

## High-throughput sequencing approaches for diagnosing hereditary bleeding and platelet disorders

Kathleen Freson and Ernest Turro

### Abstract

Hereditary bleeding and platelet disorders (BPDs) are characterised by marked genetic heterogeneity, far greater than previously appreciated. The list of genes involved in the regulation of megakaryopoiesis, platelet formation, platelet function and bleeding has been growing rapidly since the introduction of high-throughput sequencing (HTS) approaches in research. Thanks to the gradual adoption of HTS in diagnostic practice, these discoveries are improving the diagnostic yield for BPD patients, who may or may not present with bleeding problems and often have other clinical symptoms unrelated to the blood system. However, it was previously found that screening for all known aetiologies gives a diagnostic yield of over 90% when the phenotype closely matches a known BPD but drops to 10% when the phenotype is indicative of a novel disorder. Thus, further research is needed to identify currently unknown aetiologies for BPDs. Novel genes are likely to be found to be implicated in BPDs. New modes of inheritance, including digenic inheritance, are likely to play a role in some cases. Additionally, identifying and interpreting pathogenic variants outside exons is a looming challenge that can only be tackled with an improved understanding of the regulatory landscape of relevant cell types and with the transition from targeted sequencing to whole-genome sequencing in the clinic.

### Overview of bleeding and platelet disorders

Hereditary bleeding and platelet disorders (BPDs) are a heterogeneous set of diseases affecting approximately 3 million people worldwide. The most common inherited bleeding disorders are von Willebrand disease (VWD), with a prevalence of 1 in 100 to 1,000 people, and haemophilia A, which affects 1 in 5,000 to 10,000 males at birth. Other BPDs are globally very rare, with some having been described in fewer than five unrelated families worldwide, although certain disorders have an elevated prevalence in specific subpopulations due to founder effects. Over the past five decades, the genetic bases of some of these disorders have been identified. Most of the genes harbouring variants responsible for BPDs (henceforth, BPD genes) have been identified through linkage studies across informative pedigrees or using candidate gene Sanger sequencing following thorough clinical and laboratory work-up. However, over the last decade, high-throughput sequencing (HTS) has become the primary means of identifying genetic variants and experimental designs have expanded to include case/control comparisons.

BPDs may be categorised into coagulation defects and platelet defects. Table 1 lists bleeding disorders due to defects in 21 genes that regulate the coagulation pathways and platelet disorders due to defects in 52 genes important for platelet formation, morphology or function. In this review, we do not describe these diseases in detail and instead refer the reader to other recent reviews of BPDs [1], [2], [3], [4].

As the catalogue of known BPD genes has expanded and the cost of HTS has decreased, it has become apparent that Sanger-based sequencing of specific genes suspected of harbouring pathogenic variants in particular patients is, in many situations, a slower, more expensive and less sensitive means of facilitating a molecular diagnosis than HTS of many genes at a time. Sanger-based approaches are particularly ill-suited for the diagnosis of diseases that are genetically heterogeneous but for which a molecular diagnosis would help assess the risk of specific comorbidities. For example, thrombocytopenia may have a concomitant increase in the risk of leukaemia and Hermansky-Pudlak syndrome may have a concomitant increase in the risk of lung fibrosis, depending on the specific variants responsible. In these cases, a molecular diagnosis is required in order to identify such increased risks of developing life-threatening conditions. Furthermore, assaying a large panel of genes during sequencing reduces the need for some of the painstaking assays typically required to narrow down the set of potential aetiologies needed to select appropriate genes for Sanger sequencing. In due course, genetic testing will not be postponed until specialised tests such as light transmission aggregometry have been performed, and this will simplify and hasten the overall diagnostic process [5].

Currently, HTS is used in both research and clinical settings, and there is a rapid translation of emerging findings into diagnostic practice. HTS can be applied to the whole or only to selected parts of the genome. Whole-genome sequencing (WGS) is the simplest approach and involves indiscriminate sequencing of DNA from lysed cells. The other approach is targeted sequencing of pre-specified regions of the genome. Whole-exome sequencing (WES) targets all or virtually all exons from all genes while panel-based sequencing targets only the regions of the genome relevant to a particular disease area and may include non-coding regions not covered by typical WES platforms. Panel-based assays are typically designed specifically so that clinical-grade accuracy can be achieved at virtually all genomic bases of interest, while WES platforms tend to cover a larger proportion of targeted regions poorly due to technical biases. Compared to WGS, all targeted methods require additional steps in the library preparation protocol and the design and purchase of assays for performing enrichment, typically microarrays. On the other hand, they reduce the amount of sequencing per patient required to achieve accurate variant calling in the parts of the genome thought to be relevant to a disease area. Thus, the advantage of targeted sequencing over WGS is purely economic. As sequencing costs continue to decline, targeted sequencing will foreseeably be replaced universally by WGS.

### **Platforms for targeted DNA sequencing**

The first approaches for targeted DNA sequencing of known BPD genes have recently been initiated by the ThromboGenomics and UK Genotyping and Phenotyping of Platelets (UK-GAPP) consortia. The ThromboGenomics consortium reported on a targeted HTS panel of 76 genes (in the published version) for the diagnosis of inherited thrombotic and BPDs [6] and has thus far processed over 2,000 patient samples (personal communication). Initial testing of about 300 patients showed 100% sensitivity to detect previously identified causal variants based on 159 patients. When the phenotype was strongly indicative of a particular

disease aetiology but the variants were unknown, diagnostic sensitivity was also high, at >90% based on 61 patients. However, for patients without a recognisable hereditary disorder (e.g. those with delta storage pool disease or bleeding with normal platelet and coagulation test results), a genetic diagnosis was only possible in 10% of cases. Such BPD patients should be included in research programmes, as we describe below. The published version of the ThromboGenomics platform has already undergone a revision because novel BPD genes have since been discovered through HTS-based discovery programmes (e.g. *DIAPH1*, *SRC*, *FYB*). However, clinical reporting on any new genes must await their approval for so-called “Tier 1” status. These are genes for which there is substantial evidence that variants therein are implicated in BPDs. Generally, causal variants need to have been identified in at least three unrelated pedigrees with similar phenotypes or in a very large pedigree with strong evidence from co-segregation studies, supported by experimental evidence from a mouse model. Formally, the decision to include a gene to the Tier 1 list is made by the Scientific and Standardization Committee (SSC) on “Genomics in Thrombosis and Haemostasis (GinTH)” of the International Society on Thrombosis and Haemostasis (ISTH).

