Gold Variants: Defining a high-quality set of clinically relevant DNA variants with, and for, the Thrombosis and Haemostasis Community

Subcommittee: Genomics in Thrombosis and Hemostasis (GinTH)

Person responsible (Chair / Principal Investigator): Willem H Ouwehand, UK (Chair SSC GinTH 2014-2018); Kathleen Freson, Belgium (Chair-Elect SSC GinTH 2018-2022)

Collaboration with the following SSCs:
The proposed project is a collaboration with 5 other SSC:
- Prof Paolo Gresele, Italy, SSC Platelet Physiology
- Prof Guy Young, USA, SSC Factor VIII, Factor IX and Rare Coagulation Disorders
- Prof Joost Meijers, the Netherlands, SSC Factor XI and the Contact System
- Dr Verena Schroeder, Switzerland, SSC Factor XIII and Fibrinogen
- Ass Prof Maha Othman, Canada, SSC Women’s Health Issues in Thrombosis and Hemostasis

The 6 SSC will collaborate with the ClinGen/ClinVar teams at University of North Carolina (UNC), Chapel Hill, the National Centre for Biotechnology Information (NCBI), Bethesda and the Locus Reference Genomic (LRG) database team at EMBL European Bioinformatics Institute (EBI), Hinxton, UK
- Ass Prof Jonathan S Berg, USA, ClinGen hub leader at UNC and member of ClinGen steering committee
- Dr Melissa Landrum, USA, ClinVar Scientist at NCBI and member of ClinGen steering committee
- Dr Fiona Cunningham, UK, Leader Variation Annotation Team at EMBL EBI

Description Abstract
This project will lead to improvements in the care of patients by increasing the reliability of reaching a molecular diagnosis for inherited bleeding, platelet and thrombotic disorders (BPTD). Reaching a conclusive molecular diagnosis is important because it informs (i) the choice of treatment (e.g. gene correction, factor replacement, stem cell transplantation, watch & see), (ii) disease prognostication (e.g. emergence of deafness, kidney failure, myelo- and lung fibrosis, malignancy) and (iii) may inform family planning. For the vast majority of BPTD diagnosis relies on cascade laboratory testing requiring patients to attend out-patients multiple times and often resulting in a ‘diagnostic odyssey’.¹ With the advent of next generation sequencing (NGS) the ability to reach a molecular diagnosis by a single test for all 97 known BPTD genes is within reach.²⁻⁵ The main objective of this project is to provide a framework of reference, which is freely available to deliver NGS-based diagnosis safely to the clinic. This will be achieved by intercalating the disciplines and expertise from the 6 participating ISTH-SSC and by strengthening the already existing collaborations with ClinGen, ClinVar and LRG teams. The proposed SSC-cross-cutting project is timely and has 3 main aims: (i) maintaining an evidence-based catalogue of known (so called TIER 1) pathogenic BPTD genes (i.e. genes encoding coagulation factors, platelet receptors for coagulation factors and more); (ii) delivering an integrated reference database of gold-standard disorder-causing DNA variants (V), which are clearly pathogenic (VCP), likely pathogenic (VLP) and of unknown significance (VUS); and (iii) delivering a ISTH community-based REDCap solution for the safe-sharing of geno- and phenotype data between clinicians, scientists and the public to support international collaboration, which will be required to achieve the first two project aims. The solutions to be pioneered by this
SSC Subcommittee Project/Collaborative Project

project will result in the rapid and affordable DNA-based diagnosis, thereby bringing immediate substantial benefits to the estimated 2 million BPTD patients and their close relatives worldwide\(^2\).

The accurate categorization of DNA variants into the 3 groups of VCP, VLP and VUS is an essential requirement for realizing these clinical benefits. Historically the main reason preventing this of happening was the paucity of sequence data in large population cohorts to estimate the minor allele frequency (MAF) of a certain variant. This is changing rapidly with \(\sim 211,000\) DNA samples from cohorts already analyzed by Whole Exome or Genome Sequencing (WES/WGS) (gnomAD, WES,123K, WGS, 15,4K, http://gnomadexac.broadinstitute.org; TOPMed, WGS, 62,7K, https://bravo.sph.umich.edu/freeze5; NIHR Bio-Resource, WGS, 10K, http://bioinfodev.hpc.cam.ac.uk/web-apps/bridge/#home), which is set to increase to at least 1 million during the lifetime of the proposed project, thanks to large cohorts such as the UK Biobank, the 100 000 Genomes Project linking genotype obtained by WGS analysis of 0.6M samples to clinical and laboratory phenotypes retrieved from health and social care data. These initiatives provide for the first time ever a method to assign pathogenicity levels to variants using quantitative population-based metrics. This newly generated quantitative pathogenicity information will be used worldwide by Multi-Disciplinary Teams (MDT), typically consisting of a clinician, geneticist, scientist and bioinformatician to generate clinical NGS reports for BPTD patients\(^2\). The power of the Big Data revolution has already benefited approximately 5,000 BPTD patients whose DNA samples have been analysed by NGS\(^2\)\(^-\)\(^5\) and this revealed that \(\sim 50\%\) of the clinically reported VLP and VCP were hitherto unobserved and require therefore to be systematically catalogued. By fostering a culture of rapid REDCap facilitated data exchange about disease-causing variants and new BPTD genes, the ISTH community can remain the key driver of this cataloguing and will lead internationally on the labelling of DNA variants with pathogenicity status for BPTD. This will rapidly result in tangible benefits for patient care by reducing the turn-around time between sample taking and reporting and increased accuracy of clinical NGS reports. Typically information about variant pathogenicity status was maintained at the level of a specific or set of BPTD genes. Past and present members of 4 of the 6 SSC participating in this application have been responsible for initiating and maintaining some of the ‘gold-standard’ variant databases (e.g. for coagulation factors: EAHAD [http://eahad.org], for Glanzmann Disease: [https://glanzmann.mcw.edu], etc.).

