**A whole genome approach to platelet and bleeding disorders**

**Prof Mike Laffan on behalf of the BRIDGE Bleeding and Platelet Disorders Consortium**

**Address for correspondence**

Mike Laffan

Professor of Haemostasis and Thrombosis,

Honorary Consultant in Haematology

Faculty of Medicine

Imperial College

Hammersmith Hospital

Ducane Road

London W12 ONN

**m.laffan@imperial.ac.uk**

**Summary**

The catalogue of human genetic variation is being progressively expanded by the vast amount of data resulting from developments in sequencing technology and its application to whole-exome and whole-genome sequencing projects. We have determined to apply these techniques to patients with bleeding disorders and combine the results with the data from hundreds of thousands of human genomes in order to identify genetic variants putatively responsible for the disorders. In order to facilitate the analysis by accurately matching clinical and genetic data, we first extended the Human Phenotype Ontology to accommodate 80 new relevant terms and then developed a cluster analysis to group patients with similar phenotypes. The application of these techniques to known genetic platelet disorders provided validation of the approach and further analyses have resulted in identification of a number of genes responsible for normal and pathogenic platelet physiology.

**The genome and sequence variation**

The sequencing of hundreds of thousands of human exomes and hundreds of thousands of whole genomes is providing a progressively accurate and complete catalogue of human genetic variation. Most common amongst these changes are single nucleotide variants (SNVs) and to date roughly 63 million SNVs have been identified, the majority of which are rare. The two principal other forms of variation are copy number variants (CNV) and structural variants (SV). CNV are much less frequent and comprise about one CNV for each 10 SNVs. The continuing reduction in sequencing costs has enabled initiatives like the 1000 Genomes project ([1](#_ENREF_1)), the UK10K project (<http://www.uk10k.org/>) ([2](#_ENREF_2)) and the NHS 100,000 whole genome sequencing (WGS) project. The NHS 100,000 genomes project aims to achieve WGS analysis of 100,000 NHS patients by 2008 (<http://www.genomicsengland.co.uk/>) ([3](#_ENREF_3)) . The key objective of this project is to establish WGS as a part of standard care within the NHS. The resulting increase in the accumulated number of genome sequences will be enormous and will greatly increase the accuracy of minor allele frequencies (MAFs) for each of the types of variants. Initially the DNA samples and sequences will derive predominantly from Northern-European populations, but in time equivalent data from populations of different ethnicities will be obtained.

**Genome wide association studies and platelet formation**

The initial studies to use genome wide data to help understand platelet disorders performed genome wide association studies to identify loci linked to variations in blood cell parameters. In 2009 two such studies, using DNA samples from a relatively small number of healthy individuals, identified four SNVs linked to the volume and number of platelets. Three of these SNVs were located in or in close proximity to the *ARHGEF3*, *TAOK1* and *WDR66* genes which were therefore to be considered possible regulators of platelet formation ([4](#_ENREF_4), [5](#_ENREF_5)). The remaining SNV rs342293 at 7q22.3, which also exerted an effect on platelet function, lay between the hypothetical gene *FLJ36031* and *PIK3CG*. Of the four genes identified, only *PIK3CG* encoded a protein previously studied in myeloid biology. Thus the application of techniques taking advantage of our expanding knowledge of the human genome has enabled identification of novel and important genes. These studies used normal variation to find corresponding genetic variation. We next wished to investigate the genetic basis of bleeding disorders which may also provide a key to novel genes regulating platelet and haemostatic functions.

Some major platelet disorders such as Glanzmann’s thrombasthenia and Bernard Soulier syndrome are examples of well characterised disorders arising from mutations in single genes giving rise to a distinct and easily recognised phenotype. The same is largely true of disorders of coagulation such as haemophilia A and B and although von Willebrand disease (VWD) is more complex, the tools for characterisation of the laboratory phenotype are well enough developed to allow systematic classification. Nonetheless, approximately one third of patients diagnosed with VWD were found to have no linkage of their bleeding phenotype to the VWF gene.

