

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

UK GAPP Study Group; Harrison, Paul

DOI:

[10.3324/haematol.2016.146316](https://doi.org/10.3324/haematol.2016.146316)

License:

Creative Commons: Attribution (CC BY)

*Document Version*

Publisher's PDF, also known as Version of record

*Citation for published version (Harvard):*

UK GAPP Study Group & Harrison, P 2016, 'Whole exome sequencing identifies genetic variants in inherited thrombocytopenia with secondary qualitative function defects', *Haematologica*, vol. 101, pp. 1170-1179. <https://doi.org/10.3324/haematol.2016.146316>

[Link to publication on Research at Birmingham portal](#)

**Publisher Rights Statement:**

Published as detailed above and available online at: <http://dx.doi.org/10.3324/haematol.2016.146316>

**General rights**

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

**Take down policy**

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact [UBIRA@lists.bham.ac.uk](mailto:UBIRA@lists.bham.ac.uk) providing details and we will remove access to the work immediately and investigate.



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

by Ben Johnson, Gillian C. Lowe, Jane Futterer, Marie Lordkipanidze', David MacDonald, Michael A. Simpson, Isabel Sánchez Guiu', Sian Drake, Danai Bem, Vincenzo Leo, Sarah J. Fletcher, Ban Dawood, Jose' Rivera, David Allsup, Tina Biss, Paula H.B. Bolton-Maggs, Peter Collins, Nicola Curry, Charlotte Grimley, Beki James, Mike Makris, Jayashree Motwani, Sue Pavord, Katherine Talks, Jecko Thachil, Jonathan Wilde, Mike Williams, Paul Harrison, Paul Gissen, Stuart Mundell, Andrew Mumford, Martina E. Daly, Steve P. Watson, and Neil V. Morgan

Haematologica 2016 [Epub ahead of print]

*Citation: Johnson B, Lowe GC, Futterer J, Lordkipanidze' M, MacDonald D, Simpson MA, Guiu' IS, Drake S, Bem D, Leo V, Fletcher SJ, Dawood B, Rivera J, Allsup D, Biss T, Bolton-Maggs PH, Collins P, Curry N, Grimley C, James B, Makris M, Motwani J, Pavord S, Talks K, Thachil J, Wilde J, Williams M, Harrison P, Gissen P, Mundell S, Mumford A, Daly ME, Watson SP, and Morgan NV. Whole exome sequencing identifies genetic variants in inherited thrombocytopenia with secondary qualitative function defects. Haematologica. 2016; 101:xxx  
doi:10.3324/haematol.2016.146316*

### *Publisher's Disclaimer.*

*E-publishing ahead of print is increasingly important for the rapid dissemination of science. Haematologica is, therefore, E-publishing PDF files of an early version of manuscripts that have completed a regular peer review and have been accepted for publication. E-publishing of this PDF file has been approved by the authors. After having E-published Ahead of Print, manuscripts will then undergo technical and English editing, typesetting, proof correction and be presented for the authors' final approval; the final version of the manuscript will then appear in print on a regular issue of the journal. All legal disclaimers that apply to the journal also pertain to this production process.*

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

Ben Johnson,<sup>1</sup> Gillian C. Lowe,<sup>1</sup> Jane Futterer,<sup>1</sup> Marie Lordkipanidzé,<sup>1</sup> David MacDonald,<sup>1</sup> Michael A. Simpson,<sup>2</sup> Isabel Sanchez-Guiú,<sup>3</sup> Sian Drake,<sup>1</sup> Danai Bem,<sup>1</sup> Vincenzo Leo,<sup>4</sup> Sarah J. Fletcher,<sup>1</sup> Ban Dawood,<sup>1</sup> José Rivera,<sup>3</sup> David Allsup,<sup>5</sup> Tina Biss,<sup>6</sup> Paula HB Bolton-Maggs,<sup>7</sup> Peter Collins,<sup>8</sup> Nicola Curry,<sup>9</sup> Charlotte Grimley,<sup>10</sup> Beki James,<sup>11</sup> Mike Makris,<sup>4</sup> Jayashree Motwani,<sup>12</sup> Sue Pavord,<sup>13</sup> Katherine Talks,<sup>6</sup> Jecko Thachil,<sup>7</sup> Jonathan Wilde,<sup>14</sup> Mike Williams,<sup>12</sup> Paul Harrison,<sup>15</sup> Paul Gissen,<sup>16</sup> Stuart Mundell,<sup>17</sup> Andrew Mumford,<sup>18</sup> Martina E. Daly,<sup>4</sup> Steve P. Watson,<sup>1</sup> Neil V. Morgan<sup>1</sup> on behalf of the UK GAPP Study Group