The UK-GAPP study reported its first findings in 2012 using a targeted panel of 216 genes that included genes with hypothetical roles in BPDs [7]. More recently, the UK-GAPP used WES to analyse 329 potential candidate BPD genes in 18 unrelated index patients with Gi signaling or secretion defects [8]. Although the performance and details of other targeted platforms have not yet been published, one that includes 71 platelet-related genes has been developed by a Spanish team and used in a published case study [9], [10]. In addition, a team from Denmark is using WES with subsequent analysis of 87 genes related to bleeding, thrombocytopenia and thrombocytopathia, as presented at last year’s American Society of Hematology meeting [11]. Their approach evaluates several genes implicated in Ehlers Danlos syndrome (*COL3A1*, *COL5A2*, *COL5A1*) but does not include all coagulation genes. Lastly, a team in Germany has developed a platform to target the entire *F8* gene in order to detect deep intronic variants that are typically missed by WES [12], an approach also incorporated into the revised ThromboGenomics platform primarily to identify the recurrent intron 22 inversion.

### **Approaches to patient phenotyping**

Aggregating genetic and phenotypic patient data is essential to obtain the statistical power required to identify the genetic defects responsible for rare disorders. Data sharing is also important to enable a better understanding of the phenotypic heterogeneity of rare disorders. A standardised method of phenotyping is therefore key for these analyses and also for entering data into and interchanging data between clinical registries, genotype-phenotype databases and biobanks. The two most widely used clinical coding systems are ICD10 (the World Health Organisation’s International Classification of Diseases version 10, <http://www.who.int/classifications/icd>) and SNOMED-CT (Systematized Nomenclature of Medicine Clinical Terminology, <http://www.ihtsdo.org>). These systems are primarily aimed at codifying diagnoses rather than describing phenotypes and lack the flexibility required to describe novel rare inherited disorders. For example, they would not have been capable of capturing dominant

myelofibrosis with mild thrombocytopenia and bone defects, which we recently found to be caused by a germline gain-of-function variant in *SRC* [13]. Thus, ICD10 and SNOWMED-CT in their current forms are not well suited for rare disease research and diagnostics and their use could hamper data-sharing efforts. Instead, researchers and clinicians have opted for structured phenotype vocabularies called ontologies to describe rare diseases and their manifestations in particular patients. A phenotype ontology is a catalogue of specific signs, symptoms or laboratory findings seen in the clinic. Individual items (e.g. thrombocytopenia or abnormal APD aggregation) are referred to as terms with optional definitions, synonyms and translations. Additionally, ontologies provide hierarchical relations between the terms, which may range from the very specific to the general. While disease ontologies contain terms for known diseases such as Hermansky Pudlack syndrome (HPS), phenotype ontologies contain terms for individual manifestations of diseases, which in the case of HPS might include *Ocular albinism*, *Abnormal ATP secretion* and *Abnormal bleeding*. The Human Phenotype Ontology (HPO) was developed for the annotation of phenotypic abnormalities [14], [15] and is available under an open-source licence (<http://www.human-phenotype-ontology.org>). The general term that is most pertinent to BPDs is *Abnormality of blood and blood-forming tissues* and has the following HPO subclasses, among others: *Abnormality of thrombocytes*, *Abnormal bleeding*, *Abnormality of coagulation* and *Abnormal thrombosis* [15], [16]. Different types of phenotypic abnormalities are represented within these subclasses that are particularly relevant for BPDs, including morphological, cellular, physiological and laboratory-based functional abnormalities. Similarity values between terms and between sets of terms can be calculated, allowing computational comparison between patients, between patients and diseases and between patients and animal models, for example. We have used HPO to annotate approximately 1,000 cases with BPDs [17] and have shown how they can be automatically clustered and compared with annotations for diseases in the rare disease database Orphanet [18]. We have also used HPO-based statistical and computational methods to identify novel BPDs, including the above-mentioned syndrome caused by a gain-of-function variant in the universal tyrosine kinase *SRC* [13], and macrothrombocytopenia and deafness caused by gain-of-function variants in *DIAPH1* [19]. Further on, we describe how HPO can be used for variant prioritisation in a diagnostic setting.

### **Prioritisation of genetic variants**

A typical HTS experiment, even if limited to a few dozen genes, will typically result in hundreds or thousands of variant calls per individual, depending on the ancestry of the patient. The vast majority of pathogenic variants are kept at low allele frequency by negative selection, and therefore it is important to annotate variants by their population allele frequency. The largest and most widely used databases of variant allele frequencies include 1000 Genomes (2,504 individuals by low-depth WGS), UK10K (4,732 individuals by low-depth WES and 3,621 individuals by low-depth WGS) and ExAC (60,706 individuals by WES as part of various disease-specific and population genetic studies). These databases allow filtering or down-weighting of variants above particular allele frequency thresholds, but the vast majority of rare variants, even those absent entirely from these databases, are not relevant to rare disease. A consequence of this is that longer genes such as *VWF*

tend to harbour more rare variants than shorter genes such as *GP1BB* and this should be kept in mind when assessing a particular variant's relevance to a disease.

Variants also need to be annotated according to their predicted consequences, particularly with respect to gene transcripts (e.g., missense, frameshift, intronic, 5' untranslated region variant), and this is typically achieved using the Variant Effect Predictor (VEP) [20], which estimates consequences using data from Ensembl and other sources. Given that genes on Ensembl tend to have a large number of transcripts of which only some are relevant to a particular disease (e.g. those transcribed in the megakaryocyte for BPDs), it is important to take into account the specific transcripts against which each annotation is made. To this end, the ThromboGenomics gathered input from experts in different BPDs to select the relevant transcript for each gene.

It is also essential to annotate variants against databases of variants already implicated in rare diseases. Unfortunately, variants identified through rare disease research over the last few decades and published in scientific journals have not been systematically entered into publicly accessible databases, and this has allowed commercial curators to extract information from the literature and re-sell it in a structured form for a fee. The most widely used commercial database of published variants is the Human Gene Mutation Database (HGMD), which classifies variants according to a curator's assessment of the evidence presented in the relevant publications. Recently, the National Center for Biotechnology Information has released the online ClinVar database of variant interpretations for reported conditions [21], which accepts electronic submissions from clinical testing laboratories, research laboratories and other groups. ClinVar aggregates interpretations for a particular variant and condition into a single record, which is given an overall level of clinical significance. Due to the different sources for these two databases, the level of overlap is rather low and there is often disagreement in the purported pathogenicity between the two databases (Figure 1).