The remaining platelet and bleeding disorders are heterogeneous, arising from abnormalities of plasma factors, platelets and disorders of the vessel wall. The usual progression from clinical assessment to laboratory testing enables diagnosis at the level of a specific haemostatic component in only 50% of cases ([6](#_ENREF_6), [7](#_ENREF_7)). Furthermore the genetic basis is identified in an even smaller fraction; even in the more prevalent groups such as those with platelet granule secretion defects where the full extent of the physiological mechanisms are not known. Some progress has been made in subgroups such as the MYH9-Rare Disease disorders and Hermansky-Pudlak syndrome (HPS), but here the wide variety of phenotypes and in the case of HPS the large number of complementation groups, are both barriers to accurate diagnosis.

To some extent the problems of diagnosis in this group are inevitable given firstly the relatively crude nature of platelet function tests routinely available (aggregation and nucleotide pool analyses) and secondly that some haemostatic components (notably the vessel wall) cannot readily be sampled for analysis. Finally it has proved extremely difficult to reproduce the physiological circumstances of haemostasis in the laboratory to develop properly functional tests of individual patient samples. We therefore proposed as part of the BRIDGE project to use new techniques and large scale collaboration to overcome these problems and to identify novel genes responsible for haemostatic defects in patients with undefined bleeding and platelet disorders (BPD).

**The BRIDGE project**

The BRIDGE consortium (www.bridgestudy.org) is funded by the NIHR and brings together 13 rare disease gene discovery projects. The aim of these projects is to investigate as yet undiagnosed rare inherited diseases and identify the underlying mutational basis. Originally this also included the additional intention of evaluating the sensitivity and specificity of whole exome sequencing (WES), but with decreasing costs of the techniques involved this has advanced to an evaluation of WGS. An important principle of the studies is the sharing of data between participating centres and between different BRIDGE groups, which is recognised as essential for accelerating gene discovery. Participating centres in the UK and beyond have enrolled patients with bleeding and platelet disorders (BPD) identified using standard clinical methods (Table 1). A number of known syndromic platelet defect patients have also been included and provide a measure of sensitivity and validation. To date approximately 800 patients from 700 kindreds have been enrolled and over 600 have been sequenced.

Thus the central component of the BRIDGE study is the use WGS analysis by next generation sequencing (NGS) to give a complete genetic profile of all patients. This approach immediately extends analysis beyond known platelet components and beyond the candidate gene approach to diagnosis. The downside of this is the identification of very large numbers of variants, but the increasing availability of genetic information including the NHS 100,000 genomes project, the UK10 and ExAC Databases and WGS results from patients in different projects of the BRIDGE Consortium make it relatively easy to identify mutations that are likely to be causally linked to the abnormal phenotypes.

The second component is the coding of patients using the Human Phenotype Ontology (HPO). This allows the grouping of patients according to their phenotypic characteristics and importantly includes all clinical characteristics and not only the laboratory platelet and coagulation phenotypes. It is important to note that other coding systems such as the World Health Organisation International Classification of Diseases (ICD)([8](#_ENREF_8)) (which is based on organ systems); the International Health Terminology Standards Development Organisation Systematized Nomenclature of Medicine Clinical Terminology (SNOMED-CT)([9](#_ENREF_9)) ( which is disease based and is not designed for laboratory measures such as coagulation parameters, platelet function and morphology) or the Online Mendelian Inheritance in Man (OMIM)([10](#_ENREF_10)) and Orphanet ([11](#_ENREF_11)) systems (which are restricted to known genetic disorders not phenotypes) are not suitable for this purpose. Clustering of patients with common phenotypic characteristics helps validate the significance of genetic variants and reduces the need for large kindreds. However we note that segregation of putative promising variants in informative pedigrees using Sanger sequencing remains an important and valuable test of imputed pathogenicity.

**The Human phenotype ontology (HPO)**

The HPO is an open source project established to develop a system of phenotypic annotation for genetic disorders. The more than 10000 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 BRIDGE-BPD project has required the addition of 80 more terms to the HPO tree, the vast majority of which provided further definitions within the *abnormality of blood and blood forming tissue* category and dependent terms.

