Suthesh Sivapalaratnam

Willem H Ouwehand

ThromboGenomics: comprehensive high-throughput sequencing test for the diagnosis of inherited bleeding, thrombotic and platelet disorders

In this article the team from the ThromboGenomics consortium describe the use of high throughput sequencing to identify rare variants in genes encoding proteins involved in disorders of haemostasis. This can provide a definite genetic basis for patients with platelet functional disorders of uncertain aetiology. What is particularly impressive is the speed with which this technology has been translated from the research environment to the clinic.

Background

The human genome comprises 3.2 billion base pairs, of which only 65 million bases (2%) are protein-coding regions, collectively called the exome. In 2009 the reading of the coding fraction by whole exome sequencing (WES) had become technically feasible. Since, at least 100,000 human DNA samples have been analysed by WES. With the reduction of the cost of high throughput sequencing (HTS), reading of the entire genome by whole genome sequencing (WGS) is also increasingly being applied and an estimated 15,000 DNA samples have been analysed thus far. One of the important results from these large-scale sequencing efforts is the compilation of a progressively more complete and accurate catalogue of human genetic variation. The most common changes identified are single nucleotide variants (SNVs), of which so far 150 million have been identified, with the vast majority being rare ie with a minor allele frequency (MAF) below 0.1% . The principal two other forms of variation are copy number variants (CNVs) and structural variants (SVs). These are much less frequent; one CNV occurs for each 10 SNVs, and the current tally for SVs stands almost at 4.4 million.

In 2013 the 100,000 Genomes Project was launched, with the primary aim being to analyse by WGS 100,000 DNA samples of NHS patients. The key objectives are to improve the molecular classification of cancers, to support the detection and classification of infectious agents and to reduce the diagnostic delay for rare genetic diseases. There are an estimated 7,000 rare diseases in humans and 1 in 20 individuals are affected by a genetic rare condition. A large number of them will experience severe morbidity early in life, resulting in early mortality for many. The causal genes for less than half of these rare diseases have been identified. The 100,000 Genomes Project will lead to the identification of a large number of new variants and novel genes underlying rare diseases.

Inherited Bleeding, Thrombotic and Platelet Disorders

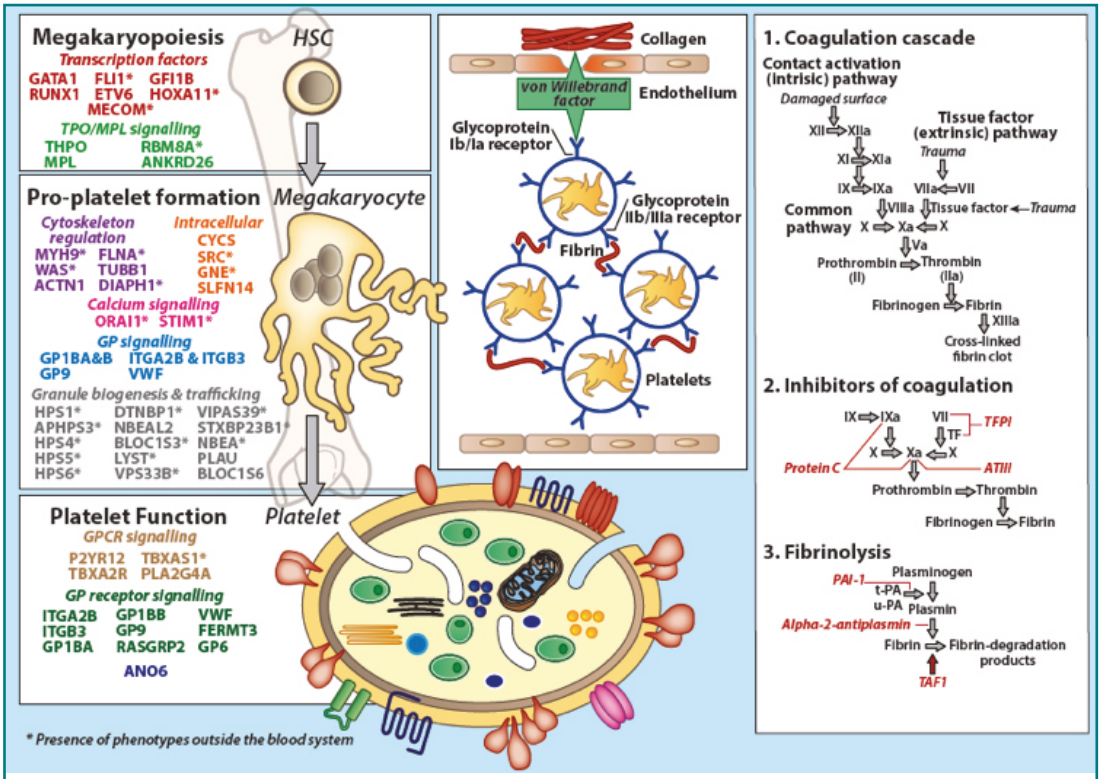
Inherited bleeding, thrombosis and platelet disorders (BPDs) are a category of rare diseases which have been well researched at the molecular level (Figure 1), affecting approximately 300 individuals per million births. With the exceptions of haemophilia A and B, and von Willebrand disease (VWD), an easily accessible molecular diagnosis for BPD patients is often unavailable. To address this unmet need, a HTS platform, 'ThromboGenomics' was designed, targetting 77 genes known to be causative of genetic BPDs. Details of the application of this platform, available across the NHS, are described below.