<sup>1</sup>Institute for Cardiovascular Sciences, College of Medical and Dental Sciences, University of Birmingham, UK

<sup>2</sup>Division of Genetics and Molecular Medicine, King's College, London, UK

<sup>3</sup>Centro Regional de Hemodonación, Universidad de Murcia, IMIB-Arrixaca, Murcia, Spain

<sup>4</sup>Department of Infection, Immunity and Cardiovascular Disease, University of Sheffield Medical School, University of Sheffield, UK

<sup>5</sup>Hull Haemophilia Treatment Centre, Hull and East Yorkshire Hospitals NHS trust, Castle Hill Hospital, Hull

<sup>6</sup>Department of Haematology, Royal Victoria Infirmary, Newcastle Upon Tyne, UK

<sup>7</sup>Department of Haematology, Manchester Royal Infirmary, Manchester, UK

<sup>8</sup>Arthur Bloom Haemophilia Centre, School of Medicine, Cardiff University, Cardiff, UK

<sup>9</sup>Oxford Haemophilia & Thrombosis Centre, Churchill Hospital, Oxford, UK

<sup>10</sup>Nottingham Haemophilia Centre, Nottingham University Hospital, Nottingham, UK

<sup>11</sup>Regional Centre for Paediatric Haematology, Leeds Children's Hospital, Leeds, UK

<sup>12</sup>Department of Haematology, Birmingham Children's Hospital, Birmingham, UK

<sup>13</sup>HaemostasisThrombosis Unit, Leicester Royal Infirmary, Leicester, UK

<sup>14</sup>Adult Haemophilia Centre, Queen Elizabeth Hospital, Birmingham, UK

<sup>15</sup>School of Immunity and Infection, College of Medical and Dental Sciences,  
University of Birmingham, UK

<sup>16</sup>Medical Research Council, Laboratory for Molecular Cell Biology, University  
College London, London, UK

<sup>17</sup>School of Physiology, Pharmacology and Neuroscience, University of Bristol,  
Bristol, UK

<sup>18</sup>School of Cellular and Molecular Medicine, 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](mailto:n.v.morgan@bham.ac.uk)

### 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  $11-186 \times 10^9/L$ . 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  $150 \times 10^9/L$  in whole blood. Platelet counts are considered normal when maintained at levels between  $150-450 \times 10^9/L$ .

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](http://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  $>1 \times 10^8/\text{mL}$  (n=13).

