Published in final edited form as:
Hum Genet. 2011 July ; 130(1): 149–166. doi:10.1007/s00439-011-0984-y.

Delineating the Hemostaseome as an aid to individualize the analysis of the hereditary basis of thrombotic and bleeding disorders

Kim Fechtel and **Marika L. Osterbur**
3rd Millennium Inc., 395 Totten Pond Road, Suite 201, Waltham, MA 02451, USA

Hildegard Kehrer-Sawatzki
Institute of Human Genetics, University of Ulm, Albert-Einstein-Allee 11, Ulm, Germany

Peter D. Stenson and **David N. Cooper**
Institute of Medical Genetics, School of Medicine, Cardiff University, Heath Park, Cardiff CF14 4XN, UK

Abstract

Next-generation sequencing and genome-wide association studies represent powerful tools to identify genetic variants that confer disease risk within populations. On their own, however, they cannot provide insight into how these variants contribute to individual risk for diseases that exhibit complex inheritance, or alternatively confer health in a given individual. Even in the case of well-characterized variants that confer a significant disease risk, more healthy individuals carry the variant, with no apparent ill effect, than those who manifest disease. Access to low-cost genome sequence data promises to provide an unprecedentedly detailed view of the nature of the hereditary component of complex diseases, but requires the large-scale comparison of sequence data from individuals with and without disease to deliver a clinical calibration. The provision of informatics support remains problematic as there are currently no means to interpret the data generated. Here, we initiate this process, a prerequisite for such a study, by narrowing the focus from an entire genome to that of a single biological system. To this end, we examine the `Hemostaseome,' and more specifically focus on DNA sequence changes pertaining to those human genes known to impact upon hemostasis and thrombosis that can be analyzed coordinately, and on an individual basis, to interrogate how specific combinations of variants act to confer disease predisposition. As a first step, we delineate known members of the Hemostaseome and explore the nature of the genetic variants that may cause disease in individuals whose hemostatic balance has become shifted toward either a prothrombotic or anticoagulant phenotype.

kfechtel@3rdmill.com.

Electronic supplementary material The online version of this article (doi:10.1007/s00439-011-0984-y) contains supplementary material, which is available to authorized users.

Introduction

Achieving the goal of personalized medicine, by incorporating an individual's fully interpreted diploid genome sequence into their electronic medical record, will require the application of multiple methodologies to describe the combined hereditary vulnerabilities and protective defenses which are of clinical relevance to that individual. The initial data gathering exercise requires the application of advanced sequencing technologies (Next Generation Sequencing, NGS), improved copy number variant detection using arrays and sequencing, as well as the development of standards for genotyping from genomic data, but will also require automated methods of genome sequence analysis. Given the large number of diseases currently known to have a causal genetic basis (whether exhibiting Mendelian or complex inheritance) that have been mapped to specific gene loci [4,012 different genes collated in the Human Gene Mutation Database, HGMD®, http://www.hgmd.org; March 2011 (Cooper et al. 2010)] and the large number of variants shown to be associated with these disease states [110,772 different variants logged in HGMD®, March 2011 (Cooper et al. 2010)], the associated genotypic and phenotypic information should be presented in a disease-specific context in order to be utilized effectively. However, automated methods for comprehensive genetic analysis with interpretation and reporting of clinically relevant results for multiple individual genomes are yet to be described.

Early in 2010, the first example appeared of a clinical case report that included a personal genome analysis (Ashley et al. 2010); more are now following apace. However, the methods employed for individual analysis and interpretation in the above example would be very costly to apply to hundreds or even thousands of individual genomes. This conundrum has been well articulated by Mardis in her recent editorial, “The $1,000 genome, the $100,000 analysis” (2010). In addition, across the many individual genome sequences and analyses published to date, 2–3 million variants are consistently identified in each individual genome examined, while only a handful of those variants reported in these studies are likely to have clinical impact; clearly comprehensive analysis, in which the full spectrum of identified variants are placed into their biological context, has not yet been achieved.

Currently employed DNA analysis pipeline methods allow for the comparison of a test genome sequence against the reference human genome sequence from which variants can be readily identified. However, these methods do not support the partitioning of these identified variants and their respective genes into pathways where their combined impact can be assessed comprehensively. Also lacking are validated collations of subsets of related genes supporting comprehensive genetic analysis of targeted biological systems in individual genome sequences. In addition, there are no automated methods that interrogate the scientific literature for the current state of knowledge of the pathogenic or benign nature of variants that have been previously studied. Thus, the present state of interpretation of the impact of the readily identified variants found in a novel individual genome sequence is presently a rather laborious, manual and incomplete process, which limits the reach of these whole-genome analyses. As we enter a new era in which the comprehensive dissection of the genetic contribution to diseases with complex inheritance is now feasible, these become compelling requirements and represent an area where significant advances must occur before whole-genome sequencing can deliver on its potential for broad clinical application.

In this review, we have initiated the process of compiling the requisite information to develop a software system capable of the comprehensive, automated analysis of an individual genome with respect to the genes that comprise the “Hemostaseome,” the genetic repertoire responsible for hemostasis and thrombosis. This gene set encodes the proteins required for normal hemostasis and hence, by implication, the proteins underlying a variety of different bleeding and thrombotic disorders. These genes/proteins may also influence, albeit indirectly, various cardiovascular and other diseases that are complicated by thrombosis. The gene set includes the presently characterized genes in the human genome that are known to encode proteins with functions in hemostasis, or which yield bleeding or thrombotic phenotypes when mutated. Here, we (i) compile the relevant gene set by reference to the scientific literature, (ii) review the current knowledge base that supports each gene's inclusion in the Hemostaseome, (iii) collate the gene mutations and polymorphic variants in genes already known to be associated with hemostatic or thrombotic disease and (iv) propose, on the basis of recently published data, that copy number variations involving these genes are very likely to play an important accessory role in modulating the hemostatic balance at the level of the individual. This effort has been undertaken to lay the foundation for the development of a prototype software system that will support such an in-depth analysis, the focus of our current work that will be the subject of a future report.