The BPDs can broadly be divided into the following categories, according to their molecular basis:

1. **Abnormalities in plasma coagulation proteins**. A large proportion of the plasma coagulation proteins are involved in clotting by creation of a fibrin network into which the platelets adhere at sites of vascular injury. Other plasma coagulation proteins are involved in the process of clot lysis (fibrinolysis). Variants in the genes encoding this group of proteins can result in bleeding, e.g. haemophilia or thrombotic disorders.
2. **Abnormalities in platelets**. Platelets are the source of primary haemostasis by adhesion and aggregation at sites of vascular injury. Abnormalities could be in their:

- membrane proteins, including cell surface receptors
- cytoplasmic proteins implicated in their formation
- granules – alpha, delta and lysosomal
- transcription factors and other proteins implicated in transcription and translation.

Figure 1: Overview of primary and secondary haemostasis



This gives a schematic overview of the different aspects of primary and secondary haemostasis. The left panel gives an overview of the process of platelet formation, starting with the haematopoietic stem cell (HSC) in the bone marrow, the generation of megakaryocytes and platelets. The genes implicated in these steps are mentioned, all of which are sequenced on the ThromboGenomics platform. The middle upper panel depicts the process of primary haemostasis. The far right panel shows the process of secondary haemostasis, including inhibitors and fibrinolysis. All genes coding for these proteins are present on the ThromboGenomics platform.

Below we will discuss some of the known disorders, focussing on platelet disorders.

Plasma coagulation proteins

Pathogenic variants in 26 genes encoding plasma coagulation proteins have been identified as causal of an increased propensity to bleeding (Figure 1). The two most frequent bleeding disorders in the category of coagulation proteins are haemophilia A and B and VWD. The severity of the bleeding coagulation disorders varies widely from mild bleeding to conditions, which if not managed properly may cause life-threatening bleeding episodes. Other inherited coagulation factor disorders are extremely rare.[1]

Abnormalities in platelet membrane proteins

In total, 50 genes involved in platelet disorders have been identified. Classic examples of the platelet membrane glycoprotein disorders are Glanzmann's Thrombasthenia (GT) and Bernard Soulier syndrome (BSS). Both disorders are autosomal recessive with some exceptions. Platelets from BSS patients are giant in appearance and lack a functional GP1B/IX/V receptor complex for von Willebrand Factor (Vwf). BSS is caused by pathogenic variants in *GP1BA*, *GP1BB* and *GP9* genes, but no cases caused by variants in the *GP5* gene have been reported. Platelets from BSS cases do not aggregate in response to ristocetin, but aggregation responses to other platelet agonists are normal. This is in strong contrast with the platelets from patients with GT, which are normal in morphological appearance, but fail to aggregate in response to both strong (collagen, thrombin) and weak (ADP, epinephrine) platelet agonists, but their ristocetin-induced agglutination is intact. The ThromboGenomics HTS platform also contains newly identified genes like the tyrosine kinase SRC, in which a rare gain-of-function variant results in a bleeding propensity because of poorly functioning 'grey platelets' lacking alpha-granules , which contain Vwf and thrombospondin 1 (Thbs 1).[2] Myelofibrosis and severe early onset osteoporosis appear to be associated manifestations of this disorder.