Anticipating that NGS-based diagnosis for the **7,000 currently known rare inherited diseases** will become routine over the next decade a Herculean task is required to achieve variant annotation to international guidelines, to safeguard that this information is integrated across rare diseases in a single database, which is sustainable and remains freely accessible. An ongoing challenge is to update the information in ‘real-time’ by incorporating new gene and variant discoveries. The NIH has recognized this challenge and called for urgent action supported with millions of new funding for the ClinGen initiative\(^6\), which aims to ‘clean-up’ the content of the ClinVar gene variant catalogue (https://www.ncbi.nlm.nih.gov/clinvar). This is being achieved by establishing ClinGen panels of clinician-scientists and by linking these gene experts with ClinVar bioinformaticians at the NCBI in Bethesda and LRG bioinformaticians at EMBL-EBI in Hinxton (UK). One of the recently established and first ClinGen panels will be focusing on hemostasis and thrombosis and is coordinated by the ClinGen hub at UNC, Chapel Hill (Jonathan S Berg, UNC; KF and WHO are invited members of this ClinGen group).

The main objectives of the project are not only to provide framework of reference to improve NGS-based diagnosis but also to facilitate counselling and care delivery. Altogether this project will deliver a network connecting healthcare professionals with patients, exemplars of how to build an efficient MDT and training modules in diverse aspects of genome-based precision medicine (specialized Training Sessions during ISTH congress [see below]) for the next generation of young clinicians, scientists and genetic counsellors.
SSC Subcommittee Project/Collaborative Project

Design and methodology

The proposed objectives will be delivered by 3 project aims:

1. Maintain an up-to-date database of TIER 1 (known) BPTD genes and aggregate NGS data from BPTD patients to aid the classification of DNA variants\textsuperscript{2-5},

2. Work with the ClinGen, ClinVar and LRG teams at UNC, NCBI and EMBL-EBI respectively to deliver pathogenicity variant classification for the current 97 BPTD genes, already ratified by the GinTH SSC\textsuperscript{6},

3. Develop an ISTH-REDCap application to facilitate the use of MatchMaker Exchange (which includes GeneMatcher and DECIPHER) for the safe sharing of data on VCP, VLP, and VUS, and ‘gold star’ VCP within the ISTH community and beyond.

To achieve these aims, the following actions are planned.

1. Variant Call Files (VCFs) from at least 6,000 BPTD patients (resulting from panel tests, WES, WGS) will be aggregated between NGS providers who have committed to share data under this application (BRIDGE-BPD, ThromboGenomics, GAPP, Mini-BRIDGE)\textsuperscript{2-5}. The ISTH SSC website will be used to place an ‘open invitation’ for other NGS-providers for BPTD to join the data aggregation initiative. Variants will be called from the aggregated VCF and variants passing filtering on population MAF will be reviewed by at least 3 MDT (providing training opportunities for young ISBT members), which will work in tandem to assign VCP, VLP and VUS status. The members of the 3 MDT will receive training to work according to the most up-to-date international guidelines for reporting NGS results.

2. We anticipate at least 1,500 VCP, VLP and VUS in the aggregated NGS data with 50% having been hitherto unobserved. We will work with ClinGen\textsuperscript{6} and ClinVar (Melissa Landrum) to deposit into ClinVar: i) the variant pathogenicity labels, ii) the number of patients in whom a particular variant has been observed, and iii) a minimal set of Human Phenotype Ontology (HPO) terms\textsuperscript{7}. We will agree with ClinGen/ClinVar teams the assignment of ‘star ratings (incl. gold star)’ to variants to indicate the accuracy level of the variant-appended pathogenicity status.

