Article:
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. ISSN: 0390-6078

https://doi.org/10.3324/haematol.2016.146316

Whole exome sequencing identifies genetic variants in inherited thrombocytopenia with secondary qualitative function defects

Ben Johnson,[1] Gillian C Lowe,[1] Jane Futterer,[1] Marie Lordkipanidzé,[1] David MacDonald,[1] Michael A Simpson,[2] Isabel Sánchez Guiú,[3] Sian Drake,[1] Danai Bem,[1] Vincenzo Leo,[4] Sarah J Fletcher,[1] Ban Dawood,[1] José Rivera,[3] David Allsup,[5] Tina Biss,[6] Paula HB Bolton-Maggs,[7] Peter Collins,[8] Nicola Curry,[9] Charlotte Grimley,[10] Beki James,[11] Mike Makris,[4] Jayashree Motwani,[12] Sue Pavord,[13] Katherine Talks,[6] Jecko Thachil,[7] Jonathan Wilde,[14] Mike Williams,[12] Paul Harrison,[15] Paul Gissen,[16] Stuart Mundell,[17] Andrew Mumford,[17] Martina E. Daly,[4] Steve P Watson,[1] Neil V Morgan[1] on behalf of the UK GAPP Study Group

[1]Institute for Cardiovascular Sciences, College of Medical and Dental Sciences, University of Birmingham, UK

[2]Division of Genetics and Molecular Medicine, King's College, London, UK

[3]Centro Regional de Hemodonación, Universidad de Murcia, IMIB-Arrixaca, Murcia, Spain

[3]Division of Genetics and Molecular Medicine, King's College, London, UK

[4]Department of Infection, Immunity and Cardiovascular Disease, University of Sheffield Medical School, University of Sheffield, UK

[5]Hull Haemophilia Treatment Centre, Hull and East Yorkshire Hospitals NHS trust, Castle Hill Hospital, Hull

[6]Department of Haematology, Royal Victoria Infirmary, Newcastle Upon Tyne, UK

[7]Department of Haematology, Manchester Royal Infirmary, Manchester, UK

[8]Arthur Bloom Haemophilia Centre, School of Medicine, Cardiff University, Cardiff, UK

[9]Oxford Haemophilia & Thrombosis Centre, Churchill Hospital, Oxford, UK

[10]Nottingham Haemophilia Centre, Nottingham University Hospital, Nottingham, UK

[11]Regional Centre for Paediatric Haematology, Leeds Children's Hospital, Leeds, UK

[12]Department of Haematology, Birmingham Children's Hospital, Birmingham, UK

[13]HaemostasisThrombosis Unit, Leicester Royal Infirmary, Leicester, UK

[14]Adult Haemophilia Centre, Queen Elizabeth Hospital, Birmingham, UK

[15]School of Immunity and Infection, College of Medical and Dental Sciences, University of Birmingham, UK

[16]Medical Research Council, Laboratory for Molecular Cell Biology, University College London, London, UK

[17]School of Physiology, Pharmacology and Neuroscience, University of Bristol, Bristol, UK

Address correspondence to:

Dr Neil V. Morgan, Institute for Cardiovascular Sciences, University of Birmingham, Institute of Biomedical Research, Edgbaston, Birmingham, B15 2TT, United Kingdom.

Tel: (+44) 121 414 6820 Fax: (+44) 121 415 8817

e-mail: n.v.morgan@bham.ac.uk

Counts

Text: 3,897 words; Abstract: 219 words

Figures: 2 Tables: 3 (one additional supplementary)

References: 37

Key points

1. “Pathogenic” or “likely pathogenic” classified variants in known genes were discovered in 46% of index cases with an inherited thrombocytopenia of unknown aetiology.

2. Whole exome sequencing combined with platelet phenotyping is a valuable research tool for discovering potentially pathogenic variants in known and novel genes for further research.

Abstract

Inherited thrombocytopenias are a heterogeneous group of disorders characterised by abnormally low platelet counts which can be associated with abnormal bleeding. Next generation sequencing has previously been employed in these disorders for the confirmation of suspected genetic abnormalities, and more recently in the discovery of novel disease causing genes. However its full potential has not previously been utilised. Over the past 6 years we have sequenced the exomes from 55 patients, including 37 index cases and 18 additional family members, all of whom were recruited to the UK Genotyping and Phenotyping of Platelets study. All patients had inherited or sustained thrombocytopenia of unknown aetiology with platelet counts varying from 11-186x10^9/L. Of the 51 patients phenotypically tested, 37 (73%), had an additional secondary qualitative platelet defect. Using whole exome sequencing analysis we have identified “pathogenic” or “likely pathogenic” variants in 46% (17/37) of our index patients with thrombocytopenia. In addition, we report variants of uncertain significance in 12 index cases which include novel candidate genetic variants in previously unreported genes in four index cases. These results demonstrate that whole exome sequencing is an efficient method for elucidating potential pathogenic genetic variants in inherited thrombocytopenia. Whole exome sequencing also has the added benefit of discovering potentially pathogenic genetic variants for further study in novel genes not previously implicated in inherited thrombocytopenia.

Introduction

Inherited thrombocytopenias (IT) are a heterogeneous group of disorders characterised by platelet counts of less than $150x10^9$/L in whole blood. Platelet counts are considered normal when maintained at levels between $150\text{-}450x10^9$/L. This is achieved by homeostatic processes controlling platelet production (thrombopoiesis), platelet senescence and platelet consumption/destruction. Pathogenic mutations can result in a disruption of these balanced processes causing inherited thrombocytopenia. However, the clinical manifestation of bleeding is often dependent on both a decreased platelet count and a qualitative or acquired platelet defect. Clinical complications can vary dramatically from severe and potentially life threatening bleeding to being asymptomatic. This variation is noted amongst individuals shown to have the same underlying genetic causes of disease, suggesting that bleeding risk and phenotype is a complex trait (1).

The average incidence of IT is estimated to be approximately 270 cases per 1 million live births (2). To date there are 27 individual IT disorders with known causative mutations registered within OMIM, although 33 disease causing genes have been described (3).

Genetic studies have played a major role in the diagnosis and progressive understanding of IT. The genes implicated in the disease encode proteins that vary widely in function and include transcription factors (*ETV6, FLI1, GATA1, GFI1B* and *RUNX1*) and proteins involved in cytoskeleton rearrangement and organisation (*ACTN1, FLNA, GP1BA, GP1BB, GP9, TUBB1* and *WAS)*. However, some protein functions currently remain unknown (*SLFN14* and *GNE*) (4-9). Although our

knowledge of the causes of IT continues to grow, presently a genetic diagnosis is only reported in approximately 50% of individuals (10-12).

Previously, genetic investigation into IT has focused on candidate gene sequencing and individual cases of whole exome sequencing (WES) when a causative gene is not obvious (9). With 50% of patients currently undiagnosed, a change in the way we approach genetic diagnosis is necessary. Here we present the first large scale WES only approach to patients with suspected IT. We demonstrate its application in determining possible genetic origins of IT including identification of variants in novel candidate causative genes. We combine this with an approach implemented by the Genotyping and Phenotyping of Platelets (GAPP) study, which combines WES analysis with extensive platelet phenotyping to create a complete method of diagnosis and gene discovery in this subset of patients.

Methods

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 GAPP study is included in the National Institute of Health Research Non-Malignant Haematology study portfolio (ID9858).