Platelet cytoplasmic proteins

Pathogenic variants in to date eight genes encoding cytoplasmic proteins generally lead to an abnormality of the count and volume of platelets. These aberrant values are the result of atypical megakaryopoiesis often associated with abnormal formation of proplatelets, sometimes also leading to abnormal morphology of platelets on electron microscopic inspection. Clinically it is important to separate these conditions into those cases which are limited to abnormalities of count and volume and sometimes morphology, versus the more syndromic cases in whom other organ systems may also be affected. The two classic examples of syn-

dromes caused by pathogenic variants in cytoplasmic proteins are Wiskott-Aldrich syndrome (WAS) and the non-muscle myosin disorders, caused by pathogenic variants in the *WAS* and *MYH9* genes respectively. WAS presents at a young age with a reduction in the number of platelets, which are also smaller. The syndrome is typified by childhood eczema and there is an increased prevalence of lymphoid disorders warranting consideration of stem cell transplant. The *MYH9*-related disorders are characterised by large platelets, often but not always accompanied by inclusion bodies in neutrophils. Patients with *MYH9* disorders may bleed, but premature loss of hearing or a reduction in kidney function are the primary reasons for clinical presentation. The phenotype association with *MYH9* variants is quite heterogeneous. Therefore, if found coincidentally because of a routine full blood count, clinical follow-up for early symptoms is warranted. Many of the other platelet disorders caused by rare variants in genes like *ACTN1*, *ANKRD26*, *CYCS*, *DIAPH1*, *FLNA* and *TUBB1* seem to uniquely modify the count of platelets and their volume, although the causal SNVs in the 5' untranslated region (UTR) in *ANKRD26* and a gain-of function variant in *DIAPH1* is associated with a mildly increased risk of leukaemia and early-onset deafness respectively.[3]

Disorders of the platelet granules

Platelets have alpha, dense and lysosomal granules, which are released upon cell activation. The cargo of alpha-granules is essential to the development of the platelet plug (factor V, fibronectin, Thbs1, Vwf) and wound healing (angiopoietin and insulin-like-, platelet-derived- and vascular-endothelial cell growth factors, etc.) at sites of vascular damage. Grey platelet syndrome (GPS) is an autosomal recessive BPD characterised by the lack of alpha-granules and large platelets of a greyish appearance. Patients may experience severe bleeding and myelofibrosis accompanied by extramedullary haematopoiesis frequently complicates this extremely rare condition. Rare variants in *NBEAL2*, which encodes a multi-domain scaffolding protein, are causal for typical GPS. *NBEAL2* is one of the seven members of the family of BEACH (BEige and Chédiak–Higashi syndrome [CHS]) genes, which are thought to have a critical function in the fusion, fission and trafficking of vesicles. Pathogenic variants in *LYST*, one of the other members of the BEACH family, underlie CHS. It mainly presents as an immune-deficiency disorder, but inadequate function of platelets may also result in a bleeding phenotype. NBEA, another member of the BEACH family, is implicated in dense-granule ontogeny, and deletion of a single copy of *NBEA* leads to the formation of atypical platelet dense-granules and severe autism because the assumed role of *NBEA* in neuronal cell function.

Hermansky Pudlack syndrome (HPS) affects the delta-granules, which contain among others ADP for paracrine activation, plus serotonin. The platelet and bleeding phenotypes vary, but HPS is generally typified by a mild bleeding tendency, with mild thrombocytopenia in some. The genetic architecture of HPS is diverse and so far nine genes have been linked to this syndrome, which encompasses ocular albinism and lung fibrosis on occasions. Genetic characterisation informs clinical management because the risk of fibrosis is linked to certain HPS genes but not to others. Hence patients with HPS caused by pathogenic variants in genes of the latter category do not require regular follow-up by respiratory physicians.