Although concerns have been expressed that such an attempt at comprehensive genetic analysis of whole-genome sequence may simply result in the advanced study of the `incidentalome' (Kohane et al. 2006), the potential for such studies to stratify patients with significant hemostatic disease more accurately into specific treatment classes by comprehensively characterizing their individual genetic repertoire seems to provide ample justification for the pursuit of this ambitious goal. Moreover, a comprehensive, all-inclusive approach has the potential to reveal the complex interplay of heritable variants within a biological system. It is currently quite impossible to predict whether the repertoire of hereditary variants in individuals with thrombotic or bleeding disorders will be quantitatively or qualitatively different from those of healthy controls. It may well be that refined models will be required to distinguish individuals in whom hemostasis is chronically out of balance (due to hereditary factors) from those in whom potentially deleterious variants are present but fail to shift the hemostatic balance sufficiently to cause disease. It is also possible that, when considering the entirety of an individual's genome sequence, clinical significance will simply not be discernible against the background noise of polymorphic variation. This notwithstanding, distinguishing these different possibilities will necessarily require automated tools, since the required studies cannot be completed using current manual methods of analysis owing to the laborious nature of the task. We therefore seek to develop automated approaches that should provide clinically relevant interpretation of individual genome data within a systems context. By reducing the scope of the query to a single gene set, the feasibility of developing automated systems becomes greater. Hence, the foundations can be laid to allow the future goal of fully automating the comprehensive reporting of clinically relevant genomic data on an individual basis to be realized.

Hemostasis and thrombosis as a pilot biological system for genomic analysis

We selected the "Hemostaseome" as the pilot biological system for automated genetic analysis because of its advanced state of functional characterization (at genetic, biochemical, molecular and cellular levels), and also because of the broad clinical applicability of the results. Thus, for example, the availability of an automated method to assess comprehensively the genetic status of the Hemostaseome can be (i) applied to investigate the full repertoire of genes that can influence anticoagulant (warfarin) dosing, (ii) used to characterize the combination of hereditary factors present when recurrent thrombotic events occur in patients with thrombotic disease, and (iii) employed to determine the full profile of genetic variants in individuals with clinically significant bleeding disorders. Comprehensive assessment of variant status in an individual may provide the opportunity to determine the net effect on the hemostatic balance of multiple genetic variants with potentially opposing (or synergizing) effects. Within the context of whole-genome medicine, such approaches are likely to provide a much more detailed and potentially informative view of the hemostatic balance in a given individual than conventional laboratory phenotypic tests such as the activated partial thromboplastin time or prothrombin time.

Given that over 2 million patients initiate warfarin treatment each year in the USA (see Daly 2009 for review), and that over 2.5 million US women with menorrhagia are estimated to have underlying bleeding disorders (ACOG Committee Opinion 2009; National Hemophilia Foundation: http://www.projectredflag.org/), these technologies, combined with structured information for the Hemostaseome required for automated analysis, have the potential to impact upon a large number of patient groups, far beyond simply advancing our understanding of basic physiological mechanisms.

For the 18,000 US patients with hemophilia, the most severe clinical complication is the formation of antibody inhibitors, which results in a decline in both the length and quality of life and impacts upon 25–33% of these patients. The single strongest predictor of which patients will produce inactivating antibodies (that render their coagulation factor replacement treatment ineffective) is the nature of the specific gene mutation they harbor (Astermark et al. 2008; Zhang et al. 2009). Routine assessment of individual mutations in these patients could therefore lead to the identification of those patients who should utilize early bypass treatment to avoid the generation of inhibitors. The practical utility of this approach was recently demonstrated by performing genomic sequencing for ten hemophilia A patients, which readily identified causal *F8* mutations for six of the ten individuals examined (Pelak et al. 2010).

Millions of patients could also potentially benefit from low-dose aspirin therapy to prevent cardiovascular disease. Consensus recommendations for the use of aspirin, to prevent myocardial infarction in normal men aged 45–79 years and to prevent ischemic stroke in women aged 55–79 years where the risk of the disease exceeds risk from gastrointestinal bleeding, suggest broad clinical utility where the risk of bleeding is low (US Preventive Services Task Force 2009; Wolff et al. 2009). However, at present, sensitive tools to stratify patients into high- and low-risk categories are not yet available, even though such tools would allow those likely to experience the greatest benefit to benefit from this low-cost

therapy. Similarly, utilizing comprehensive individual genome analysis for patient stratification to assess thrombotic risk has the potential to allow the application of effective therapies to the correct patient class with the aim of preventing cardiovascular disease.

In addition to the broad clinical applicability of this gene set, various additional characteristics make it ideally suited to our purposes: the Hemostaseome is of manageable size since this set currently comprises 109 gene members. The clinical phenotype is also relatively simple (and binary), with polymorphic variants and/or mutations shifting the pathway either in a prothrombotic or an anticoagulant direction, thereby making clinical application simpler. In the vast majority of described cases, patients suffer either unexpected or uncontrolled bleeding or a tendency toward thrombosis, but not usually both (Girolami et al. 2005). Thus, beginning with the Hemostaseome as a “proof-of-concept” gene set, we describe the members of the set, the criteria employed to select them and the spectrum of currently understood means through which variations in these genes impact hemostasis to cause human disease.

The necessity of a systems approach

We anticipate that a systems approach to genetic analysis will lead to advances in our understanding of the combinatorial role of hereditary factors in contributing to hemostatic disease. For both thrombotic and bleeding phenotypes, individual variants with significant clinical impact have been identified that contribute to disease risk [e.g., factor V Leiden (R506Q) causing activated protein C (APC) resistance (Bertina et al. 1994) and prothrombin (G20210A) affecting mRNA function (Poort et al. 1996)], whereas a missense mutation Tyr1584Cys (4751A>G) in the von Willebrand factor (*VWF*) gene has been identified in 14.3% of Canadian families with type 1 von Willebrand disease (O'Brien et al. 2003); von Willebrand disease is a common bleeding disorder present in nearly 1% of individuals in populations investigated. Although each of the above examples displays a statistically significant association with disease, all are characterized by incomplete penetrance. Conventional genome-wide association (Arellano et al. 2010; Gohil et al. 2009; Lotta 2010; Smith et al. 2009b) and case–control studies (Mannuccio Mannucci et al. 2010; Smith et al. 2009a) have identified many other heritable factors that contribute to an increased risk of thrombosis, but these factors so far only account for 1–10% of the measurable risk (Yang et al. 2007). Hence, the true hereditary basis of both thrombotic and bleeding disorders (at least those which do not resolve to one of the characterized single-gene hemophilias) is still largely enigmatic. Although next-generation sequencing and genome-wide association studies have the power to identify variants and mutations that confer substantial risk, these methods provide no information as to which hereditary risk factors in which combinations are likely to result in disease in a specific individual.