Platelet counts, morphology and white blood cell counts

Patient samples were compared to the range of healthy volunteers for the specific method of morphology used. Platelet counts for light transmission aggregometry (LTA) and flow cytometry analysis as well as mean platelet volume (MPV) in platelet rich plasma (PRP) were originally measured using the Beckman coulter counter (n=44). Subsequently, platelet counts, morphology and white blood cell counts in whole blood were measured using the Sysmex XN-1000 (n=11). The PLT-F channel was used to determine platelet counts in whole blood and the immature platelet fraction (IPF). MPV was determined from the impedance PLT-I channel. White blood cell counts were determined using the Sysmex XN-DIFF channel. All samples were tested against a normal range which was established by measuring the counts for 40 healthy individuals using the Sysmex XN-1000.

Platelet preparation and platelet function testing

Platelet function was assessed by light transmission aggregometry, including lumiaggregometry, for samples having platelet counts in PRP of $>1x10^8$/mL (n=13).

An in-house flow-cytometry assay was developed to assess platelet function in patients having platelet counts in PRP <1x10^8/mL (n=22). Platelets from individuals with borderline platelet counts in PRP between 1.0 and 1.5x10^8/mL were assessed using both assays (n=16).

Aggregometry was performed as previously described (13, 14). For flow cytometry, resting surface levels of CD42b, CD41 and GPVI were assessed. PRP was then stimulated with ADP (3 and 30 μM), CRP (0.3 and 3μg/ml) and PAR-1 peptide (10 and 100 μM). Membrane expression of P-selectin (FITC-conjugated mouse anti-human CD62P antibody, BD Pharmingen), a marker of platelet alpha granule release, as well as fluorescent fibrinogen binding (marker of integrin activation) was assessed by flow cytometry on an Accuri C6 flow cytometer. Incubation took place at 37ºC for 2 min and was terminated by adding a fivefold excess of ice cold PBS.

Whole exome sequencing

WES and bioinformatics analysis was performed as described previously (8, 15, 16) (Figure 1).

Pathogenicity of variants was determined and called using the consensus guidelines as set out by the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG guidelines) (17). Segregation was determined by Sanger sequencing of candidate variants in both affected and unaffected family members, when available, and classification was adapted appropriately for the specific study and small sample size.

Sanger sequencing

To verify candidate mutations and examine their segregation among family members Sanger sequencing was performed using standard methods on an ABI 3730 automated sequencer as described previously (8).

Results

Patient recruitment

To date, 55 patients with a suspected IT or sustained reduced platelet counts have been enrolled from 25 UK Haemophilia Care Centres and investigated as part of the GAPP study. Before enrolment in the study, all patients underwent clinical and genetic work-up to exclude known platelet disorders (including Bernard-Soulier Syndrome and *MYH9*-related disorders, analysed initially by blood film), idiopathic thrombocytopenic purpura (ITP) and other non-platelet disorders including von Willebrand disease and inherited coagulation factor deficiencies. Patient bleeding phenotypes are displayed in Table 1. WES was performed on genomic DNA from all patients, including 37 index cases, all of whom met the study's entry criteria. All patients, excluding F35.I and F35.II, were of white British or mixed British ethnicity. All results following platelet function testing and WES were reported back to the referring haematological consultants to aid in genetic counselling and disease management.

Platelet counts, morphology and function testing

Patients were recruited with a platelet count in whole blood, at the time of enrolment, of less than $150x10^9$/L. Patients having platelet counts in the range of $150\text{-}200x10^9$/L remained enrolled in the study if they showed a similar phenotype to related affected family members and a platelet count below $150x10^9$/L had been observed prior to enrolment (patients F4.II, F11.III, F13.I and F30.II). Platelet counts, MPVs and IPFs are displayed in Table 1. Of the 55 recruited patients, 12 were deemed to have a macrothrombocytopenia and three a microthrombocytopenia (Table 1). White cell

counts were within the normal range (normal range to 2SD 3.78 - 10.11x10^9/L, n=40) across all patients (n=11) analysed.

Platelet function studies revealed the presence of a secondary qualitative defect in addition to the reduction in platelet count in 37/51 (73%) of the 55 patients whose DNA underwent WES and who were also available for platelet function testing (PFT) (Table 1). Of the 37 patients with a secondary qualitative defect, 89% (33/37) displayed defects in both alpha and dense granule secretion. Five of these patients with an observed granule secretion defect were also suspected to have an additional Gi defect due to a reduction in response to all concentrations of ADP. The remaining four patients without an observable granule secretion defect showed abnormalities in alternative pathways (Integrin activation, cyclooxygenase pathway and GPVI surface levels) in addition to the reduction in platelet count (Table 1).

Whole exome sequencing

WES was performed on genomic DNA from all 55 patients, comprising of 37 index cases, following PFT. Average fold coverage of 111 was observed across all DNA samples analysed by WES with an average of 91% of target sequences having >20x coverage. Areas of poor coverage were analysed manually when occurring in previously IT associated genes.

WES revealed between 24,000 and 25,000 variants (SNVs, small scale insertions/deletions, and splice site variations) in the DNA from each patient, with an average of 197 novel variants per exome. On average, per individual, 2401 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. Percentage of the gene with

≤20x coverage for panel of 358 platelet-related genes is included within Supplementary Table 1.

CNVs were detected using ExomeDepth (18) and analysis revealed an average of 137 CNVs per exome (n=32). No CNVs were deemed potential candidates either due to a high allele frequency or a lack of expression or functional role of the gene within the megakaryocyte/platelet lineage.

Variants in known thrombocytopenia causing genes

WES and downstream analysis identified variants in 25 index cases (68%) within the 33 known IT causing genes. All variants exceeded 30x sequence coverage at the point of variation and have been confirmed by Sanger sequencing. Variants were selected from positive hits to genes within the panel of 358 IT associated genes (Supplementary Table 1). On average, 37 variants per individual (range 11-52) were noted in genes from the panel of 358 IT associated genes, of which on average four (range 0-7) variants were significant per exome analysed.

In total 28 variants were noted in 14 genes previously known to cause IT (Table 2). Twenty-one index cases possessed a single variant in a gene previously known to cause IT. Four index cases possessed two variants in genes previously known to cause IT. One variant, *RUNX1*; c.270+1G>T, was noted in two index cases (F13.I and F14.I). Candidate variations were present within; *ACTN1*, the 5'-UTR of *ANKRD26*, *CYCS*, *FLI1, GFI1B*, *ITGB3*, *GP1BA* (heterozygous), *MYH9*, *NBEAL2*, *RUNX1, SLFN14, STIM1*, *TPM4* and *TUBB1*. All but six variants were novel and not present within the variant databases previously mentioned. Three variants, *ANKRD26;* c.-126T>G in F2.I, *MYH9;* c.3493C>T (rs80338829) in F9.I and RUNX1; c.530G>A in F19.I and F19.II have been previously associated with IT (1, 19, 20).

The remaining three variants that have been previously observed occurred at frequencies in available databases < 0.005 (0.05%). One of the databases scrutinised was that of the ExAC consortium (http://exac.broadinstitute.org) which may include data from individuals with low platelet counts who were either undiagnosed or recruited through an unrelated study (Table 2). Seven variants have previously been published as part of two separate publications from the UK-GAPP study group (8, 15).