One of the main pitfalls of these databases is that they are contaminated with benign variants incorrectly labelled pathogenic. Sanger sequencing studies of single genes in the past typically checked whether putative disease-causing variants were absent from small groups of healthy controls, often containing as few as 100 individuals. For example, the variant in *MYH9* encoding p.S1114P is labelled in the HGMD database as a disease-causing mutation for Alport syndrome with macrothrombocytopenia based on a report from 2001 in which a large group of families suspected of having *MYH9*-related disorder were Sanger sequenced for variants in *MYH9* [22]. ClinVar reports this variant as having 'Conflicting interpretations of pathogenicity' as the variant has an allele frequency of 1 per 2,000 non-Finnish Europeans in ExAC. Moreover, recently the French *MYH9* network found this variant to be in linkage with another variant encoding p.D1424N, which had previously been shown to be pathogenic [23], suggesting p.S1114P was a red herring. Repeat submission to variant interpretation databases of incorrectly labelled pathogenic variants that are observed by chance in patients with a matching phenotype can result in a feedback loop that inflates confidence in variant pathogenicity. One of the tasks of the ClinGen initiative from the National Institutes

of Health is to organise arbitration on the clinical significance of variants with conflicting interpretations or with unknown significance [24]. However, arbitration can be compromised by a lack of reliable data acquired alongside the variant and phenotype submissions. For example, there is currently no mechanism in place to detect repeat submission of a variant from the same patient or of close relatives by different submitters and this can distort the evidence supporting an association between a variant and a particular phenotype. A possible method to tackle this problem would be to facilitate and encourage submission of a non-identifiable genetic fingerprint of each patient alongside the putative pathogenic variant. Such a fingerprint could be a set of genotypes at a limited set of common single nucleotide polymorphisms, or a projection of a standard set of genotypes onto a few principal components with pre-determined loadings. Disease-focused databases, such as the European Association for Haemophilia and Allied Disorders (EAHAD) coagulation factor variant databases for *VWF*, *F7*, *F8* and *F9*, tend to be more dependable than general databases as they are grounded on expert knowledge and specific measurements of coagulation factor levels, reducing the risk of erroneous entries. However, for most BPDs, such databases do not exist.

While the variant annotations described above are required for weighting or shortlisting variants for consideration, they cannot on their own provide a direct diagnostic answer because they need to be assessed in the context of a patient's phenotype. The ThromboGenomics consortium reported that after stringent variant filtering, there remained more than 5 candidate variants on average per patient. As we describe below, variant classification for reporting is typically done at this stage by a dedicated multidisciplinary team. However, this process can be supported by computational methods that compare the HPO-coded phenotype of the patient with the canonical HPO-coded phenotype of each disease linked to the genes that harbour the variants. The ThromboGenomics consortium coded each BPD with a set of characteristic HPO terms and then showed that in 85% of cases (93/109), the correct variant could be identified by automated comparison of HPO phenotypes. For example, a patient coded with *Decreased platelet glycoprotein Ib-IX-V*, *Increased mean platelet volume*, *Abnormal platelet morphology*, *Thrombocytopenia*, *Abnormality of leukocytes*, *Hearing impairment* and *Coronary artery disease* was sequenced and had candidate rare variants in *F5*, *PLA2G4A*, *VWF* and *GP1BB*. Despite atypical presence of *Hearing impairment* and the rather common phenotype *Coronary artery disease*, the homozygous variant in *GP1BB* was ranked highest because it was linked to Bernard-Soulier syndrome (BSS), which was coded with a similar phenotype, which comprised the terms *Abnormal bleeding*, *Decreased platelet glycoprotein Ib-IX-V*, *Impaired ristocetin-induced platelet aggregation*, *Increased mean platelet volume* and *Thrombocytopenia*.

### **Classifying and reporting variants**

The generation of clinical reports of variants identified by HTS for feedback to clinicians and patients should only be performed for variants affecting Tier 1 genes (established BPD genes). HTS-based platforms can adapt fairly easily to emerging evidence supporting the addition of new genes to the Tier 1 list. However, the implementation of HTS approaches in clinical genetic testing raises new challenges in variant interpretation. The average patient screened by the ThromboGenomics

platform has a mean of 5 candidate variants in BPD genes after filtering and therefore caution in the interpretation of these variants is important. A multidisciplinary team (MDT) comprising a geneticist, a clinician, a bioinformatician and a study coordinator produce the reports after classifying the variants. Guidelines for the interpretation and classification of gene variants have been suggested by the American College of Medical Genetics and Genomics and the Association for Molecular Pathology [25] and similar recommendations have been formulated by EuroGentest and the European Society of Human Genetics [26]. Guidelines propose that variants are classified using standard terminology: the terms ‘pathogenic’, ‘likely pathogenic’, ‘uncertain significance’, ‘likely benign’, and ‘benign’ should be used to describe variants identified in patients with rare hereditary disorders. These variants can have a ‘full contribution’ or ‘partial contribution’ to the phenotype. If a disease is recessive, a homozygous variant has a full contribution to the phenotype while compound heterozygous variants each have a partial contribution. In addition, this terminology can also be used when one variant cannot fully explain the entire clinical phenotype. For example, a patient with severe bleeding symptoms and mildly reduced FXI levels (by about 35%) received a report in which the heterozygous variant in *F11* had been assigned the label ‘pathogenic’ with a ‘partial contribution’ to the phenotype. This was because the specific variant had been shown previously to result in low FXI levels but it could not, on its own, cause severe bleeding. Thus, in this patient, a second unknown gene defect probably exists. For a variant to be considered ‘pathogenic’, it must have been described in at least three unrelated families with similar phenotypes. Variants can be labelled ‘likely pathogenic’ if they are found in a patient with a phenotype that is consistent with a defect in the gene it is found in and that variant has a low frequency (or is absent) in control databases (these databases are discussed in the next part). A ‘variant of uncertain significance’ (VUS) cannot be ruled out to play a role in disease but its consequence is too uncertain to label it ‘pathogenic’ or ‘likely pathogenic’ based on the information available to the MDT. It is nevertheless worthwhile to keep a record of VUS because as more controls and patients are sequenced, new evidence can help reclassify VUS as ‘likely pathogenic’ or ‘likely benign’. Information on the ethnicity of patients should also be recorded because variants can be common among particular ethnic groups but rare in the reference control databases, which tend to have a bias towards individuals with European ancestry. However, it is not advisable to report VUS or benign variants back to clinicians and patients.