Full details of the HPO programme will be published elsewhere. However a key development was the use of HPO terms to perform a clustering analysis of the enrolled patients. This was done by defining the information content of the terms according to their rareness within the cohort such that the rarer the term the higher the information content. The closeness of any two individuals could then be defined using an algorithm to assess the information available from their most informative shared terms. Individuals who share relatively rare terms thus tend to cluster as a group whose closeness to each other and distance from other groups can be defined. We hypothesised that this would aid the discovery of genes on the grounds that clusters of patients with shared phenotypic traits are likely to also share defects in the same gene (or interacting genes).

**Recruitment and enrolment of patients**

Entrants were identified primarily from lists of patients registered at haemophilia centres in the UK, but also from specialist centres abroad. A specific ethical approval and consent form permitting follow up and recall as well as extended laboratory investigation was developed for the study. At entry, demographic, clinical and laboratory data were recorded as well as a standardised bleeding score. Pseudonymised data were then entered using the standardised vocabulary onto the BRIDGE-BPD study database, including the HPO terms.

**Sequencing**

For WES analysis DNA library capture was performed using ROCHE NimbleGen SeqCap EZ 64Mb Human Exome Library version 3.0 (ROCHE NimbleGen, Inc. Madison, WI, USA). The libraries were sequenced on an Illumina Hiseq 2000 instrument. Preliminary analysis comprised removal of: variants with an allele frequency of >0.1% in any of the reference cohorts, variants unlikely to alter the protein by snpEff 3.4, variants not present in other BRIDGE recruits, <3 reads supporting the variant allele; variants present in any of the 2,000 in house controls and allele count >10. Further assessment comprised review of likely mutation effect, case similarities, gene function and transcription data from the Blueprint epigenome study([12](#_ENREF_12), [13](#_ENREF_13)). Candidate genes and variants therein were then chosen for segregation studies.

**Overview of preliminary results**

The large number of patients enrolled provides an overview of the BPD population. Approximately 75% of index cases have an abnormal bleeding phenotype, 25% an isolated platelet abnormality with no associated bleeding, 32% a platelet count <100x10^9/l and 54% have defective aggregation by light transmission technique. The number of encoded HPO terms varied from 1 to 22 with a median of 7 per case. Notably, many cases had entries in terms outside the blood forming tissues and dependent terms. Preliminary validation was provided by demonstrating the ability of HPO to cluster patients who were members of a kindred and by the clustering of patients with known bleeding related syndromes (Fig 1).

Analysis of cases with Grey Platelet Syndrome (GPS) and Thrombocytopenia with Absent Radii (TAR) identified *NBEAL2* ([14-16](#_ENREF_14))and *RBM8A(*[*17*](#_ENREF_17)*,* [*18*](#_ENREF_18)*)*as the causative genes for these two syndromes. The GPS discovery is a classic example of an autosomal recessive platelet bleeding disorder. Loss-of-function mutations on both *NBEAL2* haplotypes leads to the formation of platelets which are devoid of α-granules and myelofibrosis develops in nearly all GPS cases at later age. Studies in *Nbeal2* knockout mice have recapitulated the phenotype observed in humans, although further molecular studies are required to define the function of the NBEAL2 protein in granule formation and retention([19-21](#_ENREF_19)) The study of TAR cases revealed an entirely novel genetic mechanism where a submicroscopic deletion of 1q21.1 on one parental haplotype is compounded by the minor allele of a SNV present in 1 in 30 healthy controls, on the other haplotype. This SNV is localised in a megakaryocyte-specific enhancer element in the 5’ UTR of the *RBM8A* gene and the minor allele introduces a binding site for the transcription factor EVI1. The binding of EVI1 at this position is associated with less effective transcription from the apparently intact *RBM8A* gene causing an insufficiency of Y14, the protein encoded by *RBM8A.* Gene ablation in mice and knockdown in zebrafish of *Rbm8a* are not compatible with life. The relative insufficiency of Y14 in the megakaryocyte lineage, caused by the coming together of a deletion and the repressive 5’UTR-SNV, leads to a selective maturation block with a paucity of megakaryocytes, which is a classic hallmark of TAR. Because the effect of the minor allele is in haematopoietic cells only present in the megakaryocyte lineage, the differentiation of the other myeloid lineages can proceed normally.