Abnormalities in transcription factors

Pathogenic variants in the transcription factor encoding genes *FLI1*, *GATA1*, *GFI1B*, *HOXA11* and *RUNX1* have been reported in relation with BPDs. The importance of the transcription factor Fli1 was illustrated in patients with Paris-Trousseau syndrome, which is characterised by several congenital anomalies including ineffective megakaryopoiesis and is caused by a hemizygous microdeletion on 11q24.3, which includes the *FLI1* gene.[4] Variants in *GATA1* underlie an X-linked macrothrombocytopenia, which may be accompanied by anaemia of variable severity attributable to dyserythropoiesis, but there are no other obvious pathobiologies.[5] Interestingly *GFI1B* variants can cause a dominant form of GPS, which is not associated with myelofibrosis. The important role of *RUNX1* in the regulation of megakaryopoiesis was first revealed by the thrombocytopenia-acute myeloid leukaemia (AML) syndrome. In this extremely rare disorder, the AML is preceded by a severe thrombocytopenia at early age, which may initially be diagnosed as being of autoimmune nature. *HOXA11* is causally linked with amegakaryocytic thrombocytopenia with radio-ulnar synostosis and is one of the few examples of link between myeloid haematopoiesis and bone development. Thrombocytopenia with absent radii or TAR syndrome is another example of this link and is generally caused by a compound inheritance of a microdeletion encompassing the *RBM8A* gene and a regulatory variant, typically but not always localised in the 5' UTR of *RBM8A*. The gene encodes Y14, one of the four proteins of the exon-junction complex, which performs essential RNA processing tasks. In TAR syndrome cases, the level of Y14 protein is strongly reduced in the megakaryocytic lineage, leading to a paucity of megakaryocytes, thrombocytopenia and increased risk of bleeding. The other blood cell lineages are not affected, which is explained by the TAR-causing regulatory variants being localised in DNA elements which are specifically nucleosome depleted in megakaryocytes but not in erythroblasts.

Genetic diagnosis of BPDs

After the cloning of the genes for Factors VIII and IX in the 1980s and with the advent of the

polymerase chain reaction, it became customary in cases of haemophilia A and B to determine the causal variants by Sanger sequencing. The known disease-causing variants have been catalogued and deposited in gene-specific databases for use by genetics laboratories. For VWD, the cost of Sanger sequencing was generally deemed prohibitive and laboratory tests for Vwf antigen level and function (RICOF) are used without genetic confirmation. For the other extremely rare genetic BPDs, achieving a molecular diagnosis has remained far more cumbersome. A plethora of specialised assays, often not all available from a single laboratory, are required, often not all available from a single laboratory, and analysis by Sanger sequencing is only available under the UK Genetic Testing Network for a few of the 77 known genes. This lack of comprehensive and affordable DNA tests was the main reason for the development of the ThromboGenomics platform. The development of this HTS application required the following steps to be made:

- The knowledge of the 77 relevant genes were curated and deposited in the Locus Reference Genomics database (www.lrg-sequence.org)
- Information about pathogenic and likely pathogenic variants from gene-specific databases were transferred to a publically accessible database like ClinVar (www.ncbi.nlm.nih.gov/clinvar)
- An HTS pull-down platform was designed and validated using DNA samples from patients with BPDs with known causal variants.
- Human phenotype ontology (HPO) coding was established for the BPDs of known and unknown molecular aetiology. The HPO is an open-source project established to develop a system of phenotypic annotation for genetic disorders. Over 10,000 terms in the HPO are connected via a hierarchy of 'is-a' relationships. A specific example would be: '*thrombocytopenia* **is-a** *abnormal platelet count* **is-a** *abnormality of thrombocytes*'. Development of this system for application to the BPD project has required the addition of 80 more terms to the HPO tree, the vast majority of which provided further definitions within the 'Anormality of blood and blood forming tissue' category and dependent terms. Full details of the HPO programme are published elsewhere.[6]
- A website (www.thrombogenomics.org.uk) was required to share all the above information with relevant healthcare workers, patients and their close relatives, and the public at large.
- A multi-disciplinary team (MDT) with the relevant expertise was established to review variants observed in referred cases in the context of the clinical and laboratory phenotype data.