### **Novel modes of inheritance**

Table 1 reports the modes of inheritance (MOI) for the different BPDs. It is becoming apparent that certain genes are involved in both recessive and dominant disorders. This was first shown in relation to *ITGA2B* and *ITGB3*, which was originally reported to harbour variants responsible for Glanzmann Thrombasthenia (GT), an autosomal recessive bleeding disorder characterised by failure of platelet aggregation with typically normal platelet count [27]. Specific variants in these genes were subsequently shown to lead to a dominant disorder characterised by macrothrombocytopenia and very mild platelet dysfunction [28], [29], [30], [31], [32], [33], [34]. Similarly, BSS is an autosomal recessive bleeding disorder caused by a deficiency of the platelet membrane von Willebrand factor receptor due to variants

in the *GP1BA*, *GP1BB* or *GP9* genes [35]. BSS patients have very large platelets with thrombocytopenia and platelet dysfunction. A dominant variant in *GP1BA* (referred to as the Bolzano variant [36]) results in mild macrothrombocytopenia while a similar platelet phenotype was recently described for patients with dominant variants in *GP1BB* [37]. A recent study of Gray Platelet Syndrome (GPS) patients with recessive variants in *NBEAL2* found milder platelet GPS-like phenotypes in the obligate carriers [38]. Both dominant-negative, high-impact variants in *GFI1B* located in the region coding the zinc fingers [39], [40] and hypomorphic biallelic *GFI1B* variants result in thrombocytopenia with reduced numbers of platelet granules [41]. A very similar finding was made for Paris-Trousseau thrombocytopenia, which is mostly the result of a *FLI1* deletion (11q deletion) or high impact variants on *FLI1*, but it was recently shown that homozygous or compound heterozygous missense variants in the DNA-binding domain of *FLI1* can also cause this disease [42]. The discovery of novel MOI for established BPD genes raises the question of how best to classify diseases in relation to genes, a topic of interest for the SSC on GinTH.

In addition to dominant versus recessive MOI, there is emerging evidence that combinations of variants in two different BPD genes can result in phenotypes that are absent or much less severe when only one of the variants is present. Variants in *P2RY12* can result in a recessive bleeding disorder characterised by a platelet ADP aggregation defect while heterozygous carriers are not affected [43]. Interestingly, two studies have now reported a heterozygous variant in *P2RY12* in a patient with an aggregation defect and bleeding history much more severe than found for other obligate carriers [44], [45]. In these cases, an additional defect would be expected and indeed this was shown to be the case in one of these patients, who had a second pathogenic variant in *VWF* [44]. Another example of possible digenic inheritance was described recently in relation to a pedigree in which members with haemophilia A bleed more severely if they also carry a common variant in *PTGS1* (ExAC allele frequency of 14% in Africans) than if they do not [46]. The difficulty with generating a conclusive diagnosis in cases with digenic inheritance is the sparsity of cases with similar combinatorial defects to compare against, but it is nevertheless a possibility that may need to be considered during deliberation by MDTs.

### **Whole-genome DNA sequencing**

The major challenge ahead will be the analysis of the regulatory regions of the genome. Currently, there is a scarcity of large-scale population allele frequency databases spanning the entire genome. Recently, a group from Human Longevity Inc reported their results from analysing 10,000 deeply sequenced whole genomes [47] but the variant calls are available only through a restricted browser-based search function. The preliminary release of the gnomAD database (126,216 individuals by WES and 15,136 by WGS, but including individuals in ExAC) can at present only be accessed using a web browser but it will eventually be made available for download. Once population allele frequencies derived from large numbers of individuals become available, it will remain challenging to assess pathogenicity of non-exonic rare variants. First, they are far more numerous than exonic rare variants because exons cover only 1/50th of the genome and because



non-exonic regions tend to be less conserved than exonic regions and thus tend to harbour a higher density of rare variants. Second, as they do not alter protein sequence, it is often unclear which regulatory pathway they disrupt, if any. To tackle these difficulties, it will be necessary, initially, to focus attention on carefully selected segments of the non-exonic space, such as those believed to play important roles in gene expression regulation, including promoter and enhancer regions.

Within the field of BPDs specifically, various strategies could be followed to identify such regions. They could be identified using the ENCODE database of active regulatory regions and the binding sites of relevant transcription factors (e.g. GATA1, FLI1, ETS1, NFE2 and GATA2) in the megakaryocytic cell line K562. Chromatin immunoprecipitation (ChIP) sequencing data from primary megakaryocytes are also available for GATA1, GATA2, RUNX1, FLI1, and SCL [48]. A large genome-wide association study of 173,480 European-ancestry participants recently identified hundreds of low frequency (<5%) and rare (<1%) variants associated with full blood count parameters, including platelet volume, count, crit (total platelet mass) and distribution width [49]. Many of these variants are located in non-coding regions that are likely to play a role in the formation or clearance of platelets and, as such, they are regions of potential interest in the context of BPD patients. The same applies to variants associated with gene expression of nearby genes (expression quantitative trait loci (eQTL)) in platelets. Studies of the epigenomic landscape of the megakaryocyte have recently been conducted, revealing areas that are likely to be enhancers, for example (Petersen *et al.*, under review).

How large is the non-exonic space that is potentially of relevance to BPDs? To try to answer this question, we have combined information from genetic association studies, studies of transcription factor binding sites and epigenomic studies to identify non-exonic regions that may play a role in the aetiology of BPDs. Each base in the genome was labelled according to whether a particular signature was switched on. These signatures indicate features such as proximity to a locus in population association with platelet parameters measured by a haematology analyser, or presence of a peak for a particular transcription factor as measured by ChIP sequencing in megakaryocytes, for example. In total, we identified 54Mb of non-exonic regions with a regulatory signature suggestive of a functional role in megakaryocytes. This is comparable to the exonic space across all protein-coding genes. Most of the non-exonic segments are regions of open chromatin that are also marked as enhancers, promoters or binding sites for the structural protein CTCF (Figure 2, bottom panel). In contrast, the regions with the densest signatures, i.e. those in which many indicators of functional relevance are present, tend to cover a very small fraction of the genome (Figure 2, top panel). One of the densest signatures has all features except association with platelet crit, presence of a nearby eQTL and enhancer (row five in Figure 2) and is found solely in the vicinity of *GP1BB*. Interestingly, a BPD-causing non-coding variant has been reported in a BSS patient with a missense variant in *GP1BB* and another variant in the opposite haplotype of the *GP1BB* promoter that resulted in reduced GATA1-binding [50]. As our knowledge of the regulatory role of non-coding regions of the genome and the

impact of rare variants therein becomes clearer, it will become feasible to incorporate such variants in the diagnostic work-up.

There are also technical advantages to using WGS that will improve diagnostic yield as WGS becomes commonplace. In general, WGS is a more sensitive method than targeted sequencing for detecting structural variants such as large deletions, insertions, inversions, tandem repeat expansions and complex rearrangements. It may be that a certain fraction of undiagnosed patients have variants that affect known genes but in ways that require WGS to be adequately detected. Finally, although current sequencing read lengths are typically 100-150bp, emerging long-read sequencing technology is capable of generating reads tens of thousands of bp long. In addition to improving power to detect structural variants, long reads facilitate phasing of rare variants affecting the same region. For recessive diseases, this will reduce the need for parental samples to ascertain whether rare variants are in trans in the offspring.