An in-house flow-cytometry assay was developed to assess platelet function in patients having platelet counts in PRP  $<1 \times 10^8/\text{mL}$  ( $n=22$ ). Platelets from individuals with borderline platelet counts in PRP between  $1.0$  and  $1.5 \times 10^8/\text{mL}$  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 GPIIb/IIIa were assessed. PRP was then stimulated with ADP (3 and 30  $\mu\text{M}$ ), CRP (0.3 and 3  $\mu\text{g/ml}$ ) and PAR-1 peptide (10 and 100  $\mu\text{M}$ ). 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  $37^\circ\text{C}$  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  $150 \times 10^9/L$ . Patients having platelet counts in the range of  $150-200 \times 10^9/L$  remained enrolled in the study if they showed a similar phenotype to related affected family members and a platelet count below  $150 \times 10^9/L$  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  $3.78 - 10.11 \times 10^9/L$ ,  $n=40$ ) 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*; p.Trp79\* 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 CBF $\beta$  (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 ( $15 \times 10^9/L$ ). 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 ( $130 \times 10^9/L$ ) 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  $80-186 \times 10^9/L$  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](http://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  $45 \times 10^9/L$ ; 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  $50-150 \times 10^9/L$ . 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, et al. Mutations in ANKRD26 are responsible for a frequent form of inherited thrombocytopenia: analysis of 78 patients from 21 families. *Blood*. 2011;117(24):6673-66780.
2. Balduini CL. Diagnosis and management of inherited thrombocytopenias. *European Human Genetics Conference 2014*; 2014; Milan, Italy.
3. Johnson B, Fletcher SF, Morgan NV. Inherited thrombocytopenia: Novel insights into megakaryocyte maturation, proplatelet formation and platelet lifespan. *Platelets*. 2016:1-7. Epub ahead of print.
4. Canales ML, Mauer AM. Sex-linked hereditary thrombocytopenia as a variant of Wiskott-Aldrich syndrome. *N Engl J Med*. 1967;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;113(2):458-461.
6. Kunishima S, Okuno Y, Yoshida K, et al. ACTN1 mutations cause congenital macrothrombocytopenia. *Am J Hum Genet*. 2013;92(3):431-438.
7. Nichols KE, Crispino JD, Poncz M, et al. Familial dyserythropoietic anaemia and thrombocytopenia due to an inherited mutation in GATA1. *Nat Genet*. 2000;24(3):266-270.
8. Stockley J, Morgan NV, Bem D, et al. Enrichment of FLI1 and RUNX1 mutations in families with excessive bleeding and platelet dense granule secretion defects. *Blood*. 2013;122(25):4090-4093.
9. Zhang MY, Churpek JE, Keel SB, et al. Germline ETV6 mutations in familial thrombocytopenia and hematologic malignancy. *Nat Genet*. 2015;47(2):180-185.
10. Balduini CL, Pecci A, Noris P. Inherited thrombocytopenias: the evolving spectrum. *Hamostaseologie*. 2012;32(4):259-270.
11. Balduini CL, Savoia A. Genetics of familial forms of thrombocytopenia. *Hum Genet*. 2012;131(12):1821-1832.
12. Savoia A. Molecular basis of inherited thrombocytopenias. *Clin Genet*. 2016;89(2):154-162.
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;18(5):329-345.
14. Dawood BB, Lowe GC, Lordkipanidze M, et al. Evaluation of participants with suspected heritable platelet function disorders including recommendation and validation of a streamlined agonist panel. *Blood*. 2012;120(25):5041-5049.
15. Fletcher SJ, Johnson B, Lowe GC, et al. SLFN14 mutations underlie thrombocytopenia with excessive bleeding and platelet secretion defects. *J Clin Invest*. 2015;125(9):3600-3605.
16. Leo VC, Morgan NV, Bem D, 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;13(4):643-650.
17. Richards S, Aziz N, Bale S, 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;17(5):405-424.
18. Plagnol V, Curtis J, Epstein M, et al. A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. *Bioinformatics*. 2012;28(21):2747-2754.
19. Seri M, Cusano R, Gangarossa S, et al. Mutations in MYH9 result in the May-Hegglin anomaly, and Fechtner and Sebastian syndromes. The May-Hegglin/Fechtner Syndrome Consortium. *Nature genetics*. 2000;26(1):103-105.
20. Preudhomme C, Renneville A, Bourdon V, et al. High frequency of RUNX1 biallelic alteration in acute myeloid leukemia secondary to familial platelet disorder. *Blood*. 2009;113(22):5583-5587.
21. Kamachi Y, Ogawa E, Asano 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;64(10):4808-4819.
22. Chong JX, Buckingham KJ, Jhangiani SN, et al. The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities. *Am J Hum Genet*. 2015;97(2):199-215.
23. Yang Y, Muzny DM, Reid JG, et al. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N Engl J Med*. 2013;369(16):1502-1511.
24. Marconi C, Di Buduo CA, Barozzi S, et al. SLFN14-related thrombocytopenia: identification within a large series of patients with inherited thrombocytopenia. *Thromb Haemost*. 2016;115(5):1076-1079.
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;283(36):24718-24728.
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;64(6):1563-1569.

27. Eisenberg I, Avidan N, Potikha T, et al. The UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase gene is mutated in recessive hereditary inclusion body myopathy. *Nat Genet.* 2001;29(1):83-87.
28. Izumi R, Niihori T, Suzuki N, et al. GNE myopathy associated with congenital thrombocytopenia: a report of two siblings. *Neuromuscul Disord.* 2014;24(12):1068-1072.
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;35(9):1479-1481.
30. Cheng EC, Luo Q, Bruscia EM, et al. Role for MKL1 in megakaryocytic maturation. *Blood.* 2009;113(12):2826-2834.
31. Smith EC, Thon JN, Devine MT, et al. MKL1 and MKL2 play redundant and crucial roles in megakaryocyte maturation and platelet formation. *Blood.* 2012;120(11):2317-2329.
32. Halene S, Gao Y, Hahn K, et al. Serum response factor is an essential transcription factor in megakaryocytic maturation. *Blood.* 2010;116(11):1942-1950.
33. Record J, Malinova D, Zenner HL, et al. Immunodeficiency and severe susceptibility to bacterial infection associated with a loss-of-function homozygous mutation of MKL1. *Blood.* 2015;126(13):1527-1535.
34. Watanabe Y, Nunokawa A, Kaneko N, et al. A two-stage case-control association study of PADI2 with schizophrenia. *J Hum Genet.* 2009;54(7):430-432.
35. McElwee JL, Mohanan S, Griffith OL, 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):423-429.