In diseases for which the genetic basis is complex, phenotypic modification probably occurs in a given individual through the combined action of a spectrum of genetic variants including well-characterized risk factors as well as common and rare variants of lower impact. We propose that the myriad currently known hereditary factors, along with novel mutations that confer disease predisposition, act in various combinations and permutations so as to influence the clinical phenotype even though each variant may not individually

attain the level of statistical association required to be regarded as an independent risk factor using current epidemiological methods. Nonetheless, we posit that in the presence of multiple variants of low impact, the coagulation pathway could be shifted further toward either the anticoagulant or procoagulant (prothrombotic) end of the hemostatic spectrum. We anticipate that these variants may not achieve standard statistical metrics in population studies, either because their individual contribution is small and current studies are underpowered to determine their impact, or because their expression is also influenced by a variety of different environmental factors and hence they do not achieve the required level of statistical significance as independent variables. We nevertheless anticipate that they will each be capable of making a small contribution to disease predisposition. Since the number of such mutations and polymorphisms that can potentially impact upon the hemostatic balance in any given individual is likely to be quite large (see Tables 1, 2, 3, 4, 5, 6, 7, 8), the dissection of their contributions in population studies (where the number of possible permutations of variant profiles is going to be very high) will require a range of different analytical methods. We therefore propose that the systems analysis of individual genome sequences will be necessary to tease out the contribution of hereditary factors to thrombotic and bleeding disorders as even for relatively common variants with significant impact (e.g., factor V Leiden), individual expression of thrombotic phenotypes among heterozygous carriers is highly variable (Samama et al. 1995) and double heterozygosity or homozygosity is insufficient to the task of accounting for a predisposition to recurrent thrombosis (Lijfering et al. 2010). We also anticipate that, even in the presence of multiple prothrombotic variants that would otherwise be expected to result in a procoagulant phenotype (a `hypercoagulable state'), the simultaneous presence of anticoagulant variants could potentially shift the hemostatic balance back toward equilibrium. Thus, we postulate that the comprehensive genetic analysis of a biological system, such as the Hemostaseome, will have the potential to reveal recognizable patterns of variants that could account for the presence or absence of disease in individual cases once the standard is calibrated by clinical studies. This would then potentiate comprehensive, automated, genome-wide individual studies (GWIS) that could serve to complement the more familiar genome-wide association studies (GWAS), thereby extending a novel and potentially very powerful tool for the assessment of multifactorial inheritance as expressed in an individual genome and correlated with an individual clinical phenotype. Thus, at least for the genes and variants that contribute to the hemostatic balance, we anticipate that the common disease–common variant hypothesis, as well as the common disease–rare variant hypothesis will both apply (see Schork et al. 2009 for review).

Finally, the necessity of a systems approach derives from the limiting factor in achieving automated analysis—that of developing study annotations from the scientific literature. Although DNA sequence and original citation(s) are available through the HGMD® database, information on study size, clinical significance, laboratory demonstration of functional polymorphisms, replication studies and information on study populations must be developed for each variant. By reducing the focus from an entire genome to a distinct physiological system, this task can be carried out efficiently and then applied to multiple genomes, thereby permitting the automation of larger-scale analyses, with the focus being firmly placed on a specific biological system.

Genetic mutations, variations and hemostasis

The majority of the genetic variants that influence hemostatic and thrombotic phenotypes, which have so far been characterized at the DNA sequence level, comprise single-nucleotide mutations/polymorphisms resulting in missense/nonsense mutations, or small micro-insertions and micro-deletions which serve to disrupt the reading frame. For most of these variants, sequence data, phenotypic annotations and original literature citations are available from the Human Gene Mutation Database (HGMD®) in a structured manner that is compatible with automated analysis. However, the large number of variants, 5,692 at the present count (August 2010), identified within the genes of the Hemostaseome, precludes manual analysis as a means to compile a comprehensive assessment of variants within an individual genome. At the time of initial assessment, only 188 of these variants were present in dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP); hence, to consider variants known to play a role in hemostasis, we utilized the mutation data compiled in HGMD®.

The potential impact of copy number variation

Although hereditary factors causing hemostatic disease have been described for over 100 years (e.g., hemophilia B in the male descendants of Queen Victoria; Rogaev et al. 2009) and their molecular characterization achieved within the past 30 years, it has only been appreciated very recently that copy number variations are a common form of structural variation in individual genomes and should also be considered for their potential to impact upon hemostatic disease (Lee and Scherer 2010). Copy number variations (CNVs) are defined as DNA sequences greater than 1 kb, the presence of which varies by reduced or increased numbers of copies in a given genome. In some cases, CNVs encompass entire genes or multiple adjacent genes and hence their impact and nature do not differ much from conventional gross deletions or duplications.