### **Concluding remarks**

Intensive research activity in rare diseases continues to reveal new genes and regulatory regions of potential importance as well as novel modes of inheritance for established genes. As the evidence supporting new aetiologies accumulates and new findings gain Tier 1 status, diagnostic practice needs to react to these developments. The most advanced assays in routine use today target a relatively small, pre-determined part of the human genome, but can be re-versioned regularly. Over the next few years, however, WGS will replace targeted platforms universally. The complexity and size of the sequencing data, the phenotypic data and the breadth of potential aetiologies that must be accounted for, along with the multiple layers of genomic information needed to interpret the possible effects of non-coding variants, require changes in the diagnostic process. Multidisciplinary teams composed of clinicians, geneticists and bioinformaticians need to work together to provide a rapid yet reliable service that utilises all the evidence available effectively. The careful aggregation of clinical and genetic data from across the globe in freely accessible databases will prove essential to properly assess the ever-changing evidence for pathogenicity of particular variants. We are amidst a transition of clinical genetics into a globally networked, computer-aided and multidisciplinary discipline that will increase diagnostic yield and improve patient care substantially.

### **Figure legends**

Figure 1: Venn diagram illustrating the degree of overlap between pathogenic (DM) or likely pathogenic (DM?) variants in HGMD and variants in ClinVar. The overlap is broken down into four categories, according to whether variants are labelled pathogenic in ClinVar, have conflicting interpretations, are labelled benign, or lack any interpretations. We thank Steven Harrison for kindly providing the data for this figure, which are accurate as of August 2016.

Figure 2: Each column in the grids represents a regulatory feature, which may be part of a regulatory signature (blue) or absent from it (grey) at any location in the genome. PCT, PDW, PLT, MPV: proximity of a variant in population association with platelet crit, platelet distribution width, platelet count or mean platelet volume,

respectively; eQTL: proximity of a variant associated with expression of a nearby gene in platelets (data not published); GATA1, GATA2, FLI1, RUNX1, SCL, MEIS1: transcription factor binding sites in MKs as identified by ChIP sequencing; CTCF: binding site for the transcription repressor CTCF in MKs; Enhancer, Promoter: region assigned as having enhancer or promoter activity, respectively, in Petersen et al.; ATAC: regions of accessible chromatin identified using Assay for Transposase-Accessible Chromatin sequencing of DNA from MKs (data not published). The top grid and associated bar plot show the signatures with the highest number of regulatory features and the number of bases in the genome in which those signatures are present. The bottom grid and associated bar plot show the signatures which cover the largest fractions of the genome.

### Acknowledgements

We thank Steven Harrison for sharing a table comparing variants in HGMD and ClinVar. We thank Mattia Frontini for providing early access to the data on open chromatin in MKs. We thank Michel Georges for early access to the data on eQTLs in platelets. We acknowledge the support of the NIHR BioResource—Rare Diseases.

### Table legends

Table 1: Table of BPDs with known implicated genes, as of January 2017.

### References

- 1 Lentaigne C, Freson K, Laffan MA, Turro E, Ouwehand WH, Consortium B-B, the ThromboGenomics C. Inherited platelet disorders: toward DNA-based diagnosis. *Blood*. 2016; **127**: 2814-23. 10.1182/blood-2016-03-378588.
- 2 Gresele P, Bury L, Falcinelli E. Inherited Platelet Function Disorders: Algorithms for Phenotypic and Genetic Investigation. *Semin Thromb Hemost*. 2016; **42**: 292-305. 10.1055/s-0035-1570078.
- 3 Lambert MP. Update on the inherited platelet disorders. *Curr Opin Hematol*. 2015; **22**: 460-6. 10.1097/MOH.000000000000171.
- 4 Savoia A. Molecular basis of inherited thrombocytopenias. *Clin Genet*. 2016; **89**: 154-62. 10.1111/cge.12607.
- 5 Gresele P, Subcommittee on Platelet Physiology of the International Society on T, Hemostasis. Diagnosis of inherited platelet function disorders: guidance from the SSC of the ISTH. *Journal of thrombosis and haemostasis : JTH*. 2015; **13**: 314-22. 10.1111/jth.12792.
- 6 Simeoni I, Stephens JC, Hu F, Deevi SV, Megy K, Bariana TK, Lentaigne C, Schulman S, Sivapalaratnam S, Vries MJ, Westbury SK, Greene D, Papadia S, Alessi MC, Attwood AP, Ballmaier M, Baynam G, Bermejo E, Bertoli M, Bray PF, Bury L, Cattaneo M, Collins P, Daugherty LC, Favier R, French DL, Furie B, Gattens M, Germeshausen M, Ghevaert C, Goodeve AC, Guerrero JA, Hampshire DJ, Hart DP, Heemskerk JW, Henskens YM, Hill M, Hogg N, Jolley JD, Kahr WH, Kelly AM, Kerr R, Kostadima M, Kunishima S, Lambert MP, Liesner R, Lopez JA, Mapeta RP, Mathias M, Millar CM, Nathwani A, Neerman-Arbez M, Nurden AT, Nurden P, Othman M, Peerlinck K, Perry DJ, Poudel P, Reitsma P, Rondina MT, Smethurst PA, Stevenson W, Szkotak A, Tuna S, van Geet C, Whitehorn D, Wilcox DA, Zhang B, Revel-Vilk S, Gresele P, Bellissimo DB, Penkett CJ, Laffan MA, Mumford AD, Rendon A, Gomez K, Freson K, Ouwehand WH, Turro E. A high-throughput sequencing test for

diagnosing inherited bleeding, thrombotic, and platelet disorders. *Blood*. 2016; **127**: 2791-803. 10.1182/blood-2015-12-688267.

7 Jones ML, Murden SL, Bem D, Mundell SJ, Gissen P, Daly ME, Watson SP, Mumford AD, group UGs. Rapid genetic diagnosis of heritable platelet function disorders with next-generation sequencing: proof-of-principle with Hermansky-Pudlak syndrome. *Journal of thrombosis and haemostasis : JTH*. 2012; **10**: 306-9. 10.1111/j.1538-7836.2011.04569.x.

8 Leo VC, Morgan NV, Bem D, Jones ML, Lowe GC, Lordkipanidze M, Drake S, Simpson MA, Gissen P, Mumford A, Watson SP, Daly ME, Group UGS. Use of next-generation sequencing and candidate gene analysis to identify underlying defects in patients with inherited platelet function disorders. *Journal of thrombosis and haemostasis : JTH*. 2015; **13**: 643-50. 10.1111/jth.12836.