Classification of the 28 variants occurring within the known IT-related genes, following the interpretation guidelines as set out by Richards *et al.* (17), revealed four variants to be “pathogenic”, 13 to be “likely pathogenic” and 11 to be of “uncertain significance”. Variants classified as “pathogenic” were either already known to be a genetic cause of IT; *ANKRD26;* c.-126T>G in F2.I and *MYH9;* p.Arg1165Cys in F9.I, or were predicted loss of function variants in genes where a loss of function is known to cause disease; *FLI1;* p.Asn331Thr fs*4, in F4.I and F4.II and *RUNX1;* pTrp79* in P12.I.

On average, less than one novel variant was expected to be observed in the known IT causing genes in which variants were observed. The number of variants occurring also exceeds the expected number when extending the analysis to cover variants with a MAF of <0.01.

Of the 37 index patients, four presented with two candidate variations in known disease linked genes, which in one case were present in the same gene. These were as follows: F6.I (*GFI1B*; c.676+1G>A and *STIM1*; p.Ala610Thr), F10.I (*NBEAL2*; (p.Leu459Arg fs*13 and p.Asn2298Ser), F12.I (*RUNX1*; p.Trp79* and *ITGB3* p.Arg117Trp) and F23.I (*TPM4;* p.Ala183Val and *TUBB1*; p.Phe242Leu).

Of the 25 index cases having variants in known disease causing genes, nine were observed to have variants within the RUNT1-related transcription factor gene; *RUNX1*. One variant, *RUNX1*; p.Arg177Gln, observed in F19.I and F19.II has been previously reported as causative germline mutation of familial platelet disorder in two individuals from the same pedigree (20). The variations consisted of five missense variants, two splice-site variants and one nonsense variant. One splice-site variation, c.270+1G>T, was present within three affected individuals from two separate families (F13 and F14). All variants, with the exception of a missense substitution (p.D6N), lie within the genetic region encoding the RUNT homology domain (RHD) which mediates DNA binding and heterodimerization with CBFβ (Figure 2) (21). Platelets from the majority of these patients (10/13) demonstrated a reduction in ATP secretion and, in keeping with previous reports, several of these patients displayed additional clinical features. Variations in *RUNX1* are associated with a propensity to myelodysplastic syndrome and acute myeloid leukaemia (AML). To date, haematological malignancies have not been reported in any patients; however, the brother of F16.1 did have a history of AML but was unavailable for testing.

Potentially damaging variants in novel candidate genes

After scrutinising individuals for variants within the panel of 358 platelet associated genes (Supplementary table 1), individuals without a variant in a previously IT associated gene were analysed for variants in novel genes. WES analysis revealed potentially damaging candidate variants occurring within three families with currently unknown genetic aetiology (Table 3). All candidate variants are novel (excluding a previously annotated variant in *MKL1;* p.Val575Met, which occurs at a frequency of 0.007718 within the ExAC consortium), segregate with the disease status and have been confirmed by Sanger sequencing.

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

WES analysis of two related patients (F35.I and F35.II) of South Asian ethnicity was approached differently to other patients in this study. Both patients displayed a similarly severe clinical phenotype with a significant reduction in circulating platelets ($15x10^9$/L). PFT revealed a reduction in P-selectin (CD62P) expression upon stimulation and variable fluorescent fibrinogen binding which was consistent across both affected individuals. The patients were cousins born from consanguineous relationships within a single consanguineous kindred so analysis was focused on identification of a shared homozygous variant due to the recessive segregation of disease. Three variants occurring within *ANKRD18A;* p.Glu799del, *GNE;* p.Gly447Arg and *FRMPD1;* p.Ala509Val were present within both affected individuals and within a tightly linked region of homozygosity on chromosome 9p. The variations within *ANKRD18A* and *GNE* were novel within the databases previously mentioned whereas the variant in *FRMPD1* has been observed at a frequency of 0.0003708 including 39 times within the South Asian population (rs571037699). There is no ClinVar entry for this variant and all three variants are classified as variants of "uncertain significance".

One missense variant in the recently proposed IT linked gene; *MKL1*

One individual was shown to harbour a rare frequency (<0.01) missense variant within the Megakaryoblastic Leukaemia (translocation) 1 gene; *MKL1.* The variant was the only variant occurring within a gene of haemostatic relevance within 109 significant novel variants. The variant; *MKL1;* c.1723G>A, p.Val575Met present in patient F37.I has been noted previously at a frequency of 0.0007718 (allele count of

6/7774 in the ExAC consortium). The patient has a mild reduction in platelet count ($130x10^9$/L) with no secondary qualitative defects in platelet function observed. The variant is classified as of “unknown significance”.

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

A large kindred with three affected individuals and four unaffected related individuals were recruited to the study. A mild thrombocytopenia was observed within the family with platelet counts ranging from 80-$186x10^9$/L in the three affected individuals. All three affected individuals presented with a normal platelet size (7.9-8.6fL) and a mild reduction in secretion was observed in F30.I and F30.III which was not shared with F30.II. All affected individuals shared a similar bleeding phenotype, suffering from spontaneous epistaxis, excessive bruising and prolonged bleeding from minor wounds. WES analysis revealed 14 novel or rare frequency (<0.01) variants shared between the three affected individuals. Sanger sequencing of all 14 variants in four unaffected related individuals narrowed down candidates to only two missense variants; *PADI2* (p.Lys499Arg) and *TTF2* (p.His1089Asp). Both variants segregate with disease, not being present in the unaffected individuals. Both variants have been observed at a low frequency previously (<0.01) within the EXaC database (Table 3) and are currently classified as “uncertain significance”.

Discussion

Here we present the first large scale application of WES analysis to patients with inherited bleeding diatheses presenting with thrombocytopenia of unknown aetiology.

Platelet counts and phenotypic presentation vary among our patients considerably which is consistent with the variability observed in the spectrum of IT. However, the majority of patients (73%) were noted to have a secondary qualitative defect in platelet function which may explain the disproportionate bleeding when compared to the patient's platelet counts. A lack of consistency was noted in families 13 and 30 where affected individuals are observed both with and without defects in platelet function. Clinical complications are shared among the affected family members so this most likely represents limitations in the sensitivity of platelet function testing or intra-familial variability.

Overall, when considering pathogenicity WES analysis revealed 46% of index cases (17/37) to have a positive prediction of pathogenicity (classified "pathogenic" or "likely pathogenic" in a gene consistent with the patients phenotype and zygosity consistent with expected inheritance). Twenty-two percent of the index cases (8/37) were of uncertain/possible pathogenicity (results classified of "uncertain significance" in known IT causing genes). The remaining 32% of index cases (12/37) had a negative prediction of pathogenicity (no convincing variants identified in known causing genes). WES is not without its limitations and like with any genetic analysis all variants must be functionally confirmed as deleterious to the coded protein. However, our positive variant discovery rate is comparable to or exceeds previous large scale WES clinical multicentre studies of Mendelian disorders (22, 23).

Focusing our genetic analysis on patients with unknown aetiology of disease with minor prior genetic testing has produced a spectrum of variants different from previous large scale targeted genetic studies of IT. Patients were recruited to the study with clinically diagnosed bleeding disorders of unknown aetiology. One caveat of this approach results in possible exclusion of individuals with known BSS and MYH9-related disorders as these two forms of IT are routinely tested for in many haematological centres within the UK. However, 3 index cases have been noted with variants in either *GP1BA* or *MYH9* in our analysis representing cases with atypical presentation of BSS or MYH9-related disorder and therefore potentially falsely-negatively reported cases. The individuals with variants within *GP1BA* and *MYH9* showed a slight increase in MPV; however this was not at the magnitude of giant platelets normally attributed to this group of disorders and only patient F9.I showed any secondary syndromic symptoms with the individual suffering from congenital cataracts.