Multiple studies have demonstrated that elevated levels of coagulation factor proteins are associated with thrombosis, whereas their deficiencies contribute to a variety of bleeding disorders. With the report that the factor VIII (*F8*) gene was duplicated in the genome sequence of J. Craig Venter, as detected by Illumina HumanHap650Y BeadChip arrays (Levy et al. 2007), we began to explore whether other genes of the Hemostaseome might also exhibit copy number variations that would result in potentially significant gene duplications or deletions. Given the wide range of protein levels ascertained for the different coagulation factors and related proteins in healthy populations, and the clinical significance of elevated levels of certain proteins that correlate with thrombotic disease, CNVs could represent an important additional heritable element of combinatorial risk for either bleeding or thrombotic disorders, a possibility that is only just beginning to be explored (see Kamstrup 2010 for review; Kraft et al. 1996). Although CNV status has not been ascertained at the fine-structural level in most studies (or functionally characterized/fully sequenced at the genomic level) owing to the complexity of assembling sequence reads from multiple gene copies, it is likely that a complete gene gain or loss through CNV has the potential to impact significantly upon coagulation factor protein levels and, by implication, clinical phenotypes. Although the CNVs compiled from the literature have typically not been identified in individuals for whom medical records (including coagulation factor activity/antigen levels) are available, the presence of CNVs for so many members of the

Hemostaseome gene set (see Tables 1, 2, 3, 4, 5, 6, 7, 8) suggests that they have the potential to be major contributors to bleeding and thrombotic disorders. Although genome-wide association studies focused on traditional SNP analysis have successfully identified heritable factors contributing to increased risk of hematological phenotypes, these factors account for only 1–10% of the measurable risk (Yang et al. 2007). We propose that the inclusion of CNVs alongside both rare and common factors of low impact, and in combination with presently known independent risk factors that meet rigorous statistical metrics, will allow a more comprehensive description of the hereditary factors, the combined action of which leads to disruption of hemostasis in subjects with clinical disease, thereby potentially helping to resolve the issue of missing heritability (Clarke and Cooper 2010; Eichler et al. 2010; Manolio et al. 2009).

The availability of individual whole-genome sequence data and array-based CNV detection on a low-cost basis has now made possible the study of a reasonable number of subjects with extremes of phenotype and multifactorial heritability. Of particular interest would be subjects with recurrent thrombosis or recurrent bleeding whose pathology cannot be attributed to a single-gene mutation. The challenge facing us is therefore our limited ability to analyze the resulting data. By delineating the Hemostaseome, a targeted analysis can be undertaken, which essentially comprises a systems-based, multifactorial, candidate gene sequencing and analysis approach. By reducing the focus of the analysis to the 109 genes that we have selected, the comprehensive analysis of presently known mutations and variants becomes feasible for the automated analysis of tens to hundreds of individual genome sequences.

Through review of the literature, the Online Mendelian Inheritance in Man (OMIM) and the Human Gene Mutation Database (HGMD), we describe an initial set of 109 genes that together comprise the Hemostaseome. Online Resource 1 presents the initial selection information utilized to compile this gene set. Online resource 2 provides a detailed narrative considering each gene and discusses evidence from the literature associating each gene with hemostatic disease. In addition, a description is provided for each gene of currently described copy number variants reported from structural studies of individual genomes. Online Resource 3 summarizes data from the literature with specific citations demonstrating elevated protein levels of the proteins encoded by these genes and the association of these elevated protein levels with hemostatic disease. This detailed information is summarized in Tables 1, 2, 3, 4, 5, 6, 7, and 8.

Discussion

We have compiled a set of 109 genes comprising the `Hemostaseome' that function in a variety of pathways to bring about hemostasis. The majority of these genes have already been linked to hemostatic disease either through association studies, case–control analyses or by sequencing studies of patients with Mendelian disorders. Variants and mutations have been described that also show disease association and for which DNA sequences are available in HGMD®. Many of these variants and mutations are associated with phenotypes that are inherited in Mendelian fashion, whereas others have been implicated in diseases with complex inheritance (e.g., cardiovascular disease and thrombosis). The compilation of

these genes into a cohesive set with attendant genomic information, DNA sequences, literature citations and annotations permits the structured organization of this information, thereby allowing the comprehensive assessment of individual genome sequences for the presence/absence of these disease-associated variants so that the automated assessment of individual genome sequence can be performed rapidly for multiple genomes. This enterprise offers an opportunity to begin to study, on a larger and more comprehensive scale, diseases with either Mendelian or complex inheritance patterns utilizing individual genome sequence to understand how a constellation of genetic variables contributes to the hemostatic balance in a given individual and to diseases of hemostasis when that balance is lost. This effort also sets the stage for what will be required, from an information perspective, to fully analyze an individual genome sequence for all diseases and biological systems (including all variants presently characterized) and report this information from within a specific disease context. At the time of writing (August 2010), 5,692 distinct mutations and polymorphisms in the 109 genes of the Hemostaseome have been characterized in the research literature and compiled in HGMD®. The distribution of these variations across the genes of the Hemostaseome is annotated in Tables 1, 2, 3, 4, 5, 6, 7, and 8.

Impact of copy number variation on the genes of the Hemostaseome

Copy number variations have been commonly reported for the members of the Hemostaseome gene set. Full gene gains would be expected to be equivalent to a duplication of the gene and hence, in all likelihood, to have the potential to result in an increase in protein level; we found 12 reported cases of complete gene gains among the set of 109 genes. For many of the genes of the Hemostaseome, elevated protein levels are a well-characterized risk factor for thrombosis or cardiovascular disease. Thus, in principle, either regulatory mutations or complete gene gains could underlie the heritable elevation of protein levels. For partial-gene gain copy number variations, the impact of which is more difficult to predict, we also found reported 12 examples. A recent report of a partial duplication of *PROS1* associated with a disease-associated deficiency phenotype (Pintao et al. 2009) suggests that partial duplication can constitute a significant mutational mechanism. Complete gene loss copy number variations were also common with 20 genes of the Hemostaseome exhibiting complete loss CNVs; these variants would be expected to be identical to deficiencies in terms of their consequent phenotypes. For the majority of these genes, a significant deficiency phenotype associated with disease has been characterized, either in human patients or in knockout mouse models. Thus, it is likely that complete gene loss would result in a clinically overt protein deficiency from the affected allele. Finally, partial-gene loss copy number variations were also found to be common, with 12 genes showing partial loss involving at least one exon. In such cases, these variants would also be likely to result in a deficiency phenotype depending on the number and relative positions of exons that are deleted.

