New gene functions in megakaryopoiesis and platelet function

Willem H. Ouwehand[1-3], Nicole Soranzo[3,1], Augusto Rendon[1,4], on behalf of the BRIDGE, HAEMGEN and THROMBOGENOMICS consortia

[1]Department of Haematology, University of Cambridge, [2]National Health Service Blood and Transplant, Cambridge CB2 0PT, UK; [3]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK; [4]Biostatistics Unit, Medical Research Council, Cambridge CB2 0SR, UK.

The sequencing of tens of thousands of human genomes provides an exhaustive and increasingly accurate catalogue of the types of DNA variants and their minor allele frequencies (MAFs) in population samples of different ethnicities. Single nucleotide variants (SNVs) are the primary source of sequence variation and about 30 million have been identified too date. Based on their MAFs, SNVs are categorised in arbitrary groups of common (MAF>0.05) and rare ones (<0.005 and >0.001) and mutations (<0.001). Most of the observed variants are assumed to have no effect on phenotype and therefore a relatively small fraction of variants may modify the risk of diseases or regulate disease-quantitative traits. Common SNPs typically have small effects on a trait, with less than 10% variation between the three genotype groups, whilst about 30,000 mutations and rare variants have large effects and are causative of Mendelian disorders.

In 2007 we set out to explore the relationship between DNA variants and the processes of megakaryopoiesis and the formation of platelets. The initial studies focussed on common variants in large population samples and more recently we have initiated the BRIDGE Bleeding and Platelet Disorders sequencing programme, which aims to discover rare variants and mutations causative of this class of rare inherited disorders.

The Genome Wide Association Study meta-analysis (GWAS) by the HaemGen consortium in a population sample of about 68,000 healthy individuals identified 68 common sequence variants that are associated with the volume and count of platelets at genome-wide significance level1. Computational biology approaches and wet-lab genome-wide and locus-focussed functional studies were used to better appreciate the mechanisms that underlie the observed associations2. The results of chromatin immunoprecipitation combined with massive parallel sequencing for 6 transcription factors (GATA1/2, FLI1, MEIS1, RUNX1, SCL (TAL1)) in primary megakaryocytes3,4 and formaldehyde assisted identification of regulatory elements (FAIRE) in primary erythroblasts, megakaryocytes and monocytes5,6 resulted in a better understanding of how non-coding GWAS variants effect the platelet phenotypes of volume and count. Expression studies in platelets and monocytes from genome-wide typed healthy individuals were used to delineate the cis-effects of association SNPs on transcript levels. From these studies a picture is emerging of allelic differences in the binding of several key transcription factors, including MEIS1 and EVI1 at association SNPs. These result point to allelic differences in transcript levels between genotype groups for a category of genes adjacent to or at the GWAS signals, for example at the PIK3CG and DNM3 loci. For the latter the GWAS association SNP together with a MEIS1 ChIP-seq binding site mark an alternative promoter in the DNM3 gene which is only operational in megakaryocytes but not in any other cell type known to transcribe this gene4, illustrating the lineage-specific effect of the common GWAS variants on phenotype.

Furthermore two thirds of the 68 genes that were identified in the GWAS for platelet volume and count were hitherto unknown to be implicated in megakaryopoiesis and platelet biology. We postulated that proteins that regulate the formation of platelets by megakaryocytes may also play an important role in platelet function. To test this hypothesis we performed a further candidate gene association analysis for the 68 GWAS SNPs on function in the Cambridge Platelet Function Cohort of 1,000 healthy individuals. In spite of the limited power of this study to delineate associations we did observe several common GWAS SNPs in different volume and count loci that also modified function.

With the advent of next generation sequencing we are now able to search the coding fraction of the genome, or so called exomes, for causative variants with large effect sizes and we applied this to

50 index cases with inherited bleeding and platelet disorders. This effort identified NBEAL2[7] and RBM8A[8,9] as the causative genes for Grey Platelet (GPS) and Thrombocytopenia with Absent Radii (TAR) syndromes, respectively. We also discovered another three genes underlying rare platelet disorders. This early success confirms the power of next generation sequencing to discover mutations for Mendelian disorders but the challenges of analysis of the results obtained by exome sequencing another 700 cases are substantial. This collection of cases is clinically far more heterogeneous than GPS and TAR. To maintain power of gene discovery there is a need to segregate patients in clusters based on detailed clinical phenotype information and completeness of laboratory results. The former information is gathered by assigning Human Phenotype Ontology (HPO) terms to each patient and by grading the severity of bleeding. An integrated analysis, which included full blood count results, did show clustering of patients across centres. It is hoped that this clustering will be an effective method to maintain power for gene discovery.