9 Bastida JM, Del Rey M, Revilla N, Benito R, Perez-Andres M, Gonzalez B, Riesco S, Janusz K, Padilla J, Benito-Sendin AH, Bueno D, Blanco E, Hernandez-Rivas M, Vicente V, Rivera J, Gonzalez-Porrás R, Lozano ML. Wiskott-Aldrich syndrome in a child presenting with macrothrombocytopenia. *Platelets*. 2016: 1-4. 10.1080/09537104.2016.1246715.

10 Lozano ML, Cook A, Bastida JM, Paul DS, Iruin G, Cid AR, Adan-Pedroso R, Ramon Gonzalez-Porrás J, Hernandez-Rivas JM, Fletcher SJ, Johnson B, Morgan N, Ferrer-Marin F, Vicente V, Sondek J, Watson SP, Bergmeier W, Rivera J. Novel mutations in RASGRP2, which encodes CalDAG-GEFI, abrogate Rap1 activation, causing platelet dysfunction. *Blood*. 2016; **128**: 1282-9. 10.1182/blood-2015-11-683102.

11 Leino EB, Zetterberg E, Oestrup O, Kampmann P, Eva N, Andersson NG, Kinalis S, Qvortrup K, Rossing CM. Targeted Whole Exome Sequencing of 87 Predisposition Genes in 156 Patients from the Oresund Region with Bleeding Disorders. *58th American Society of Hematology Annual Meeting*. 2016.

12 Bach JE, Oldenburg J, Muller CR, Rost S. Mutational spectrum and deep intronic variants in the factor VIII gene of haemophilia A patients. Identification by next generation sequencing. *Haemostaseologie*. 2016; **36**: S25-S8.

13 Turro E, Greene D, Wijgaerts A, Thys C, Lentaigne C, Bariana TK, Westbury SK, Kelly AM, Selleslag D, Stephens JC, Papadia S, Simeoni I, Penkett CJ, Ashford S, Attwood A, Austin S, Bakchoul T, Collins P, Deevi SV, Favier R, Kostadima M, Lambert MP, Mathias M, Millar CM, Peerlinck K, Perry DJ, Schulman S, Whitehorn D, Wittevrongel C, Consortium B-B, De Maeyer M, Rendon A, Gomez K, Erber WN, Mumford AD, Nurden P, Stirrups K, Bradley JR, Raymond FL, Laffan MA, Van Geet C, Richardson S, Freson K, Ouwehand WH. A dominant gain-of-function mutation in universal tyrosine kinase SRC causes thrombocytopenia, myelofibrosis, bleeding, and bone pathologies. *Sci Transl Med*. 2016; **8**: 328ra30. 10.1126/scitranslmed.aad7666.

14 Robinson PN, Mundlos S. The human phenotype ontology. *Clin Genet*. 2010; **77**: 525-34. 10.1111/j.1399-0004.2010.01436.x.

15 Kohler S, Vasilevsky NA, Engelstad M, Foster E, McMurry J, Ayme S, Baynam G, Bello SM, Boerkoel CF, Boycott KM, Brudno M, Buske OJ, Chinnery PF, Cipriani V, Connell LE, Dawkins HJ, DeMare LE, Devereau AD, de Vries BB, Firth HV, Freson K, Greene D, Hamosh A, Helbig I, Hum C, Jahn JA, James R, Krause R, SJ FL, Lochmuller H, Lyon GJ, Ogishima S, Olry A, Ouwehand WH, Pontikos N, Rath A, Schaefer F, Scott RH, Segal M, Sergouniotis PI, Sever R, Smith CL, Straub V, Thompson R, Turner C, Turro E, Veltman MW, Vulliamy T, Yu J, von Ziegenweid J, Zankl A, Zuchner S, Zemojtel T, Jacobsen JO, Groza T, Smedley D, Mungall CJ, Haendel M, Robinson PN. The Human Phenotype Ontology in 2017. *Nucleic Acids Res*. 2017; **45**: D865-D76. 10.1093/nar/gkw1039.

- 16 Kohler S, Doelken SC, Mungall CJ, Bauer S, Firth HV, Bailleul-Forestier I, Black GC, Brown DL, Brudno M, Campbell J, FitzPatrick DR, Eppig JT, Jackson AP, Freson K, Girdea M, Helbig I, Hurst JA, Jahn J, Jackson LG, Kelly AM, Ledbetter DH, Mansour S, Martin CL, Moss C, Mumford A, Ouwehand WH, Park SM, Riggs ER, Scott RH, Sisodiya S, Van Vooren S, Wapner RJ, Wilkie AO, Wright CF, Vulto-van Silfhout AT, de Leeuw N, de Vries BB, Washington NL, Smith CL, Westerfield M, Schofield P, Ruef BJ, Gkoutos GV, Haendel M, Smedley D, Lewis SE, Robinson PN. The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. *Nucleic Acids Res.* 2014; **42**: D966-74. 10.1093/nar/gkt1026.
- 17 Westbury SK, Turro E, Greene D, Lentaigne C, Kelly AM, Bariana TK, Simeoni I, Pillois X, Attwood A, Austin S, Jansen SB, Bakchoul T, Crisp-Hihn A, Erber WN, Favier R, Foad N, Gattens M, Jolley JD, Liesner R, Meacham S, Millar CM, Nurden AT, Peerlinck K, Perry DJ, Poudel P, Schulman S, Schulze H, Stephens JC, Furie B, Robinson PN, van Geet C, Rendon A, Gomez K, Laffan MA, Lambert MP, Nurden P, Ouwehand WH, Richardson S, Mumford AD, Freson K, Consortium B-B. Human phenotype ontology annotation and cluster analysis to unravel genetic defects in 707 cases with unexplained bleeding and platelet disorders. *Genome Med.* 2015; **7**: 36. 10.1186/s13073-015-0151-5.
- 18 Rath A, Olry A, Dhombres F, Brandt MM, Urbero B, Ayme S. Representation of rare diseases in health information systems: the Orphanet approach to serve a wide range of end users. *Human mutation.* 2012; **33**: 803-8. 10.1002/humu.22078.
- 19 Stritt S, Nurden P, Turro E, Greene D, Jansen SB, Westbury SK, Petersen R, Astle WJ, Marlin S, Bariana TK, Kostadima M, Lentaigne C, Maiwald S, Papadia S, Kelly AM, Stephens JC, Penkett CJ, Ashford S, Tuna S, Austin S, Bakchoul T, Collins P, Favier R, Lambert MP, Mathias M, Millar CM, Mapeta R, Perry DJ, Schulman S, Simeoni I, Thys C, Consortium B-B, Gomez K, Erber WN, Stirrups K, Rendon A, Bradley JR, van Geet C, Raymond FL, Laffan MA, Nurden AT, Nieswandt B, Richardson S, Freson K, Ouwehand WH, Mumford AD. A gain-of-function variant in DIAPH1 causes dominant macrothrombocytopenia and hearing loss. *Blood.* 2016; **127**: 2903-14. 10.1182/blood-2015-10-675629.
- 20 McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, Cunningham F. The Ensembl Variant Effect Predictor. *Genome biology.* 2016; **17**: 122. 10.1186/s13059-016-0974-4.
- 21 Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Hoover J, Jang W, Katz K, Ovetsky M, Riley G, Sethi A, Tully R, Villamarin-Salomon R, Rubinstein W, Maglott DR. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* 2016; **44**: D862-8. 10.1093/nar/gkv1222.
- 22 Zhang S, Zhou X, Liu S, Bai T, Zhang Y, Wang J, Wang S, Zhang X, Wang B. MYH9-related disease: description of a large Chinese pedigree and a survey of reported mutations. *Acta Haematol.* 2014; **132**: 193-8. 10.1159/000356681.
- 23 Saposnik B, Binard S, Fenneteau O, Nurden A, Nurden P, Hurtaud-Roux MF, Schlegel N, French MYHn. Mutation spectrum and genotype-phenotype correlations in a large French cohort of MYH9-Related Disorders. *Molecular genetics & genomic medicine.* 2014; **2**: 297-312. 10.1002/mgg3.68.
- 24 Rehm HL, Berg JS, Brooks LD, Bustamante CD, Evans JP, Landrum MJ, Ledbetter DH, Maglott DR, Martin CL, Nussbaum RL, Plon SE, Ramos EM, Sherry ST, Watson MS, ClinGen. ClinGen--the Clinical Genome Resource. *N Engl J Med.* 2015; **372**: 2235-42. 10.1056/NEJMSr1406261.