One attribute of excluding patients with known variants in *GP1BA, GP1BB, MYH9* and potentially *GP9* is the discovery of a relatively large percentage of individuals analysed (24% of index cases) with variants in *RUNX1* as a primary likely cause of disease. With the exception of one predicted loss of function variant, the variants present within *RUNX1* are currently classified as either “likely pathogenic” or of “uncertain significance” and need functional confirmation to be disease causing. However, the presence of these variants in a large number of individuals with an often shared secondary functional defect in secretion does suggest the prevalence of *RUNX1* variants may be higher than previously thought. This raises the question whether it should be considered as clinically significant as BSS and *MYH9*-related disorders and be primarily screened for genetically upon initial diagnosis with IT.

An advantage of using WES is the lack of limitations allowing the possibility of finding candidate variations in novel genes in cases that did not possess variants in known IT genes. To determine whether these variants are pathogenic relies on functional confirmation of the deleterious effect of the variant. However, WES analysis, especially with combined segregation analysis by Sanger sequencing in extensive kindreds, can provide indications as to which may be of scientific and clinical relevance. This strategy has recently been utilised in the discovery of novel candidate variations in *SLFN14* initially as part of the GAPP study (15, 24).

Family 35 is an interesting case of two affected related individuals born from consanguineous relationships. The molecular function of ANKRD18A is currently unknown and FRMPD1 functions to regulate the subcellular localisation of activator of G-protein signalling 3 (AGS3) (25). Both genes show weak expression in cells of the haematopoietic lineage, however, *GNE*; an enzyme in the sialic acid biosynthetic pathway, is expressed within all cells of the haematopoietic lineage. There are currently 88 registered mutations in *GNE* within the Human Genome Mutation Database (www.hgmd.cf.ac.uk). Mutations are known to be the genetic cause of sialuria (OMIM269921) and Hereditary Inclusion Body Myopathy (HIBM; OMIM600737) (26, 27). Recently, two separate groups have reported patients with compound heterozygous variations in *GNE*, causing GNE related myopathy with congenital thrombocytopenia (28, 29). Platelet counts within the four reported affected individuals were below $45x10^9$/L; however no MPV measurements were recorded. None of the patients displayed signs of myopathy until mid-adolescence/early adulthood; F35.I and F35.II are currently aged 10 and 6 respectively. Without functional characterisation of the effects of each variation, we cannot definitively conclude the genetic aetiology of these two individuals' severe

thrombocytopenia. However, WES analysis has allowed us to focus our efforts on three potentially pathogenic variants in novel genes.

MKL1 was initially included in our panel of 358 genes for post WES analysis due to its role in megakaryocyte maturation elucidated via its binding partner Serum response factor (SRF) (30-32). Recently, the first case of a homozygous mutation in *MKL1* in a patient with a severe immunodeficiency and no haematological malignancies was reported (33). One interesting phenotypic presentation within the affected individual was an intermittent mild thrombocytopenia with a reduced platelet count in whole blood of 50-150x10^9/L. Here we present one novel variant within *MKL1,* at a highly conserved genetic site. The missense variant observed in F37.I represents the only variant to occur in a gene with previous haematological implications. One further variant in *MKL1* was observed in addition to a "likely pathogenic" frameshift causing insertion within *TUBB1* in patient F25.I. Due to the predicted loss of function of the frameshift causing *TUBB1* variant it is unlikely that the variant with *MKL1* is additive to patient F25.I's phenotype. However, the presence of the variant of uncertain significance in patient F37.I is an interesting candidate to take forward for functional studies.

WES and segregation determination using Sanger sequencing revealed candidate variants in *PADI2* and *TTF2* that segregate with disease in F30.I, .II and .III. Phenotypic presentation does vary between the patients but clinical presentation remains consistent which may reflect limitations in the sensitivity of platelet function testing. Neither gene has previously been implicated in haematological abnormalities with mutations in *PADI2* causing schizophrenia, breast cancer and rheumatoid arthritis, and mutations in *TTF2* associated with Thyroid dysgenesis (34-37). WES analysis has therefore provided us with the first steps in determining the impact of

these two variants of uncertain significance and whether they have the propensity to be disease causing.

In summary, we show WES can be applied to identify the underlying genetic cause in known IT causing genes for patients with thrombocytopenia and unclear aetiologies of disease. We show similar positive detection rates when compared to prior targeted studies and with the addition of complementary functional studies show an improved detection rate when compared to WES analysis of other developmental disorders. We also suggest the applicability of WES in providing preliminary insight into novel genes and their potential mechanism of action through candidate variations of unknown significance. This approach provides a foundation to enhance our current knowledge on megakaryopoiesis, platelet function and platelet senescence/death upon subsequent functional studies.

Acknowledgements

We thank the families for providing samples and our clinical and laboratory colleagues for their help. This work was supported by the British Heart Foundation (RG/PG/13/36/30275; RG/09/007), an MRC Doctoral Training Partnership grant (BJ), a Wellcome Trust Combined Training Programme Fellowship (093994) (GCL), the Healing Foundation (PH) and the Platelet Charity. We thank the NIHR Haematology Specialty Group for their help in recruiting to the study, and all our clinical investigators and collaborators. The authors also acknowledge support from the Department of Health via the National Institute for Health Research (NIHR) comprehensive Biomedical Research Centre award to Guy's & St Thomas' NHS Foundation Trust in partnership with King's College London and King's College Hospital NHS Foundation Trust. We thank the Queen Elizabeth Hospital Charity for funding the Sysmex XN-1000.

Authorship Contributions

BJ, GCL, ML, SPW and NVM designed the research. BJ, GCL, JF, ML, DM, MAS, ISG, SD, DB, VL, SJF, BD, JR, PH and NVM performed the research and analysed data. GCL, DA, TB, PHB B-M, PC, NC, CG, BJ, MM, JM, SP, KT, JT, JW and MW provided patient samples and clinical data. GCL, ML, SPW and NVM undertook the research governance of the study. BJ and NVM wrote the paper and all authors critically reviewed and edited the paper. GL, PH, PG, SM, AM, MD, SPW and NVM coordinated the GAPP study.

Disclosure of Conflicts of Interest

PH was a previous consultant for Sysmex UK.