The challenges of analysis of the results obtained by WES/WGS of the remaining 500 BRIDGE BPD cases are substantial. First, the BRIDGE collection of BPD cases is clinically more heterogeneous than GPS and TAR. Second power of gene discovery is eroded by this phenotype heterogeneity, but it is hoped that this can be recovered by clustering of patients using the structured and detailed phenotypic information entered into HPO. Third three of the inherited conditions at the *ANKRD26*, *RBM8A* and *SMIM1* ([22](#_ENREF_22)) loci could not have been revealed by WES analysis only as variants in regulatory elements are part of the observed genetic mechanisms.

**Discussion**

Gene discovery and functional genomics have revolutionised our understanding of the genetic architecture of megakaryopoiesis and platelet formation. A large number of important novel regulators of both processes have been revealed. The discovery of novel genes implicated in rare BPDs supports an international effort to bring to the clinic an affordable and comprehensive NGS platform containing the accepted BPD genes for their affordable sequencing by NGS in patient samples thus leading to a far more rapid diagnosis. This will not only remedy the considerable diagnostic delay for index cases but also provide information for family planning purpose if required.

**Table 1. Eligibility criteria for BRIDGE Bleeding and Platelet Disorders study**

**Inclusion criteria**

Platelet count less than 100 x109/L or greater than 400 x109/L, or

Mean platelet volume less than 6 fL or greater than 12 fL, or

Reproducible abnormal platelet function test results, or

Abnormal platelet morphology by light or electron microscopy, or

Pathological bleeding of unknown aetiology, and

Considered by referring clinician to be of genetic aetiology

**Exclusion criteria**

Acquired bleeding or platelet disorders, including any of the following:

Use of any medication known to affect platelet function or cause bleeding

Immune Thrombocytopenia

HIV infection

Malignancies, particularly those affecting haematopoiesis

Bone marrow aplasia

Thrombotic thrombocytopenic purpura/ Haemolytic-uremic syndrome

Acute viral infection

Splenomegaly

Uraemia or hepatic failure

**Figure 1. Clustering of index cases according to HPO terms.**

**A**) Heat map showing pairwise phenotypic similarity amongst affected members of pedigrees,

cases with classical syndromes and cases with pertinent findings in the *MYH9* gene.

The groups are ordered through complete-linkage hierarchical clustering within each

class and *p*-values of group association are shown in a scatterplot.

**B**) Subgraph of the HPO showing terms shared by at least 2 of the 9 cases (number shown in each

node) with pertinent findings in *MYH9* and the frequency of these terms amongst all

cases in the BPD collection. Arrows indicate direct (solid) or indirect (dashed) *is a*

relations between terms in the ontology. PA, phenotypic abnormality; BBT, blood and

blood-forming tissues; BMT, bleeding with minor or no trauma; TCP,

thrombocytopenia; Subcut, subcutaneous.

**References**

1. Genomes Project C, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, et al. A map of human genome variation from population-scale sequencing.[Erratum appears in Nature. 2011 May 26;473(7348):544 Note: Xue, Yali [added]; Cartwright, Reed A [added]; Altshuler, David L [corrected to Altshuler, David]; Kebbel, Andrew [corrected to Keebler, Jonathan]; Koko-Gonzales, Paula [corrected to Kokko-Gonzales, Paula]; Nickerson, Debbie A [corrected to Nickerson, Deborah A]]. Nature. 2010;467(7319):1061-73.

2. Kaye J, Hurles M, Griffin H, Grewal J, Bobrow M, Timpson N, et al. Managing clinically significant findings in research: the UK10K example. Eur J Hum Genet. 2014;22(9):1100-4.

3. Moran N. 10,000 rare-disease genomes sequenced. Nat Biotech. 2014;32(1):7-.

4. Meisinger C, Prokisch H, Gieger C, Soranzo N, Mehta D, Rosskopf D, et al. A genome-wide association study identifies three loci associated with mean platelet volume. Am J Hum Genet. 2009;84(1):66-71.