3. We will work with the ISTH office to develop and commission a REDCap application, which has as its primary aim to provide ISTH members with a simple-to-use doctor/scientist/nurse friendly gateway for rapid entry of geno- and pheno-type data from BPTD patients. The reason for developing a simplified data submission gateway under the ISTH auspices is that visiting NCBI and EMBL-EBI websites can be a ‘daunting experience’ for those not skilled in the art of genomics and bioinformatics. The REDCap solution will provide an intuitive and user-friendly application for the clinical community to enter single-patient geno- and pheno- type reports. In case of new variants or putative new genes, the application will provide ‘clickable’ options to set a ‘beacon’ for the particular patient with the purpose of identifying other patients matched for pheno- and geno- types. We will not develop the large number of applications which are used by GeneMatcher and DECIPHER, but provide the user with a simplified ISTH-branded access to these applications.

Study population of BPTD patients

NGS data on an estimated 5,000 BPTD patients has already been generated\textsuperscript{2-5} and this is expected to increase by at least 1,000 patients/year during the lifetime of the project. The following BPTD patient groups are eligible:

\textit{Inclusion criteria:} Bleeding & platelet disorders: Patients with a platelet disorder as demonstrated by ANY of the following: (i) Platelet count <100x10\textsuperscript{9}/L OR >400x10\textsuperscript{9}/L, (ii) Mean
Platelet Volume <6fL OR >12 fL, (iii) abnormal platelet function test, iv) abnormal platelet morphology OR v) any combination of the above. Any bleeding disorder of known and unknown molecular aetiology. Thrombotic disorders: Patients with (i) a laboratory abnormality indicating an abnormality of protein C, protein S, antithrombin and dysfibrinogenaemia, AND/OR (ii) a first thrombosis before the age of 40, OR (iii) where there is at least one first degree relative with thrombosis before the age of 40.

For all patients, there must be high likelihood of the condition being genetic, demonstrated by either early onset OR the presence of other affected family members OR the BTPD occurs in the context of a wider syndrome (e.g. neurodevelopmental, immunological, skeletal).

Exclusion criteria: (i) Use of drugs known to be associated with abnormal platelet phenotypes AND/OR bleeding OR thrombosis, (ii) High likelihood of autoimmune thrombocytopenia (ITP) or other autoimmune disorders associated with low platelet count, including HIV positivity OR thrombosis AND/OR (iii) Other medical conditions known to be associated with abnormal platelet count and volume, abnormal platelet function or increased risk of thrombosis.

Expected outcomes

Publications. The project will deliver at least one publication for each of the 3 following categories: (i) official SSC guideline in the Journal of Thrombosis and Haemostasis, (ii) review article and (iii) original article. In parallel knowledge will be distributed to the global community via the ClinVar and LRG databases and through REDCap-supported MatchMaker Exchange activities. Publications will be reviewed and approved by the members of the SC (see below) before submission.

Training Sessions. We will organize 1 training session/year (aligned with the ISTH main meeting) to educate the ISTH community, and specially its junior members, about genomics as becoming one of the drivers of progress in the delivery of precision medicine, which requires genomically-literate clinicians and scientists. These hand-on training sessions will focus on: (i) existing and new NGS technologies, (ii) statistical methods for the interpretation of NGS results, (iii) population-based genomics principles underlying variant pathogenicity annotation, (iv) the use of HPO terms, and (v) reporting of NGS results by a MDT. Sessions on medical ethics and genetic counselling will be organized in parallel for clinicians, specialized nurses and genetic counsellors and other healthcare workers providing care for BPTD patients and their close relatives. These training events will provide valuable opportunities not only in skills training, but also in the fostering of new collaborative interactions, encouraging networking and contributing to informed career development.

Project set-up and management, needed infrastructure and resources:

Steering Committee (SC). The project delivery will be overseen by the SC with one representative of each of the 6 participating SSCs and representatives from UNC ClinGen, NCBI-ClinVar and EBI-LRG in attendance. The SC will appoint a Chair as the responsible person to oversee the project delivery. Decisions by the SC will generally be reached by consensus but if there is no consensus then a simple majority vote will suffice (non SSC members will have no vote). The SC will review progress of the project against the 3 agreed aims (see above) and milestones.

**SSC Subcommittee Project/Collaborative Project**

**Infrastructure (IS).** There is an already an informatics strategy agreed between the members of the SC, which is characterized by ‘preventing the development of new compute applications’ but to use and possibly modify existing applications, which are characterized by interoperability, long-term sustainability, already being used in other clinical and genomics communities and free of license/access fees. REDCap will be used for integration and tailoring of the applications to the need of the ISTH community of clinicians, scientists, nurses, patients and the public at large. The project can make a rapid start because it will be based on a solid foundation of pre-existing knowledge and past deliverable by the 6 participating SSC. If funding is awarded the project will be launched at the ISTH-SSC meeting in Dublin 2018. The project will last 36 months and the finalization and reporting will be through the delivery of at least 3 publications.

**Budget.** A grant total of $49,800 is requested over three years (2018-2020) with 65% being committed to clinical bioinformatics staff capacity for data aggregation, variant deposition into ClinVar and REDCap development and the remaining 35% is earmarked for the Training Sessions (+ travel awards for young ISTH members on a competitive basis) and ‘free access costs’ for the 3 publications.

**References**