Thus, from this analysis, we conclude that copy number variations for the gene set comprising the Hemostaseome are not infrequent occurrences: indeed, on the basis of these recent studies, 45 of 109 (41.3%) gene members of the Hemostaseome have been found to exhibit copy number variations and these studies have typically screened apparently healthy individuals (specific examples are noted in Tables 1, 2, 3, 4, 5, 6, 7, 8). This is significantly

higher than the 12% of genes in OMIM™ that overlap with validated CNVs (as observed by Conrad et al. 2010). These results are corroborated by HGMD® gross deletion data (Tables 1, 2, 3, 4, 5, 6, 7, 8 del); 31 of 109 (28.4%) genes of the Hemostaseome have disease-associated gross deletions described (see HGMD® mutation data for individual citations). Thus, it is clear that gross structural variations involving the genes of the Hemostaseome are not infrequent occurrences. Indeed, they have the potential to exert a very significant inXuence on hemostatic phenotypes and hence should, in the future, be routinely assessed in comprehensive studies investigating genetic factors contributing to hemostatic diseases.

At present, our understanding of the impact of copy number variations with regard to disease association is fairly rudimentary. High-resolution breakpoint mapping is a prerequisite for the accurate assessment of CNV size, the identification of the genes and associated regulatory elements affected, and hence for the determination of the consequences of copy number variation for gene expression and the phenotypic *sequelae* (Beckmann et al. 2008; de Smith et al. 2008). Fewer than 10% of known germline CNVs have been mapped to the single-nucleotide level (Conrad et al. 2010). Owing to the complexity of genome assembly, we cannot usually make inferences about the likely consequences for gene expression from a duplication CNV, the precise chromosomal localization of which is unknown (as is the case for CNV gains detected in array studies). Even where array data span the entire gene and are suggestive of a complete duplication, inactivating mutations may well be present and the gene may therefore not be expressed. Alternatively, as recently demonstrated in the dog, copy number variation duplications can be due to retrotransposition of an mRNA rather than simple genomic duplication; in the cited case, the copy number variation involved the fibroblast growth factor 4 gene (*Fgf4*) and was caused by a retrotransposition event that gave rise to *chondrodysplasia* in various domestic dog breeds. The retro-transposed gene was shown to be expressed, resulting in a dominant dwarfism phenotype present in all dwarf breeds tested (Parker et al. 2009). That such mutational mechanisms occur in humans was confirmed by Conrad et al. (2010), who demonstrated retrotransposition to be a common mechanism generating CNV duplications that are consequently dispersed in the genome. Although the identi-fication of the presence of a CNV gain is readily accomplished with array studies, localizing the duplicated genetic material and determining its integrity have not been carried out in the majority of copy number variation studies completed to date. Thus, in most cases, the true impact of CNVs on gene/protein expression remains unknown. However, as illustrated by the examples noted above, CNV variants have the potential to account for significant (and heritable) impact that remains to be explored in the context of patients with a bleeding diathesis or thrombotic disease. Initial efforts have led to confirmation for such an approach. In Quebec platelet disorder (which causes a bleeding diathesis), a tandem duplication of *PLAU* is the likely causal mutation (Paterson et al. 2010). For deficiency of *PROS1*, a prothrombotic condition, both duplication and deletion CNVs are identified in point mutation-negative *PROS1* deficiency patients as the causal mutations (Pintao et al. 2009). Finally, expansion of Kringle IV domains through partial CNV in *LPA* (van der Hoek et al. 1993) is associated with a reduced risk of ischemic heart disease (Clarke et al. 2009; reviewed by Kamstrup 2010) indicating a positive attribute associated with an intragenic, partial CNV.

Further support for the importance of structural variation comes from studies that correlate the presence of CNV and gene expression levels with the finding that duplications and deletions have distinguishable expression ratios; Stranger et al. (2007) showed that 17.7% of gene expression variance was attributable to CNVs. Lockwood et al. (2008) integrated parallel gene expression profiles with array comparative genomic hybridization data for genes within amplification hotspots in 27 NSCLC cell lines and showed that approximately 50% of the amplified genes were expressed at a significantly higher level as a consequence of the increased dosage. Schuster-Bockler et al. (2010) replicated those findings, associating expression levels with the presence of CNVs reported in earlier studies but observing a more continuous distribution of expression levels rather than the 0.5, 1.0 and 1.5 ratios expected per copy number increase. Thus, there is already compelling evidence that CNVs influence clinical pheno-types as well as contributing to differences in gene expression. However, as stated earlier, for most studies conducted, breakpoint definition and detailed delineation of the variant at the DNA sequence level have not been performed and follow-up studies indicate that the actual CNV size may be smaller than originally reported—as found when higher-definition arrays are utilized in replication studies (Perry et al. 2008). Nonetheless, the presence of the CNVs is generally replicable; what remains to be delineated is the exact size and specific overlap with genes at the nucleotide sequence level. The studies completed to date are consistent with the view that the gene members of the Hemostaseome exhibit a greater number of overlapping CNVs than is typical for the genome as a whole, and that these variations offer the potential to lend further understanding to the still largely enigmatic hereditary basis of hemostasis and thrombosis.

Combinatorial impact of hemostatic variants

Given the large number of genes in which mutations that result in hemostatic disease can occur, the substantial numbers of conventional mutations and variations characterized to date, and taking copy number variation into consideration, it is likely that individuals with hereditary thrombophilia or a bleeding diathesis will demonstrate a complex spectrum of variations in their genome sequence. It has already been demonstrated that the risk of venous thromboembolism (VTE) is increased in patients who carry more than one genetic variant disrupting genes of the Hemostaseome. In one study, 19% of symptomatic individuals harboring a protein C (*PROC*) gene mutation were also heterozygous for factor V Leiden (Koeleman et al. 1994). In a replication study, 9.5% of symptomatic patients were found to carry both mutations (Gandrille et al. 1995). Similar findings were made in families with protein S (*PROS1*) deficiency; among symptomatic individuals, 38% also carried the factor V Leiden mutation (Koeleman et al. 1995). In a larger-scale study involving 132 thrombophilic families, the risk of thrombosis was increased and the age of onset was lower in cases where co-heterozygosity for two gene variants was observed (combinations of variants in *PROC*, *PROS1*, *F5* and *F2*) over that recorded for individuals carrying single variants of these genes (Tirado et al. 2001). In another study in which multiple thrombophilic factors were investigated, there was a trend for multiple thrombophilic factors to be associated with adverse pregnancy outcomes (Larciprete et al. 2010) with specific combinations showing a higher association with specific outcomes (e.g., *MTHFR* C677T and PAI-1 5G/5G, *SERPINE1*, associated with intrauterine demise). In a case study where three family members exhibited overt thrombosis (mesenteric-vein, ileo-femoral and deep-