5. Soranzo N, Spector TD, Mangino M, Kuhnel B, Rendon A, Teumer A, et al. A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. Nat Genet. 2009;41(11):1182-90.

6. Hayward CP, Pai M, Liu Y, Moffat KA, Seecharan J, Webert KE, et al. Diagnostic utility of light transmission platelet aggregometry: results from a prospective study of individuals referred for bleeding disorder assessments. J Thromb Haemost. 2009;7(4):676-84.

7. Quiroga T, Goycoolea M, Panes O, Aranda E, Martinez C, Belmont S, et al. High prevalence of bleeders of unknown cause among patients with inherited mucocutaneous bleeding. A prospective study of 280 patients and 299 controls. Haematologica. 2007;92(3):357-65.

8. Organisation WH. [cited 2014 September]. Available from: <http://www.who.int/classifications/icd/en/>.

9. Organisation. IHTSD. [cited 2014 September]. Available from: <http://www.ihtsdo.org/snomed-ct/>.

10. Information. NCfB. OMIM: Online Mendelian inheritance in man. [cited 2014 September]. Available from: <http://www.ncbi.nlm.nih.gov/omim>.

11. 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. Hum Mutat. 2012;33(5):803-8.

12. Chen L, Kostadima M, Martens JHA, Canu G, Garcia SP, Turro E, et al. Transcriptional diversity during lineage commitment of human blood progenitors. Science. 2014;345(6204).

13. Saeed S, Quintin J, Kerstens HHD, Rao NA, Aghajanirefah A, Matarese F, et al. Epigenetic programming of monocyte-to-macrophage differentiation and trained innate immunity. Science. 2014;345(6204).

14. Albers CA, Cvejic A, Favier R, Bouwmans EE, Alessi MC, Bertone P, et al. Exome sequencing identifies NBEAL2 as the causative gene for gray platelet syndrome. Nat Genet. 2011;43(8):735-7.

15. Gunay-Aygun M, Falik-Zaccai TC, Vilboux T, Zivony-Elboum Y, Gumruk F, Cetin M, et al. NBEAL2 is mutated in gray platelet syndrome and is required for biogenesis of platelet alpha-granules. Nat Genet. 2011;43(8):732-4.

16. Kahr WH, Hinckley J, Li L, Schwertz H, Christensen H, Rowley JW, et al. Mutations in NBEAL2, encoding a BEACH protein, cause gray platelet syndrome. Nat Genet. 2011;43(8):738-40.

17. Albers CA, Newbury-Ecob R, Ouwehand WH, Ghevaert C. New insights into the genetic basis of TAR (thrombocytopenia-absent radii) syndrome. Curr Opin Genet Dev. 2013;23(3):316-23.

18. Albers CA, Paul DS, Schulze H, Freson K, Stephens JC, Smethurst PA, et al. Compound inheritance of a low-frequency regulatory SNP and a rare null mutation in exon-junction complex subunit RBM8A causes TAR syndrome. Nat Genet. 2012;44(4):435-9, S1-2.

19. Deppermann C, Nurden P, Nurden AT, Nieswandt B, Stegner D. The Nbeal2(-/-) mouse as a model for the gray platelet syndrome. Rare dis. 2013;1:e26561.

20. Guerrero JA, Bennett C, van der Weyden L, McKinney H, Chin M, Nurden P, et al. Gray Platelet Syndrome: Pro-inflammatory megakaryocytes and α-granule loss cause myelofibrosis and confer resistance to cancer metastasis in mice. Blood. 2014;124:3624-35.

21. Kahr WH, Lo RW, Li L, Pluthero FG, Christensen H, Ni R, et al. Abnormal megakaryocyte development and platelet function in Nbeal2(-/-) mice. Blood. 2013;122(19):3349-58.

22. Cvejic A, Haer-Wigman L, Stephens JC, Kostadima M, Smethurst PA, Frontini M, et al. SMIM1 underlies the Vel blood group and influences red blood cell traits. Nat Genet. 2013;45(5):542-5.

**Figure 1A**

 



**Figure 1B**