For more details and to request a ThromboGenomics investigation, please visit the website www.thrombogenomics.org.uk

The ThromboGenomics investigation in practice

The ThromboGenomics test can be requested through the NHS; Figure 2 summarises the process. The UK Haemophilia Centre Doctors' Organisation patient information leaflet, a routine NHS genetic test consent form and referral card can be downloaded from www.thrombogenomics.org.uk. The referring clinician provides relevant clinical and

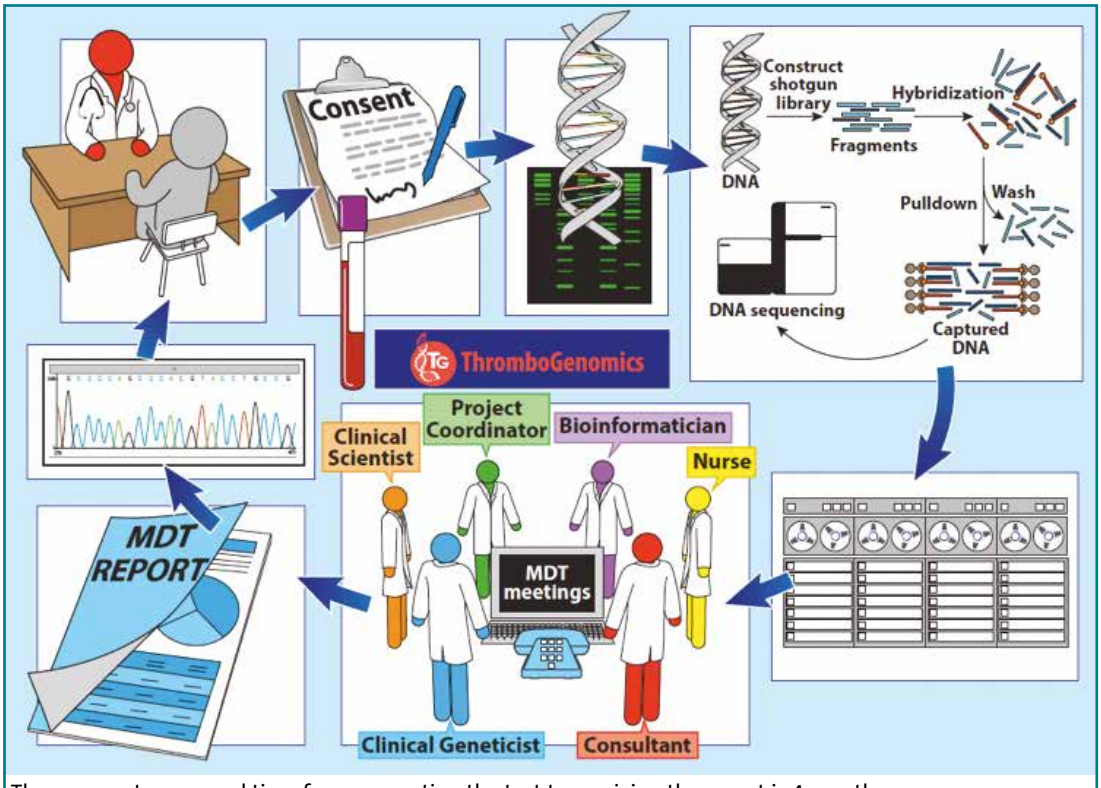


Figure 2: The processes involved in ThromboGenomics testing

The average turnaround time from requesting the test to receiving the report is 4 months