- 25 Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL, Committee ALQA. 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**: 405-24. 10.1038/gim.2015.30.
- 26 Matthijs G, Souche E, Alders M, Corveleyn A, Eck S, Feenstra I, Race V, Sistermans E, Sturm M, Weiss M, Yntema H, Bakker E, Scheffer H, Bauer P, EuroGentest, European Society of Human G. Guidelines for diagnostic next-generation sequencing. *Eur J Hum Genet*. 2016; **24**: 2-5. 10.1038/ejhg.2015.226.
- 27 Rosenberg N, Yatuv R, Orion Y, Zivelin A, Dardik R, Peretz H, Seligsohn U. Glanzmann thrombasthenia caused by an 11.2-kb deletion in the glycoprotein IIIa (beta3) is a second mutation in Iraqi Jews that stemmed from a distinct founder. *Blood*. 1997; **89**: 3654-62.
- 28 Kunishima S, Kashiwagi H, Otsu M, Takayama N, Eto K, Onodera M, Miyajima Y, Takamatsu Y, Suzumiya J, Matsubara K, Tomiyama Y, Saito H. Heterozygous ITGA2B R995W mutation inducing constitutive activation of the alphaIIb beta3 receptor affects proplatelet formation and causes congenital macrothrombocytopenia. *Blood*. 2011; **117**: 5479-84. 10.1182/blood-2010-12-323691.
- 29 Nurden AT, Pillois X, Fiore M, Heilig R, Nurden P. Glanzmann thrombasthenia-like syndromes associated with Macrothrombocytopenias and mutations in the genes encoding the alphaIIb beta3 integrin. *Semin Thromb Hemost*. 2011; **37**: 698-706. 10.1055/s-0031-1291380.
- 30 Ghevaert C, Salsmann A, Watkins NA, Schaffner-Reckinger E, Rankin A, Garner SF, Stephens J, Smith GA, Debili N, Vainchenker W, de Groot PG, Huntington JA, Laffan M, Kieffer N, Ouwehand WH. A nonsynonymous SNP in the ITGB3 gene disrupts the conserved membrane-proximal cytoplasmic salt bridge in the alphaIIb beta3 integrin and cosegregates dominantly with abnormal proplatelet formation and macrothrombocytopenia. *Blood*. 2008; **111**: 3407-14. 10.1182/blood-2007-09-112615.
- 31 Vanhoorelbeke K, De Meyer SF, Pareyn I, Melchior C, Plancon S, Margue C, Pradier O, Fondu P, Kieffer N, Springer TA, Deckmyn H. The novel S527F mutation in the integrin beta3 chain induces a high affinity alphaIIb beta3 receptor by hindering adoption of the bent conformation. *The Journal of biological chemistry*. 2009; **284**: 14914-20. 10.1074/jbc.M809167200.
- 32 Gresele P, Falcinelli E, Giannini S, D'Adamo P, D'Eustacchio A, Corazzi T, Mezzasoma AM, Di Bari F, Guglielmini G, Cecchetti L, Noris P, Balduini CL, Savoia A. Dominant inheritance of a novel integrin beta3 mutation associated with a hereditary macrothrombocytopenia and platelet dysfunction in two Italian families. *Haematologica*. 2009; **94**: 663-9. 10.3324/haematol.2008.002246.
- 33 Jayo A, Conde I, Lastres P, Martinez C, Rivera J, Vicente V, Gonzalez-Manchon C. L718P mutation in the membrane-proximal cytoplasmic tail of beta 3 promotes abnormal alpha IIb beta 3 clustering and lipid microdomain coalescence, and associates with a thrombasthenia-like phenotype. *Haematologica*. 2010; **95**: 1158-66. 10.3324/haematol.2009.018572.
- 34 Kashiwagi H, Kunishima S, Kiyomizu K, Amano Y, Shimada H, Morishita M, Kanakura Y, Tomiyama Y. Demonstration of novel gain-of-function mutations of alphaIIb beta3: association with macrothrombocytopenia and glanzmann thrombasthenia-like phenotype. *Molecular genetics & genomic medicine*. 2013; **1**: 77-86. 10.1002/mgg3.9.