vein thrombosis), multiple members were found to be co-heterozygous for factor V Leiden, the factor II G20210 mutation and a protein S deficiency (Gonzalez et al. 2003). For case studies of thrombophilia in infancy, five subjects were found to be heterozygous for both factor V Leiden and a methylenetetrahydrofolatereductase (*MTHFR*) variant while two cases were heterozygous for both factor V Leiden and a factor II variant (Koren et al. 2003). These patients had experienced either cerebrovascular events or thrombosis associated with femoral catheterization. In an additional family study, anti-thrombin deficiency (*SERPINC1*), factor V Leiden and hyperhomocysteinemia (*MTHFR*) contributed to thrombosis among different family members (Gemmati et al. 1998). Finally, a family study in which the proband died perinatally due to bilateral renal vein thrombosis as well as thrombosis in the major cerebral veins demonstrated that (i) both protein S and protein C deficiencies were transmitted from the parents and (ii) the proband was heterozygous for a protein C deficiency and compound heterozygote for two distinct protein S (*PROS1*) mutations (Formstone et al. 1996). A second pregnancy was shown to exhibit only isolated protein C deficiency and hence resulted in term delivery without the complication of venous thrombosis. Although prenatal diagnosis was undertaken to permit direct mutation detection in a third pregnancy, this infant also developed renal thrombosis in utero, although it harbored only the paternal *PROS1* and *PROC* mutations. Thrombosis in the third infant was considered by the authors to be due to the presence of unknown additional variants or due to a `hypercoagulable' maternal environment.

Although such case studies provide strong circumstantial support for the combinatorial impact of multiple mutations in thrombotic disease (Corral et al. 1999; De Stefano et al. 1999; Gonzalez Ordonez et al. 1999), ultimately they do not address the full repertoire of variations in affected individuals as they query only a small number of the most common variants. In addition, they do not identify individuals in whom the same variants are present without the manifestation of disease. It is likely that a large number of currently characterized thrombophilic variants act combinatorially to effect disease. This postulate is supported by additional reports that characterize the involvement of other genes. For example, co-heterozygosity for heparin cofactor II (*SERPIND1*) and factor V Leiden or protein C deficiency was observed in subjects who experienced juvenile thromboembolic episodes (Bernardi et al. 1996), as well as in other families with thrombosis tested for these variants (Lane et al. 1996; Sansores-Garcia and Majluf-Cruz 1998). Two sisters in one family study were found to be homozygous for a factor XII deficiency-associated mutation (*F12*) and heterozygous for factor V Leiden; one sister exhibited multiple thrombotic events, whereas the other was asymptomatic (Girolami et al. 2010) suggesting that there is significantly greater complexity underlying disease predisposition than just the presence of two or three variants. Finally, the risk of cardiovascular disease was increased significantly in the presence of both decreased *ADAMTS13* protein levels and elevated *VWF* protein levels above that contributed by either change alone (Bongers et al. 2009). Thus, there is already compelling evidence to support the view that multiple hereditary factors can lead to a substantial increase in the risk of hemostatic disease. Hence, assessing the full repertoire of variants present in individual genomes is likely to greatly clarify our understanding of disease predisposition in individual cases.

It should also be appreciated that the close physical proximity of some members of the Hemostaseome can contribute to the potential for contiguous gene disruption. Indeed, two of the CNV loss variants described involve pairs of adjacent hemostatic genes. Thus, for the 13q34 CNV loss reported by Wong et al. (2007), both *F7* and *F10* were deleted, whereas for the 1q24.2 CNV loss reported by Conrad et al. (2010), both *F5* and *SELP* were deleted. At present, the clinical impact of these variants is unknown as they were identified in apparently healthy volunteer subjects. However, CNVs that involve adjacent genes offer an additional mechanism through which multiple variants can be transmitted and can contribute to disease, or could potentially compensate for other thrombotic variants.

Given the difficulty of assessing multiple mutations/variants in a large number of genes without genomic approaches, the new advanced sequencing technologies offer the potential to carefully interrogate a large swathe of the genome for potentially interacting mutations. Given the likelihood that recurrent thrombosis will be associated with a greater burden of genetic risk, our emerging ability to identify patients at highest risk, who would benefit maximally from extended anticoagulant therapy, offers the possibility to carefully stratify patients across risk/benefit categories. In addition to additive or synergistic hereditary risk factors, the potential also exists to establish the presence of compensatory mutations; for example, the simultaneous presence of deficiencies for coagulation factors could potentially reduce the risk of thrombosis, a possibility that has not yet been fully explored. In support of this possibility is a preliminary report by Salomon et al. (2011), which noted that patients with severe factor XI (*F11*) deficiency exhibited a reduced incidence of deep-vein thrombosis. In addition, meta-analysis of the evidence supporting a protective role for the factor XIII Val34Leu variant revealed a small but significant protective effect associated with reduced occurrence of VTE (Wells et al. 2006). This does not, however, appear to be generally the case for variants that cause bleeding disorders. For example, factor VII (*F7*) deficiency does not appear to protect against thrombosis in those with other prothrombotic variants (Astermark et al. 2001; Mariani et al. 2003) and a case report of the occurrence of venous thrombosis in a woman with Hermansky–Pudlak syndrome following cardiac catherization (preceded by DDVAP treatment to correct abnormal bleeding) suggests that thrombotic risk may remain in patients with bleeding disorders (Katz and Shani 2005). Therefore, compensatory variants, if they exist, will in all likelihood exert their effect only in very specific combinations. However, given the present state of the study of combinatorial risk for thrombosis with the presence of two or more variants substantially increasing the risk of disease, we can presently envisage the potential value of more comprehensive individual genetic health assessment for the genes of the Hemostaseome in assessing risk and for developing an individual treatment plan for patients with thrombosis or a bleeding diathesis based on an understanding of the integrity of their individual hemostatic pathways.