## Table legends

### Table 1

**Platelet and bleeding phenotypes of 55 patients recruited to the UK-GAPP study.** Average platelet count =  $85 \times 10^9/L$  (normal range to 2SD  $147-327 \times 10^9/L$ ,  $n=40$ ). 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 Bayes 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 <sup>9</sup> /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	GF1B	c.814+1G>A		Splicing	Novel	Likely Pathogenic
		STIM1	c.1828G>A	p.Ala610Thr	Missense	0.00019 (1k)	Uncertain significance
	II	GF1B	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	<i>PADI2</i>	c.1496A>G	p.Lys499Arg	0.000008681	1.647	1	Disease causing	Tolerated	Neutral	Benign	PM (segregation)	<b>Uncertain Significance</b>
	<i>TTF2</i>	c.3265C>G	p.His1089Asp	1.65E-05	5.131	1	Disease causing	Damaging	Deleterious	Damaging	PM (segregation) , PP3	<b>Uncertain Significance</b>
35	<i>ANKRD18A</i>	c.2395_2397del <sup>hom</sup>	p.Glu799del <sup>hom</sup>	Novel	0.772	0.965	Polymorphism	NA	Deleterious	NA	PM2, PP (segregation), PM6	<b>Uncertain Significance</b>
	<i>GNE</i>	c.1339G>A <sup>hom</sup>	p.Gly447Arg <sup>hom</sup>	Novel	5.343	1	Disease causing	Damaging	Neutral	Damaging	PM2, PP (segregation), PM6	<b>Uncertain Significance</b>
	<i>FRMPD1</i>	c.1526C>T <sup>hom</sup>	p.Ala509Val <sup>hom</sup>	0.0003708	-1.459	0	Polymorphism	Tolerated	Neutral	Benign	PP (segregation), PM6	<b>Uncertain Significance</b>
37	<i>MKL1</i>	c.1723G>A	p.Val575Met	0.0007718	3.358	1	Disease causing	Damaging	Neutral	Damaging		<b>Uncertain Significance</b>

## 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](http://www.hgmd.cf.ac.uk)) and ClinVar ([www.ncbi.nlm.nih.gov/clinvar/](http://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**