References

1. Noris P, Perrotta S, Seri M, Pecci A, Gnan C, Loffredo G, et al. Mutations in ANKRD26 are responsible for a frequent form of inherited thrombocytopenia: analysis of 78 patients from 21 families. *Blood*. 2011 Jun 16;**117**(24):6673-80.
2. Balduini CL. Diagnosis and management of inherited thrombocytopenias. European Human Genetics Conference 2014; 2014; Milan, Italy; 2014.
3. Johnson B, Fletcher SF, Morgan NV. Inherited thrombocytopenia: Novel insights into megakaryocyte maturation, proplatelet formation and platelet lifespan. *Platelets*. 2016 **In press**.
4. Canales ML, Mauer AM. Sex-linked hereditary thrombocytopenia as a variant of Wiskott-Aldrich syndrome. *N Engl J Med*. 1967 Oct 26;**277**(17):899-901.
5. Kunishima S, Kobayashi R, Itoh TJ, Hamaguchi M, Saito H. Mutation of the beta1-tubulin gene associated with congenital macrothrombocytopenia affecting microtubule assembly. *Blood*. 2009 Jan 8;**113**(2):458-61.
6. Kunishima S, Okuno Y, Yoshida K, Shiraishi Y, Sanada M, Muramatsu H, et al. ACTN1 mutations cause congenital macrothrombocytopenia. *Am J Hum Genet*. 2013 Mar 7;**92**(3):431-8.
7. Nichols KE, Crispino JD, Poncz M, White JG, Orkin SH, Maris JM, et al. Familial dyserythropoietic anaemia and thrombocytopenia due to an inherited mutation in GATA1. *Nat Genet*. 2000 Mar;**24**(3):266-70.
8. Stockley J, Morgan NV, Bem D, Lowe GC, Lordkipanidze M, Dawood B, et al. Enrichment of FLI1 and RUNX1 mutations in families with excessive bleeding and platelet dense granule secretion defects. *Blood*. 2013 Dec 12;**122**(25):4090-3.
9. Zhang MY, Churpek JE, Keel SB, Walsh T, Lee MK, Loeb KR, et al. Germline ETV6 mutations in familial thrombocytopenia and hematologic malignancy. *Nat Genet*. 2015 Feb;**47**(2):180-5.
10. Balduini CL, Pecci A, Noris P. Inherited thrombocytopenias: the evolving spectrum. *Hamostaseologie*. 2012;**32**(4):259-70.
11. Balduini CL, Savoia A. Genetics of familial forms of thrombocytopenia. *Hum Genet*. 2012 Dec;**131**(12):1821-32.
12. Savoia A. Molecular basis of inherited thrombocytopenias. *Clin Genet*. 2015 May 7.
13. Dawood BB, Wilde J, Watson SP. 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*. 2007 Aug;**18**(5):329-45.
14. Dawood BB, Lowe GC, Lordkipanidze M, Bem D, Daly ME, Makris M, et al. Evaluation of participants with suspected heritable platelet function disorders including recommendation and validation of a streamlined agonist panel. *Blood*. 2012 Dec 13;**120**(25):5041-9.
15. Fletcher SJ, Johnson B, Lowe GC, Bem D, Drake S, Lordkipanidze M, et al. SLFN14 mutations underlie thrombocytopenia with excessive bleeding and platelet secretion defects. *J Clin Invest*. 2015 Sep;**125**(9):3600-5.
16. Leo VC, Morgan NV, Bem D, Jones ML, Lowe GC, Lordkipanidze M, et al. Use of next-generation sequencing and candidate gene analysis to identify underlying defects in patients with inherited platelet function disorders. *J Thromb Haemost*. 2015 Apr;**13**(4):643-50.
17. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. 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*. 2015 May;**17**(5):405-24.
18. Plagnol V, Curtis J, Epstein M, Mok KY, Stebbings E, Grigoriadou S, et al. A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. *Bioinformatics*. 2012 Nov 1;**28**(21):2747-54.
19. Seri M, Cusano R, Gangarossa S, Caridi G, Bordo D, Lo Nigro C, et al. Mutations in MYH9 result in the May-Hegglin anomaly, and Fechtner and Sebastian syndromes. The May-Heggllin/Fechtner Syndrome Consortium. *Nature genetics*. 2000 Sep;**26**(1):103-5.
20. Preudhomme C, Renneville A, Bourdon V, Philippe N, Roche-Lestienne C, Boissel N, et al. High frequency of RUNX1 biallelic alteration in acute myeloid leukemia secondary to familial platelet disorder. *Blood*. 2009 May 28;**113**(22):5583-7.
21. Kamachi Y, Ogawa E, Asano M, Ishida S, Murakami Y, Satake M, et al. Purification of a mouse nuclear factor that binds to both the A and B cores of the polyomavirus enhancer. *J Virol*. 1990 Oct;**64**(10):4808-19.
22. Chong JX, Buckingham KJ, Jhangiani SN, Boehm C, Sobreira N, Smith JD, et al. The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities. *Am J Hum Genet*. 2015 Aug 6;**97**(2):199-215.
23. Yang Y, Muzny DM, Reid JG, Bainbridge MN, Willis A, Ward PA, et al. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N Engl J Med*. 2013 Oct 17;**369**(16):1502-11.

24. Marconi C, Di Buduo CA, Barozzi S, Palombo F, Pardini S, Zaninetti C, et al. SLFN14-related thrombocytopenia: identification within a large series of patients with inherited thrombocytopenia. *Thromb Haemost*. 2016 Jan 14;**115**(5).
25. An N, Blumer JB, Bernard ML, Lanier SM. 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*. 2008 Sep 5;**283**(36):24718-28.
26. Seppala R, Lehto VP, Gahl WA. 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*. 1999 Jun;**64**(6):1563-9.
27. Eisenberg I, Avidan N, Potikha T, Hochner H, Chen M, Olender T, et al. The UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase gene is mutated in recessive hereditary inclusion body myopathy. *Nat Genet*. 2001 Sep;**29**(1):83-7.
28. Izumi R, Niihori T, Suzuki N, Sasahara Y, Rikiishi T, Nishiyama A, et al. GNE myopathy associated with congenital thrombocytopenia: a report of two siblings. *Neuromuscul Disord*. 2014 Dec;**24**(12):1068-72.
29. Zhen C, Guo F, Fang X, Liu Y, Wang X. A family with distal myopathy with rimmed vacuoles associated with thrombocytopenia. *Neurol Sci*. 2014 Sep;**35**(9):1479-81.
30. Cheng EC, Luo Q, Bruscia EM, Renda MJ, Troy JA, Massaro SA, et al. Role for MKL1 in megakaryocytic maturation. *Blood*. 2009 Mar 19;**113**(12):2826-34.
31. Smith EC, Thon JN, Devine MT, Lin S, Schulz VP, Guo Y, et al. MKL1 and MKL2 play redundant and crucial roles in megakaryocyte maturation and platelet formation. *Blood*. 2012 Sep 13;**120**(11):2317-29.
32. Halene S, Gao Y, Hahn K, Massaro S, Italiano JE, Jr., Schulz V, et al. Serum response factor is an essential transcription factor in megakaryocytic maturation. *Blood*. 2010 Sep 16;**116**(11):1942-50.
33. Record J, Malinova D, Zenner HL, Plagnol V, Nowak K, Syed F, et al. Immunodeficiency and severe susceptibility to bacterial infection associated with a loss-of-function homozygous mutation of MKL1. *Blood*. 2015 Sep 24;**126**(13):1527-35.
34. Watanabe Y, Nunokawa A, Kaneko N, Arinami T, Ujike H, Inada T, et al. A two-stage case-control association study of PADI2 with schizophrenia. *J Hum Genet*. 2009 Jul;**54**(7):430-2.
35. McElwee JL, Mohanan S, Griffith OL, Breuer HC, Anguish LJ, Cherrington BD, et al. Identification of PADI2 as a potential breast cancer biomarker and therapeutic target. *BMC Cancer*. 2012;**12**:500.
36. Chang X, Xia Y, Pan J, Meng Q, Zhao Y, Yan X. PADI2 is significantly associated with rheumatoid arthritis. *PLoS One*. 2013;**8**(12):e81259.
37. Castanet M, Polak M. Spectrum of Human Foxe1/TTF2 Mutations. *Horm Res Paediatr*. 2010;**73**(6):423-9.