A final enigma can be addressed by comprehensive assessment of variants in the genes of the Hemostaseome. Although statistically significant metrics have demonstrated disease association for the aforementioned variants associated with thrombotic disease, elevation of risk for each variant is individually low and incomplete penetrance is observed for all prothrombotic variants. This means that the vast majority of individuals bearing these

variants do not suffer from thrombotic disease. We anticipate, however, that individuals with recurrent thrombosis will tend to have a greater number of prothrombotic variants whereas those who have experienced a single thrombotic event will have significantly fewer variants, while those who have never experienced thrombosis will harbor even fewer prothrombotic variants. Because the number of genes that can have an impact on hemostasis is large and the number of variants with potential impact is larger still, we expect that a large number of different variant combinations and permutations will manifest significant disease association. This genetic risk will underlie each prothrombotic challenge (such as pregnancy, long haul air travel and immobilization after surgery) with a greater risk of thrombosis occurring in those with a greater number of prothrombotic variants in their genome. We also anticipate that where multiple variants impact the same point in a pathway, gene interactions may well be synergistic rather than additive.

At present, anticoagulant therapies are effective but exact a significant cost and burden from patients who must submit to routine management including frequent blood testing while not being able to avoid the occurrence of significant adverse events, such as episodes of uncontrolled bleeding. In a recent meta-analysis, Iorio et al. (2010) showed that the highest risk of recurrence of VTE occurred in patients in whom the first thrombotic event was unprovoked by surgery or other transient factors. Annual rates of reoccurrence of VTE for surgical patients were 0.7% per year, 3.3% for patients with transient factors, and 7.4% for unprovoked VTE. Similar findings were reported by Christiansen et al. (2010), who also confirmed a higher risk of recurrence among men than women. Although D-dimer assessment post-anticoagulation treatment stratifies those patients with higher risk of recurrence (positive for D-dimer: 8.8% recurrence per 100 patient years) from those with a lower risk (negative for D-dimer: 3.7% recurrence per 100 patient years; Douketis et al. 2010), it does not, however, clearly distinguish those patients who should continue indefinite anticoagulant treatment from those who should be able to discontinue anticoagulation after a finite treatment period. Comprehensive assessment of genetic predisposing factors among patients with unprovoked VTE offers the potential to identify those at greatest risk of reoccurrence for whom continued prophylactic anticoagulant treatment could be lifesaving. Among this subgroup, it would also be beneficial to identify those at lowest risk in whom treatment can be discontinued after 3 months. Thus, comprehensive genetic assessment offers the potential to stratify patients among this highest-risk group. To achieve this goal, a large-scale study would need to be undertaken in which the full spectrum of hemostatic variants are assessed and compared to control groups, e.g., those individuals who have never experienced thrombosis despite hemostatic challenge or those who have experienced a single event followed by subsequent years without reoccurrence. Such studies have the potential to establish and distinguish the highest-risk genotypes.

A case can be made for the future clinical utility of this approach. Testing variants in two genes associated with warfarin response has been demonstrated to reduce hospitalization rates during the dose-establishing first 6 months of treatment (Epstein et al. 2010) and the genes tested account for only 30% of the variance in warfarin response. Presumably, accounting for the remaining 70% would enhance this initial and very promising success. Population studies have consistently shown significant risk attributable to specific hemostatic variants, but the goal of developing informative diagnostic tests has remained

elusive. Meta-analysis has confirmed a significant role for many gene members of the Hemostaseome in venous thrombosis (Austin et al. 2011; Castaman et al. 2003; Gohil et al. 2009), but there is still no comprehensive testing panel available clinically. Meta-analysis of recurrent thrombosis in children has confirmed increased risk associated with deficiencies of protein C, protein S, antithrombin, the factor II variant and more than two genetic traits (Young et al. 2008). However, <10% of children with VTE test positive for these specific mutations, suggesting that those with unprovoked VTE probably harbor other mutations. With the prospect of greatly reduced sequencing costs and the potential for automated analysis, the tools are on the near horizon with which to determine the basis of genetic susceptibility to diseases such as thrombosis that exhibit complex inheritance. The observation that unprovoked thrombosis has the highest recurrence rates suggests that these individuals bear the greatest genetic burden of risk; this postulate can now be investigated directly, accurately and comprehensively using genomic sequencing. Meta-analysis of 66,155 cases of coronary disease found that the factor V Leiden and prothrombin 20210A variants were moderately associated with risk of disease (Ye et al. 2006). Meta-analysis of low levels of protein Z showed significant association with arterial vascular diseases, while subgroup analysis revealed effects for pregnancy complications and venous thromboembolic disease (Sofi et al. 2010). Although meta-analysis of incidence of hemostatic variants in 24 genes replicated only the factor V Leiden association with venous thrombosis (Smith et al. 2009b), assessing only specific SNPs would be expected to under-report the true number of variants occurring in these genes in cases. Meta-analysis of 126,525 cases of VTE showed statistically significant associations with VTE for two factor V (*F5*) variants, two prothrombin (*F2*) variants, a single alpha-fibrinogen (*FGA*) variant and a regulatory mutation in PAI-1 (*SERPINE1*); protective effects were found for the factor XIII (*F13*) Val34Leu and beta-fibrinogen (*FGB*) 455 G/A (Gohil et al. 2009). Meta-analysis of increased levels of LP(a) and reduced copy number was significantly associated with risk of coronary disease (Clarke et al. 2009). Thus, the benefit to be potentially derived from case–control studies where the entire spectrum of hemostatic variants can be interrogated simultaneously offers the promise of novel, highly multiplexed diagnostic testing. Given that thrombosis in children is frequently associated with cancer, autoimmune disease, congenital heart disease and the placement of intravenous lines, diagnostic screening tools can be utilized to monitor such high-risk children. Such tools then can also be applied in adult patients with elevated risk (family history, prior thrombotic event) to personalize their care.

Thrombosis is not the only potential arena for such diagnostic testing. Even for single-gene bleeding disorders, the genotype–phenotype relationship is not always clear. Among patients with von Willebrand disease, 1.5% were found to harbor multiple coagulopathies requiring different treatment strategies (Asatiani and Kessler 2007); secondary sites included at least seven different genes among these patients. Apparently, variations in the factor VII (*F7*) gene can ameliorate the phenotype in severe hemophilia (A/B) resulting in a milder phenotype (Jayandharan et al. 2009).