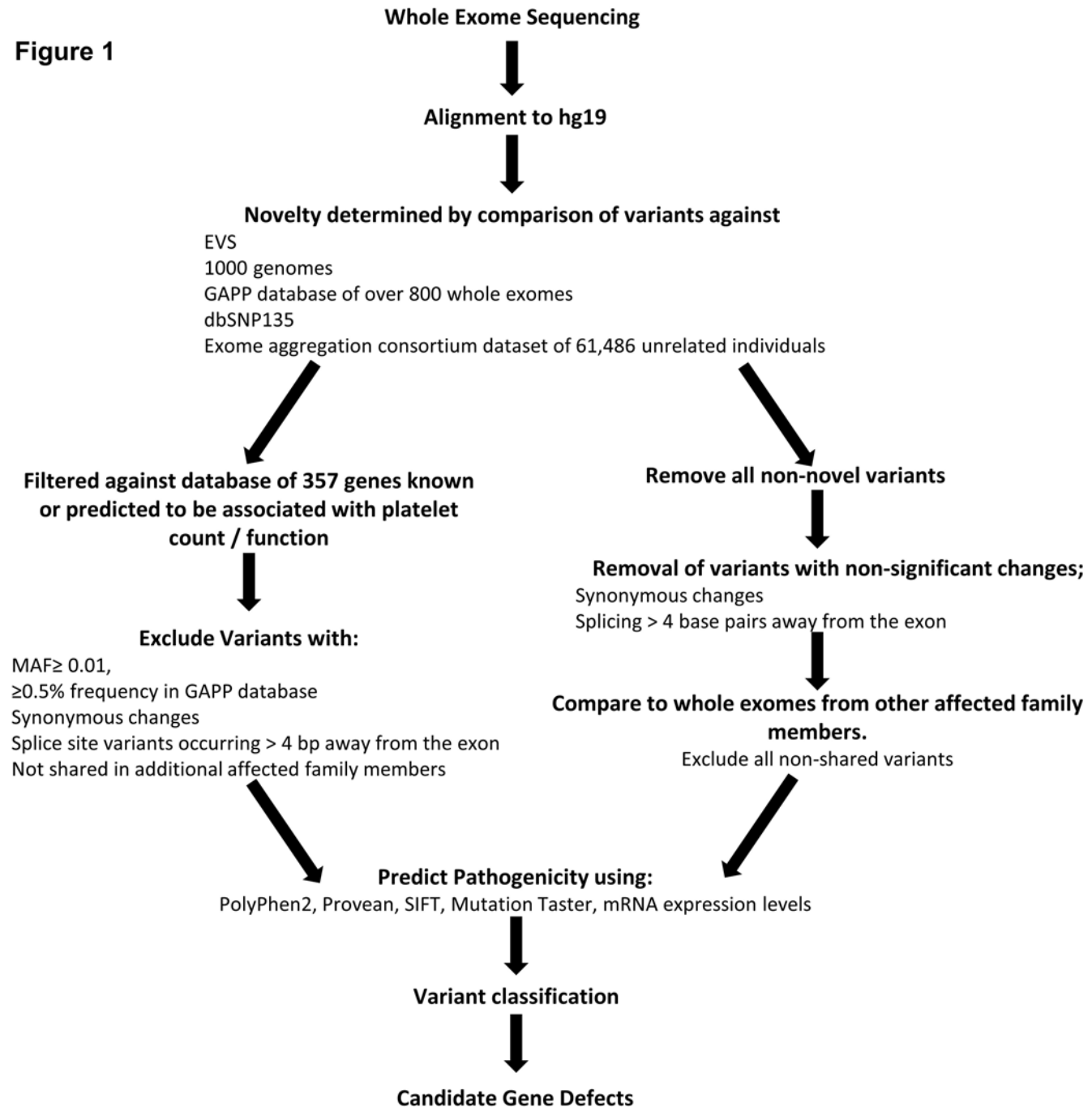
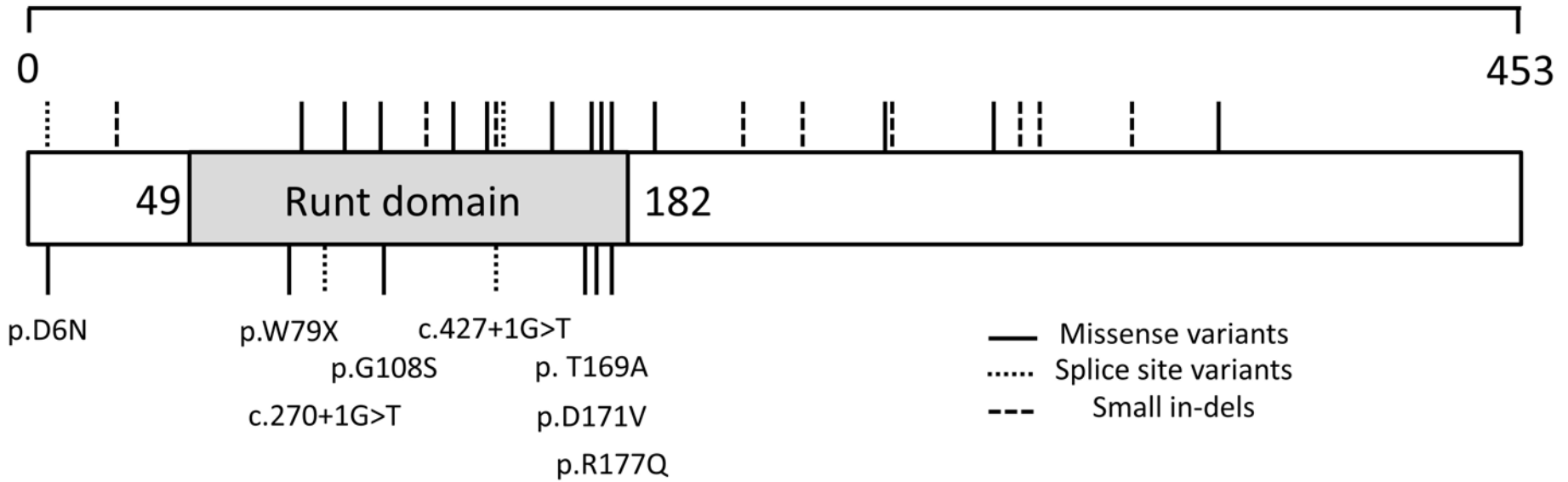




Figure 2

# RUNX1



## Supplementary Table 1

IT related panel genes utilised for WES bioinformatics analysis. Percentage of each (

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

GF11B	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



gene covered  $\leq 20x$  is also shown.