Finally the translation of next generation sequencing into the clinic needs to be addressed. Currently the diagnosis of rare inherited bleeding and platelet disorders is based on a plethora of biochemical and platelet and coagulation factor function tests, many of which have a poor Receiver Operating Characteristic and considerable laboratory expertise is required to perform these laborious tests. Ideally a single and affordable DNA-based test should be available. To develop an integrated sequencing platform, backed-up by a knowledge database about the clinical consequence of mutations and rare sequence variants in relevant candidate genes a working group, named ThromboGenomics (www.thromboggenomics.org.uk) has been established under the auspices of the International Society for Thrombosis and Haemostasis (ISTH). This ISTH working group is driving forward two parallel initiatives. First a list of 94 candidate genes known to be implicated in inherited bleeding and platelet disorders has been compiled. Second an international group of gene experts work together with the Locus Reference Genomic team at the European Informatics Team to curate the variant knowledge of these 94 genes to clinical standards. Third our department has with Nimblegen/Roche developed a pull-down platform for the sequencing of the exons of the 94 genes in pooled and indexed clinical samples. An initial validation of 95 clinical samples with a known genetic or clinical diagnosis showed an overall sensitivity of 94%, with 2 of the false-negative results being the consequence of structural variants which cannot be detected by this type of sequencing assay and the remaining two can be resolved by platform modifications.

In conclusion 6 years of platelet genomics research has revolutionised our understanding of the genetic architecture of megakaryopoiesis and platelet formation and a large number of important novel regulators of both processes has been revealed. The discovery of novel genes implicated in rare bleeding and platelet disorders supports an international effort to bring to the clinic an affordable and comprehensive next generation sequencing platform for the rapid diagnosis of most of the known inherited bleeding and platelet disorders. This will not only remedy the considerable diagnostic delay for the index case but more importantly also provide information for family planning purpose if required.

References

1 Gieger, C. et al. New gene functions in megakaryopoiesis and platelet formation. Nature 480, 201-208, doi:10.1038/nature10659 (2011).

2 Serbanovic-Canic, J. et al. Silencing of RhoA nucleotide exchange factor, ARHGEF3, reveals its unexpected role in iron uptake. Blood 118, 4967-4976, doi:10.1182/blood-2011-02-337295 (2011).

3 Tijssen, M. R. et al. Genome-wide analysis of simultaneous GATA1/2, RUNX1, FLI1, and SCL binding in megakaryocytes identifies hematopoietic regulators. Dev Cell 20, 597-609, doi:10.1016/j.devcel.2011.04.008 (2011).

4. Nurnberg S.T., et al. A GWAS sequence variant for platelet volume marks an alternative DNM3 promoter in megakaryocytes near a MEIS1 binding site. (2012) Blood, 120, 4859-4868.

5 Paul, D. S. et al. Maps of open chromatin guide the functional follow-up of genome-wide association signals: application to hematological traits. PLoS Genet 7, e1002139, doi:10.1371/journal.pgen.1002139 (2011).

6 Paul, D. S. et al. Maps of open chromatin highlight cell type-restricted patterns of regulatory sequence variation at hematological trait loci. (2013) Genome Res, 23, 1130-1141. PubMed PMID: 23570689

7 Albers, C. A. et al. Exome sequencing identifies NBEAL2 as the causative gene for gray platelet syndrome. Nature genetics 43, 735-737, doi:10.1038/ng.885 (2011).

8 Albers, C. A. 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, doi:10.1038/ng.1083 (2012).

9 Albers, C. A. et al. New insights into the genetic basis of TAR (thrombocytopenia-absent radii) syndrome. (2013) Curr Opin Genet Dev, 23, 316-323. PubMed PMID: 23602329