusion

Bleeding from minor wounds	No/trivial	- > 5/year or - more than 10 minutes	**CONSULTATION ONLY***	Surgical hemostasis	Blood transfusion, replacement therapy, or desmopressin
Oral cavity	No/trivial	Present	**CONSULTATION ONLY***	Surgical hemostasis or antifibrinolytic	Blood transfusion, replacement therapy or desmopressin
GI bleeding	No/trivial	Present (not associated with ulcer, portal hypertension, hemorrhoids, angiodysplasia)	Consultation only*	Surgical hemostasis, antifibrinolytic	Blood transfusion, replacement therapy or desmopressin

Hematuria	No/trivial	Present (macroscopic)	**CONSULTATION ONLY***	Surgical hemostasis, iron therapy	Blood transfusion, replacement therapy or desmopressin
Tooth extraction	No/trivial or none done	Reported in ≤25% of all procedures, no intervention**	**REPORTED IN >25% OF ALL PROCEDURES, NO INTERVENTION***	Resuturing or packing	Blood transfusion, replacement therapy or desmopressin
Surgery	No/trivial or none done	Reported in ≤25% of all procedures, no intervention**	**REPORTED IN >25% OF ALL PROCEDURES, NO INTERVENTION***	Surgical hemostasis or antifibrinolytic	Blood transfusion, replacement therapy or desmopressin

Menorrhagia	No/trivial	Consultation only* or - Changing pads more frequently than every 2 hours or - Clot and flooding or - PBAC score>100[#]	**- TIME OFF WORK/SCHOOL > 2/YEAR OR - REQUIRING ANTIFIBRINOLYTICS OR HORMONAL OR IRON THERAPY**	- Requiring combined treatment with antifibrinolytics and hormonal therapy or - Present since menarche and > 12 months	- Acute menorrhagia requiring hospital admission and emergency treatment or - Requiring blood transfusion, Replacement therapy, Desmopressin, or - Requiring dilatation & curretage or endometrial ablation or hysterectomy)
Post-partum hemorrhage	No/trivial or no deliveries	Consultation only* or - Use of syntocin or - Lochia > 6 weeks	- Iron therapy or - Antifibrinolytics	- Requiring blood transfusion, replacement therapy, desmopressin or - Requiring examination	- Any procedure requiring critical care or surgical intervention (e.g. hysterectomy, internal iliac artery legation, uterine artery embolization, uterine brace

				under anaesthesia and/or the use of uterin balloon/package to tamponade the uterus	sutures)
Muscle hematomas	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
Hemarthrosis	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
CNS bleeding	Never	-	**-**	Subdural, any intervention	Intracerebral, any intervention

Other bleedings^	No/trivial	Present	**CONSULTATION ONLY***	Surgical hemostasis, antifibrinolytics	Blood transfusion or replacement therapy or desmopressin

In addition to the guidance offered by the table, it is mandatory to refer to the text for more detailed instructions.

[§] Distinction between 0 and 1 is of critical importance. Score 1 means that the symptom is judged as present in the patient's history by the interviewer but does not qualify for a score 2 or more

* Consultation only: the patient sought medical evaluation and was either referred to a specialist or offered detailed laboratory investigation

** Example: 1 extraction/surgery resulting in bleeding (100%): the score to be assigned is 2; 2 extractions/surgeries, 1 resulting in bleeding (50%): the score to be assigned is 2; 3 extractions/surgeries, 1 resulting in bleeding (33%): the score to be assigned is 2; 4 extractions/surgeries, 1 resulting in bleeding (25%): the score to be assigned is 1

[#] If already available at the time of collection

^ Include: umbilical stump bleeding, cephalohematoma, cheek hematoma caused by sucking during breast/bottle feeding, conjunctival hemorrhage or excessive bleeding following circumcision or venipuncture. Their presence in infancy requires detailed investigation independently from the overall score