Table legends

Table 1

Platelet and bleeding phenotypes of 55 patients recruited to the UK-GAPP study. Average platelet count = 85x10^9/L (normal range to 2SD 147-327x10^9/L, n=40). Average MPV = 10fL (mean normal range to 2SD 7.8-12.69fL, n=40). IPF was available for 11 patients and varied between 1.8-87% (normal range 1.3-10.8%, n=40). Patients with an observed macro and micro thrombocytopenia are denoted by a + and -, respectively, following their most recent analysed MPV. Secondary qualitative defects are abbreviated to the following; (Gi) - reduction in response upon ADP stimulation indicating a possible defect in the Gi pathway, (GPVI) – reduction in surface GPVI quantity. Bleeding diathesis of each individual is summarised under bleeding phenotype.

Table 2

Results of whole exome sequencing analysis of 55 patients with inherited thrombocytopenia showing variants in known thrombocytopenia causing genes. 68% of individuals have a predicted genetic aetiology in a previously IT associated gene. When a variant has been previously observed it is annotated in the prevalence column with the database it is observed in. The ACMG consensus guideline results are also displayed in the final column (17).

Table 3

Potentially damaging variants in novel candidate genes. When a variant has been previously observed it is annotated in the prevalence column with the database it is observed in. 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. Mutationtaster uses a Beyes classifier to predict the effect of a mutation from a feed a 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.

Table 1

Family	Patient	Platelet count (x10^9/l)	MPV (fL)	IPF (%)	Secondary defect	Bleeding Phenotype
1	I	73	9.6		Yes (Fibrinogen)	Cutaneous bruising/bleeding, menorrhagia
2	I	50	8.6		No	Cutaneous bruising
3	I	80	10.3		Yes (Cyclooxygenase)	Cutaneous bruising, menorrhagia
	II	50	12		Yes (Cyclooxygenase)	Cutaneous bruising, epistaxis, purpura
	III	98	10.5	3.2	N/A	Cutaneous bruising, epistaxis, purpura
4	I	142	11.8+		Yes (Secretion and Gi)	Oral cavity bleeding, epistaxis, menorrhagia
	II	157	11.4+		Yes (Secretion)	Cutaneous bruising, oral cavity bleeding
5	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
6	I	110			Yes (Secretion)	Cutaneous bruising, excessive bleeding following surgery
	II	100	8.9		Yes (Secretion)	Cutaneous bruising, epistaxis
7	I	50	10.4		Yes (Secretion)	Cutaneous bruising, menorrhagia
8	I	70	10.7+		No	Cutaneous bruising, menorrhagia, post-partum haemorrhage
	II	70	10.1		No	Cutaneous bruising, epistaxis, haematuria, menorrhagia, post-partum haemorrhage
9	I	35	11.4+		No	Epistaxis, cutaneous bruising.
10	I	55			N/A	Cutaneous bruising, epistaxis
11	I	62			Yes (Secretion)	Cutaneous bruising
	II	N/A			Yes (Secretion)	Cutaneous bruising, epistaxis
	III	146			Yes (Secretion)	Cutaneous bruising, epistaxis, menorrhagia
12	I	100	8		Yes (Secretion)	Excessive cutaneous bleeding
13	I	163	9.1		No	Cutaneous bruising, epistaxis, haematoma
	II	45	11.9+		Yes (GPVI)	Cutaneous bruising, epistaxis
14	I	139	8		Yes (Secretion)	Epistaxis, haematoma
15	I	90	7.6-		Yes (Secretion)	Cutaneous bruising, petechiae
16	I	130	7.1-		Yes (Secretion and Gi)	Cutaneous bruising, epistaxis, oral cavity bleeding
	II	70	7.5-		Yes (Secretion and Gi)	Cutaneous bruising, epistaxis, oral cavity bleeding
17	I	N/A	N/A		N/A	Excessive bruising/bleeding
18	I	110	8.1		Yes (Secretion)	Cutaneous bruising/bleeding, petechiae, haematoma
19	I	100	9		No	Cutaneous bruising/bleeding
	II	100	9.2		No	Cutaneous bruising/bleeding
20	I	89	13+	17.5	Yes (Secretion and Gi)	Cutaneous bruising
21	I	63	11.9	19.1	Yes (Secretion)	Cutaneous bruising, epistaxis, haematoma
	II	83	11.9	24.3	Yes (Secretion)	Cutaneous bruising, epistaxis, haematoma
22	I	74	11.2		Yes (Secretion and Gi)	Cutaneous bruising/bleeding, haematuria
	II	62	12.7+	20.8	Yes (Secretion and Gi)	Cutaneous bruising/bleeding, menorrhagia, post-partum haemorrhage, haematoma
	III	109	11		Yes (Secretion)	Cutaneous bruising, haematoma, menorrhagia
23	I	119	11.1		Yes (Secretion)	Cutaneous bruising/bleeding
24	I	104	9.6		No	Menorrhagia, post-partum haemorrhage
	II	133	8.6		No	Epistaxis
25	I	11	13.4+		Yes (Secretion and Gi)	Cutaneous bruising
26	I	43	14+		No	Cutaneous bruising, menorrhagia, oral cavity bleeding
27	I	100	10.3		Yes (Secretion)	Cutaneous bruising/bleeding, epistaxis, oral cavity bleeding
28	I	25	8.5		Yes (Secretion)	Cutaneous bruising
29	I	15	9.4		Yes (Secretion)	Haematomas
30	I	137	7.9		Yes (Secretion)	Cutaneous bruising, epistaxis, menorrhagia
	II	186	8.6		No	Cutaneous bruising, menorrhagia, haematuria
	III	80	8.2		Yes (Secretion)	Cutaneous bruising
31	I	20	9.7		N/A	Cutaneous bruising, epistaxis, oral cavity bleeding
32	I	15	9.5	20.2	No	Cutaneous bruising
33	I	66	9.9	1.8	Yes (Secretion)	Cutaneous bleeding
34	I	93	14.4+	20.5	Yes (Secretion)	Cutaneous bleeding, epistaxis
35	I	15	10.4	87	Yes (Secretion and other)	Cutaneous bruising, epistaxis, haematomas
	II	14	15+	83	Yes (Secretion and other)	Cutaneous bleeding
36	I	104	13.3+	17	No	Menorrhagia
37	I	130	9.7		No	Cutaneous bruising, epistaxis, menorrhagia