laboratory data on the referral card at the time of request. A sample of EDTA anticoagulated blood, together with the completed referral card, are sent to the Clinical Genetics Laboratory at Cambridge University Hospitals NHS Trust, where HPO terms are appended to the patient's computer record and genomic DNA is extracted and its quality checked by gel electrophoresis. DNA samples from 24 patients are processed and sequenced in a batch. Variant calls and patient phenotypes (in HPO) are deposited and visualised in the Sapientia web application for assessment during MDT meetings. The MDT team consists of a consultant haematologists, clinical scientists and bioinformaticians, and is chaired by Dr Keith Gomez from the Royal Free Hospital in London. The Sapientia application (Congenica Ltd, Hinxton, Cambridge) displays variant information such as predicted effect, allele frequencies in control data sets comprising nearly 70,000 DNA samples and presence in the Human Gene Mutation Database, patient data and phenotype information in the form of HPO terms. After automated filtering for diagnostic prioritisation, an average of 5.34 candidate variants per patient require assessment by the MDT. The web application allows the MDT to annotate each variant with respect to its predicted full or partial contribution to the disease phenotype and whether it is pathogenic or likely pathogenic. Variants are assigned the pathogenic label if they have been observed in four genetically unrelated BPD cases with similar clinical features, and if the frequency in the control data is compatible with the presumed mode of inheritance e.g. high penetrance autosomal dominant DNA variants are unlikely to have a MAF of >1 in 100,000 and are generally unobserved in the control data, while for autosomal recessive disorders, a MAF ranging between 1 in 1000 and 1 in 10,000 is compatible with the observed rarity of BPDs. Finally, the MDT produces a clinical diagnostic report for the referring clinician. For the time being, HTS-identified variants deemed causal by the MDT are confirmed by Sanger sequencing and variants of unknown clinical significance (VUS) are not reported. Current NHS guidance for an acceptable turnaround time for tests like the ThromboGenomics one is 4 months, but

the laboratory makes all reasonable efforts to shorten this interval.

Conclusion

The ThromboGenomics HTS test is the first international comprehensive platform containing the accepted BPD genes for affordable sequencing by HTS in patient samples. We expect that this will lead to a far more and rapid diagnosis of a greater breadth of BPD cases, including hitherto un-sequenced extremely rare disorders. To date, all (international) samples have been processed at a single Genomic Medicine Centre, based at Cambridge University Hospitals. The test is routinely available for all clinicians in the NHS and will be free of charge in 2016. We predict that in the future, when the price has dropped for WGS, all individuals may be whole genome sequenced at least once in their lifetime. Subsequently disease-specific bio-informatics will pull down the necessary genetic sequence needed to answer disorder-related questions. So in the case of the ThromboGenomics gene panel, a computational query will be run on the whole genome sequence data, to test variation in all BPD genes when needed. Until that day comes, a platform such as ThromboGenomics will bring substantial benefits in the form of precision medicine to the care of BPD patients and their families.

Suthesh Sivapalaratnam and Willem Ouwehand
University of Cambridge
On behalf of the ThromboGenomics Consortium

Acknowledgments

The authors acknowledge Dr Kate Downes, Dr Dan Hart, Dr Keith Gomez and Prof Mike Laffan for their helpful comments.

References

1. BCSH. Guideline for the diagnosis and management of the rare coagulation disorders. *British Journal of Haematology* 2014;167:304–326. http://onlinelibrary.wiley.com/doi/10.1111/bjh.13058/full.
2. Turro E, Greene D, Wijgaerts A, Thys C, Lentaigne C, Bariana TK, Westbury SK *et al.* A dominant gain-of-function mutation in universal tyrosine kinase SRC causes thrombocytopenia, myelofibrosis, bleeding, and bone pathologies. *Sci Transl Med* 2016;8:328ra30.
3. Stritt S, Nurden P, Turro E, Greene D, Jansen SB, Westbury SK, Petersen R *et al.* A gain-of-function variant in DIAPH1 causes dominant macrothrombocytopenia and hearing loss. *Blood* 2016; Feb 24. pii: blood-2015-10-675629 [Epub ahead of print].
4. Raslova H, Komura E, Le Couedic JP, Larbret F, Debili N, Feunteun J, Danos O *et al.* FLI1 monoallelic expression combined with its hemizygous loss underlies Paris-Trousseau/Jacobsen thrombopenia. *J Clin Invest* 2004;114:77–84.
5. Freson K, Devriendt K, Matthijs G, Van Hoof A, De Vos R, Thys C, Minner K, Hoylaerts MF, Vermylen J, Van Geet C. Platelet characteristics in patients with X-linked macrothrombocytopenia because of a novel GATA1 mutation. *Blood* 2001;98:85–92.
6. Westbury SK, Turro E, Greene D, Lentaigne C, Kelly AM, Bariana TK, Simeoni I *et al.* Human phenotype ontology annotation and cluster analysis to unravel genetic defects in 707 cases with unexplained bleeding and platelet disorders. *Genome Med* 2015;7:36