- 35 Lopez JA, Andrews RK, Afshar-Kharghan V, Berndt MC. Bernard-Soulier syndrome. *Blood*. 1998; **91**: 4397-418.
- 36 De Marco L, Mazzucato M, Fabris F, De Roia D, Coser P, Girolami A, Vicente V, Ruggeri ZM. Variant Bernard-Soulier syndrome type bolzano. A congenital bleeding disorder due to a structural and functional abnormality of the platelet glycoprotein Ib-IX complex. *J Clin Invest*. 1990; **86**: 25-31. 10.1172/JCI114692.
- 37 Sivapalaratnam S, Westbury SK, Stephens JC, Greene D, Downes K, Kelly AM, Lentaigne C, Astle WJ, Huizinga EG, Nurden P, Papadia S, Peerlinck K, Penkett CJ, Perry DJ, Roughley C, Simeoni I, Stirrups K, Hart DP, Tait RC, Mumford AD, BioResource N, Laffan MA, Freson K, Ouwehand WH, Kunishima S, Turro E. Rare variants in GP1BB are responsible for autosomal dominant macrothrombocytopenia. *Blood*. 2016. 10.1182/blood-2016-08-732248.
- 38 Bottega R, Pecci A, De Candia E, Pujol-Moix N, Heller PG, Noris P, De Rocco D, Podda GM, Glembotsky AC, Cattaneo M, Balduini CL, Savoia A. Correlation between platelet phenotype and NBEAL2 genotype in patients with congenital thrombocytopenia and alpha-granule deficiency. *Haematologica*. 2013; **98**: 868-74. 10.3324/haematol.2012.075861.
- 39 Monteferrario D, Bolar NA, Marneth AE, Hebeda KM, Bergevoet SM, Veenstra H, Laros-van Gorkom BA, MacKenzie MA, Khandanpour C, Botezatu L, Fransen E, Van Camp G, Duijnhouwer AL, Salemink S, Willemsen B, Huls G, Preijers F, Van Heerde W, Jansen JH, Kempers MJ, Loeys BL, Van Laer L, Van der Reijden BA. A dominant-negative GFI1B mutation in the gray platelet syndrome. *N Engl J Med*. 2014; **370**: 245-53. 10.1056/NEJMoa1308130.
- 40 Kitamura K, Okuno Y, Yoshida K, Sanada M, Shiraishi Y, Muramatsu H, Kobayashi R, Furukawa K, Miyano S, Kojima S, Ogawa S, Kunishima S. Functional characterization of a novel GFI1B mutation causing congenital macrothrombocytopenia. *Journal of thrombosis and haemostasis : JTH*. 2016; **14**: 1462-9. 10.1111/jth.13350.
- 41 Ferreira CR, Chen D, Abraham SM, Adams DR, Simon KL, Malicdan MC, Markello TC, Gunay-Aygun M, Gahl WA. Combined alpha-delta platelet storage pool deficiency is associated with mutations in GFI1B. *Mol Genet Metab*. 2016. 10.1016/j.ymgme.2016.12.006.
- 42 Stevenson WS, Rabbolini DJ, Beutler L, Chen Q, Gabrielli S, Mackay JP, Brighton TA, Ward CM, Morel-Kopp MC. Paris-Trousseau thrombocytopenia is phenocopied by the autosomal recessive inheritance of a DNA-binding domain mutation in FLI1. *Blood*. 2015; **126**: 2027-30. 10.1182/blood-2015-06-650887.
- 43 Cattaneo M, Lecchi A, Randi AM, McGregor JL, Mannucci PM. Identification of a new congenital defect of platelet function characterized by severe impairment of platelet responses to adenosine diphosphate. *Blood*. 1992; **80**: 2787-96.
- 44 Stockley J, Nisar SP, Leo VC, Sabi E, Cunningham MR, Eikenboom JC, Lethagen S, Schneppenheim R, Goodeve AC, Watson SP, Mundell SJ, Daly ME, Group GSiCwtM-VS. Identification and Characterization of Novel Variations in Platelet G-Protein Coupled Receptor (GPCR) Genes in Patients Historically Diagnosed with Type 1 von Willebrand Disease. *PLoS One*. 2015; **10**: e0143913. 10.1371/journal.pone.0143913.
- 45 Remijn JA, MJ IJ, Strunk AL, Abbes AP, Engel H, Dikkeschei B, Dompeling EC, de Groot PG, Slingerland RJ. Novel molecular defect in the platelet ADP receptor P2Y12 of a patient with haemorrhagic diathesis. *Clin Chem Lab Med*. 2007; **45**: 187-9. 10.1515/CCLM.2007.036.
- 46 Nance D, Campbell RA, Rowley JW, Downie JM, Jorde LB, Kahr WH, Mereby SA, Tolley ND, Zimmerman GA, Weyrich AS, Rondina MT. Combined variants in factor VIII and prostaglandin synthase-1 amplify hemorrhage severity across three generations of

descendants. *Journal of thrombosis and haemostasis : JTH*. 2016; **14**: 2230-40. 10.1111/jth.13500.

47 Telenti A, Pierce LC, Biggs WH, di Iulio J, Wong EH, Fabani MM, Kirkness EF, Moustafa A, Shah N, Xie C, Brewerton SC, Bulsara N, Garner C, Metzker G, Sandoval E, Perkins BA, Och FJ, Turpaz Y, Venter JC. Deep sequencing of 10,000 human genomes. *Proc Natl Acad Sci U S A*. 2016; **113**: 11901-6. 10.1073/pnas.1613365113.

48 Tijssen MR, Cvejic A, Joshi A, Hannah RL, Ferreira R, Forrai A, Bellissimo DC, Oram SH, Smethurst PA, Wilson NK, Wang X, Ottersbach K, Stemple DL, Green AR, Ouwehand WH, Gottgens B. Genome-wide analysis of simultaneous GATA1/2, RUNX1, FLI1, and SCL binding in megakaryocytes identifies hematopoietic regulators. *Dev Cell*. 2011; **20**: 597-609. 10.1016/j.devcel.2011.04.008.

49 Astle WJ, Elding H, Jiang T, Allen D, Ruklisa D, Mann AL, Mead D, Bouman H, Riveros-Mckay F, Kostadima MA, Lambourne JJ, Sivapalaratnam S, Downes K, Kundu K, Bomba L, Berentsen K, Bradley JR, Daugherty LC, Delaneau O, Freson K, Garner SF, Grassi L, Guerrero J, Haimel M, Janssen-Megens EM, Kaan A, Kamat M, Kim B, Mandoli A, Marchini J, Martens JH, Meacham S, Megy K, O'Connell J, Petersen R, Sharifi N, Sheard SM, Staley JR, Tuna S, van der Ent M, Walter K, Wang SY, Wheeler E, Wilder SP, Iotchkova V, Moore C, Sambrook J, Stunnenberg HG, Di Angelantonio E, Kaptoge S, Kuipers TW, Carrillo-de-Santa-Pau E, Juan D, Rico D, Valencia A, Chen L, Ge B, Vasquez L, Kwan T, Garrido-Martin D, Watt S, Yang Y, Guigo R, Beck S, Paul DS, Pastinen T, Bujold D, Bourque G, Frontini M, Danesh J, Roberts DJ, Ouwehand WH, Butterworth AS, Soranzo N. The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell*. 2016; **167**: 1415-29 e19. 10.1016/j.cell.2016.10.042.

50 Ludlow LB, Schick BP, Budarf ML, Driscoll DA, Zackai EH, Cohen A, Konkle BA. Identification of a mutation in a GATA binding site of the platelet glycoprotein Ibbeta promoter resulting in the Bernard-Soulier syndrome. *The Journal of biological chemistry*. 1996; **271**: 22076-80.