Table 2

Family	Patient	Gene(s)	Genomic variation	Protein effect	Variation type	Prevalence	Classification
1	I	ACTN1	c.2647G>C	p.Gly883Arg	Missense	Novel	**Likely Pathogenic**
2	I	ANKRD26	c.-126T>G		5'-UTR variation	Known	**Pathogenic**
3	I	CYCS	c.155C>T	p.Ala52Val	Missense	Novel	**Likely Pathogenic**
	II	CYCS	c.155C>T	p.Ala52Val	Missense	Novel	
	III	CYCS	c.155C>T	p.Ala52Val	Missense	Novel	
4	I	FLI1	c.992_995del	p.Asn331Thr fs*4	Frameshift deletion	Novel	**Pathogenic**
	II	FLI1	c.992_995del	p.Asn331Thr fs*4	Frameshift deletion	Novel	
5	I	FLI1	c.1028A>G	p.Tyr343Cys	Missense	Novel	**Likely Pathogenic**
	II	FLI1	c.1028A>G	p.Tyr343Cys	Missense	Novel	
6	I	GFI1B	c.814+1G>A		Splicing	Novel	**Likely Pathogenic**
		STIM1	c.1828G>A	p.Ala610Thr	Missense	0.00019 (1k)	**Uncertain significance**
	II	GFI1B	c.814+1G>A		Splicing	Novel	**Likely Pathogenic**
		STIM1	c.1828G>A	p.Ala610Thr	Missense	0.00019 (1k)	**Uncertain significance**
7	I	GP1BA	c.1761A>C	p.Gln587His	Missense	0.00043 (EXaC) (rs570515282)	**Uncertain significance**
8	I	GP1BA	c.413G>T	p.Gly138Val	Missense	Novel	**Likely Pathogenic**
	II	GP1BA	c.413G>T	p.Gly138Val	Missense	Novel	
9	I	MYH9	c.3493C>T	p.Arg1165Cys	Missense	Known (rs80338829)	**Pathogenic**
10	I	NBEAL2	c.1376delT	p.Leu459Arg fs*13	Frameshift deletion	Novel	**Likely Pathogenic**
		NBEAL2	c.6893A>G	p.Asn2298Ser	Missense	Novel	**Likely Pathogenic**
11	I	RUNX1	c.16G>A	p.Asp6Asn	Missense	Novel	**Likely Pathogenic**
	II	RUNX1	c.16G>A	p.Asp6Asn	Missense	Novel	
	III	RUNX1	c.16G>A	p.Asp6Asn	Missense	Novel	
12	I	RUNX1	c.236G>A	p.Trp79*	Nonsense	Novel	**Pathogenic**
		ITGB3	c.349C>T	p.Arg117Trp	Missense	Novel	**Uncertain significance**
13	I	RUNX1	c.270+1G>T		Splicing	Novel	**Likely Pathogenic**
	II	RUNX1	c.270+1G>T		Splicing	Novel	
14	I	RUNX1	c.270+1G>T		Splicing	Novel	
15	I	RUNX1	c.322G>A	p.Gly108Ser	Missense	Novel	**Uncertain significance**
16	I	RUNX1	c.427+1G>T		Splicing	Novel	**Likely Pathogenic**
	II	RUNX1	c.427+1G>T		Splicing	Novel	
17	I	RUNX1	c.505A>G	p.Thr169Ala	Missense	Novel	**Uncertain significance**
18	I	RUNX1	c.512A>T	p.Asp171Val	Missense	Novel	**Uncertain significance**
19	I	RUNX1	c.530G>A	p.Arg177Gln	Missense	Known	**Likely Pathogenic**
	II	RUNX1	c.530G>A	p.Arg177Gln	Missense	Known	
20	I	SLFN14	c.652A>G	p.Lys218Glu	Missense	Novel	**Uncertain significance**
21	I	SLFN14	c.657A>T	p.Lys219Asn	Missense	Novel	**Uncertain significance**
	II	SLFN14	c.657A>T	p.Lys219Asn	Missense	Novel	
22	I	SLFN14	c.659T>A	p.Val220Asp	Missense	Novel	**Likely Pathogenic**
	II	SLFN14	c.659T>A	p.Val220Asp	Missense	Novel	
	III	SLFN14	c.659T>A	p.Val220Asp	Missense	Novel	
23	I	TPM4	c.548C>T	p.Ala183Val	Missense	Novel	**Uncertain significance**
		TUBB1	c.726C>G	p.Phe242Leu	Missense	Novel	**Uncertain significance**
24	I	TUBB1	c.721C>T	p.Arg241Trp	Missense	0.0001071 (ExAC)(rs368923302)	**Uncertain significance**
	II	TUBB1	c.721C>T	p.Arg241Trp	Missense	0.0001071 (ExAC)(rs368923302)	
25	I	TUBB1	c.1080_1081insG	p.Leu361Ala fs*19	Frameshift insertion	Novel	**Likely Pathogenic**
26	I	Unknown					
27	I	Unknown					
28	I	Unknown					
29	I	Unknown					
30	I	Unknown					
	II	Unknown					
	III	Unknown					
31	I	Unknown					
32	I	Unknown					
33	I	Unknown					
34	I	Unknown					
35	I	Unknown					
	II	Unknown					
36	I	Unknown					
37	I	Unknown					

Table 3

Family	Gene	Variant	Protein effect	Prevalence	PhyloP	PhastCons	Mutation taster	SIFT	Provean	PolyPhen-2	ACMG	Classification
30	*PADI2*	c.1496A>G	p.Lys499Arg	0.000008681	1.647	1	Disease causing	Tolerated	Neutral	Benign	PM (segregation)	**Uncertain Significance**
	TTF2	c.3265C>G	p.His1089Asp	1.65E-05	5.131	1	Disease causing	Damaging	Deleterious	Damaging	PM (segregation) , PP3	**Uncertain Significance**
35	*ANKRD18A*	c.2395_2397del^{hom}	p.Glu799del^{hom}	Novel	0.772	0.965	Polymorphism	NA	Deleterious	NA	PM2, PP (segregation), PM6	**Uncertain Significance**
	GNE	c.1339G>A^{hom}	p.Gly447Arg^{hom}	Novel	5.343	1	Disease causing	Damaging	Neutral	Damaging	PM2, PP (segregation), PM6	**Uncertain Significance**
	FRMPD1	c.1526C>T^{hom}	p.Ala509Val^{hom}	0.0003708	-1.459	0	Polymorphism	Tolerated	Neutral	Benign	PP (segregation), PM6	**Uncertain Significance**
37	*MKL1*	c.1723G>A	p.Val575Met	0.0007718	3.358	1	Disease causing	Damaging	Neutral	Damaging		**Uncertain Significance**

Figure 1

Bioinformatics pipeline analysis of whole exome sequencing data. Initial WES analysis focused upon comparison to a panel of 358 genes (supplementary table 1). After which screening of exomes variants focused upon novel variants. Variant classification was performed utilising the ACMG consensus guidelines.

Figure 2

Spatial amino acid locations of all thrombocytopenia causing variants present within RUNT transcription factor 1 (RUNX1) (RefSeq NP_001001890). Previously disease causing variants found within the HGMD (www.hgmd.cf.ac.uk) and ClinVar (www.ncbi.nlm.nih.gov/clinvar/) databases are denoted above. The eight variants found within *RUNX1* in the GAPP cohort of 54 patients who have undergone whole exome sequencing are denoted below and the effect on the protein or predicted splice-site shown.

Figure 1

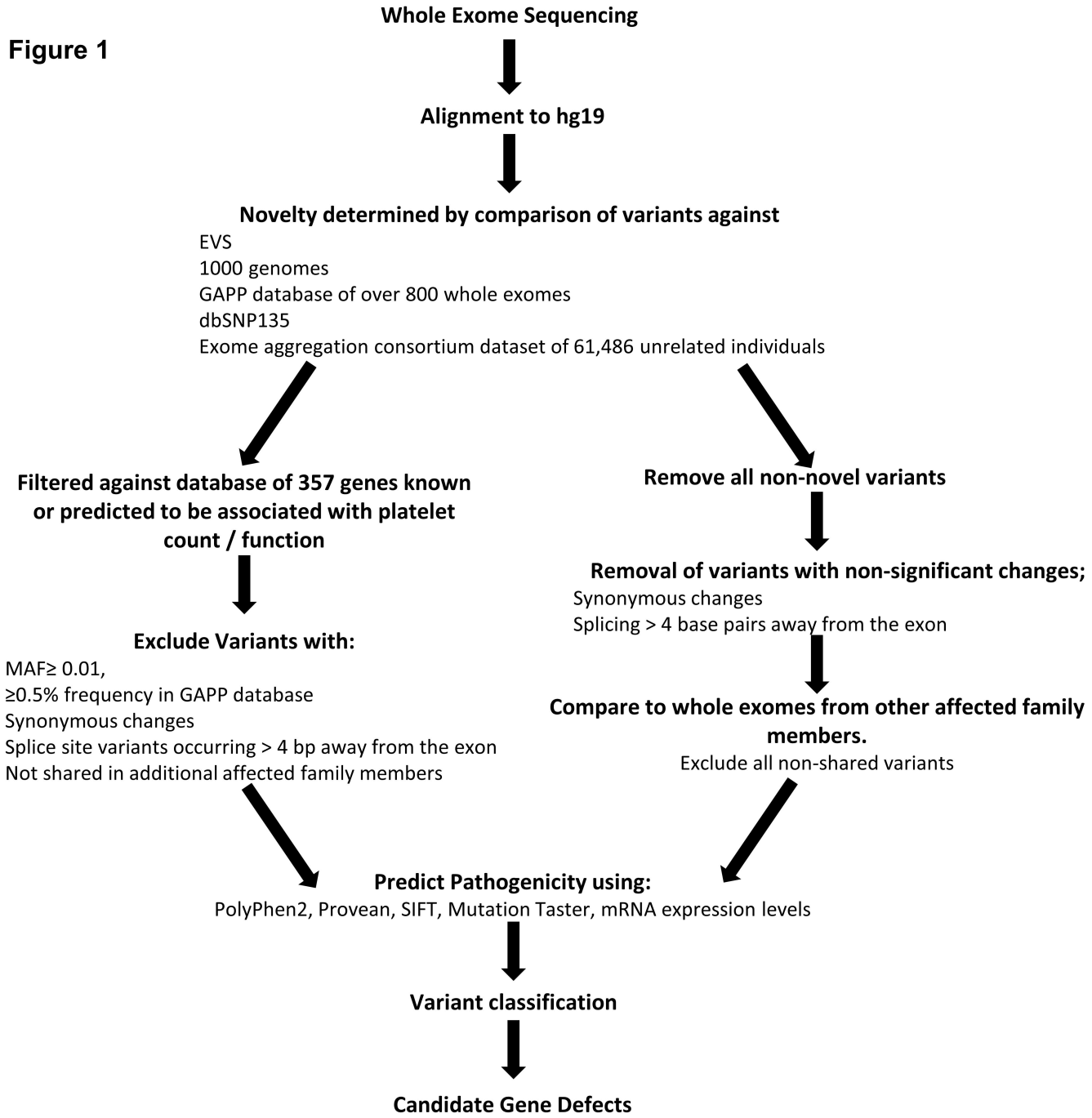
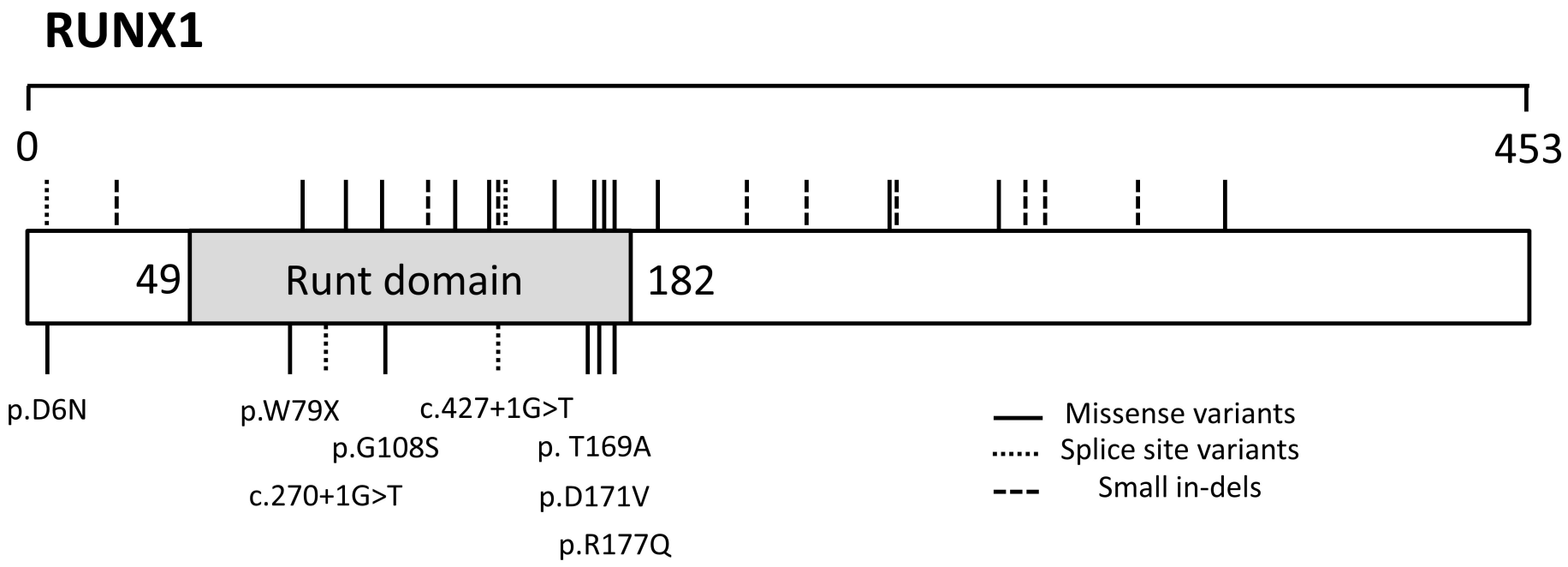

Whole Exome Sequencing
Alignment to hg19
Novelty determined by comparison of variants against
EVS
1000 genomes
GAPP database of over 800 whole exomes
dbSNP135
Exome aggregation consortium dataset of 61,486 unrelated individuals
Filtered against database of 357 genes known or predicted to be associated with platelet count / function
Exclude Variants with:
MAF≥ 0.01,
≥0.5% frequency in GAPP database
Synonymous changes
Splice site variants occurring > 4 bp away from the exon
Not shared in additional affected family members
Remove all non-novel variants
Removal of variants with non-significant changes;
Synonymous changes
Splicing > 4 base pairs away from the exon
Compare to whole exomes from other affected family members.
Exclude all non-shared variants
Predict Pathogenicity using:
PolyPhen2, Provean, SIFT, Mutation Taster, mRNA expression levels
Variant classification
Candidate Gene Defects

Figure 2


RUNX1
0
453
49
Runt domain
182
p.D6N
p.W79X
c.427+1G>T
p.G108S
p. T169A
c.270+1G>T
p.D171V
p.R177Q
Missense variants
Splice site variants
Small in-dels