Taken together, our survey of characterized mutations and polymorphic variants that impact upon hemostasis, combined with an assessment of copy number variation, suggests that the essential information needed to comprehensively delineate a matrix of variants that

corresponds to the genetic factors underlying hemostatic disease in an individual genome can be fairly readily compiled and should be amenable to automated analysis. Hemostasis is an exquisitely balanced biological system with multiple, independent pathways responsible for activation and a myriad of regulators that constrain coagulation activity only to sites of vascular damage; in addition, a suite of metabolic processes underpins its action. Thus, with over 100 underlying genes, individuals with hemostatic diseases will exhibit a complex profile of deleterious mutations and variants that can together tip the hemostatic balance in either a prothrombotic or anticoagulant direction. We predict that the system as a whole is likely to be able to tolerate a significant degree of perturbation before disease manifests due to the depth of cross-regulation. Further, individuals with hemostatic disease are likely to harbor a significant number of variants that act in concert thereby contributing to disease manifestation. Even for those diseases attributed to single-gene heredity (e.g., the hemophilias and von Willebrand disease), the severity, clinical course and response to treatment will all be influenced by other variants (Asatiani and Kessler 2007; Berg et al. 1994; Ghosh et al. 2001); in cases of hemophilia where thrombotic variants are also present, thrombosis can result from treatment designed to control bleeding (Nowak-Gottl et al. 2003). We anticipate that the clinical application of “omics” technologies such as assessment of individual genome sequences (and perhaps also serum proteomes) are on the near horizon for medically important targets such as the Hemostaseome and that the application of these technologies on an individual basis will inform patient stratification and allow personalized treatment for a variety of hemostatic disorders.

Recent studies that associate predisposition for type 1 diabetes and an anti-viral gene expression network confirm that linking biological systems, including related pathways, with genetic variation provides a powerful framework within which diseases with a complex hereditary basis can be investigated (Heinig et al. 2010). Other studies support our thesis that a broad range of gene variants ought to be considered to fully explore the hereditary basis of complex disease. In a recent analysis of the quantitative trait of adult height using genome-wide association methods and meta-analysis, Lango Allen et al. (2010) estimated that variation at 697 different genetic loci contribute to 19.6% of height variability. Although genome-wide association studies can identify variants associated with diseases exhibiting complex inheritance, these methods cannot provide information about the factors whose combined effects in an individual create a foundation for disease. By delineating the genes involved in the biological system that comprises hemostasis, we are now poised to investigate the full repertoire of genes and variants that predispose to hemostatic disease in affected individuals. This notwithstanding, until such a time as we have the capability to fully interpret a complete individual human genome sequence with automated methods, we should simply delineate the critical components of discrete biological systems such as the Hemostaseome and then develop the requisite computational methods and knowledge base to support automated analysis of this important subset of the human genome.

Rapid progress is occurring in the application of NGS technologies for the identification of causal mutations underlying diseases that exhibit Mendelian inheritance (Lalonde et al. 2010; Ng et al. 2010a, b). Indeed, a selected gene set has recently been employed to screen for heritable alterations in eye diseases. However, even in this report where 46 candidate genes were specifically screened (and five mutations were identified) in 21 families with

autosomal dominant retinitis pigmentosa, the mutations identified were single-gene alterations, the impact of which on their own may be regarded as causal (Bowne et al. 2011). In diseases with Mendelian inheritance, we understand that a single mutation (or a single mutation with modifiers that can recalibrate the phenotype slightly, up or down) is the causal basis for disease. In diseases that exhibit complex inheritance, however, we have typically considered only a genetic predisposition underlying these diseases, because many people carry mutations that are identified as significant risk factors but which are insufficient for disease to develop. Given the observation that these alterations appear to be associated with disease but are not necessarily causal, we have not discerned what may be, in fact, the same underlying causality. Recent advances in network analysis suggest that the genetic basis of diseases that exhibit complex inheritance is also causal (Schadt 2009). This causality has been masked by the way we currently study disease inheritance in large populations, as these studies do not distinguish between individuals based upon their levels of functional impairment. Clearly, with Mendelian disease, sufficient impairment of a functional pathway occurs as a consequence of a single mutation resulting in disease. We now believe that individuals who manifest diseases with complex inheritance also have sufficient pathway impairment to directly cause disease; however, in these cases, impairment may arise via the additive or synergistic effect of many mutations. Thus, in diseases with complex inheritance, this threshold of impairment of pathway function has also been crossed, but through the combined action of many mutations. In individuals who harbor only a subset of these mutations, the requisite damage required for disease to manifest does not occur, or does so only with the addition of contributing environmental factors. This new view of complex inheritance suggests that comprehensive genetic analysis of pathways that underlie disease processes will permit the development of specific assessment methods to distinguish individuals with the greatest impairment from those with lesser levels permitting the stratification of patients for diagnosis, prognosis and treatment. The first requirement is to select a suitable gene set to initiate studies to examine this possibility for the diseases of hemostasis; we have achieved this in our work to prepare this review. Producing the final list of definitive members of the Hemostaseome will require a process of triangulation; first an inclusive selection must be made from the entire genome which can then be studied and refined, with additions and removals, until a consensus is achieved and the set of contributing members is fully demonstrated. The test set we propose in this paper can now be employed as the foundation for analyses that test multiple individual genome sequences obtained from cases and controls to determine if variations are assessed comprehensively, and whether complex profiles can provide an accurate diagnosis. We have shown in this report that such a comprehensive analysis will require careful assessment of conventional mutations, functional polymorphisms and copy number variations. We must be able to rapidly identify both novel and previously characterized variations in the genomes we plan to examine. Our understanding of how biological pathways are compromised or impaired in diseases with Mendelian inheritance now establishes the framework within which we can view the level of impairment occurring in pathways of diseases with complex inheritance. We expect that individuals who carry a subset of hemostatic variants but who do not demonstrate disease will be below some notional threshold of pathway impairment that their symptomatic relatives exceed. We anticipate that the new NGS sequencing and array hybridization platforms combined with novel automated analysis tools will result in a

fundamental shift in how we understand diseases with complex inheritance and that the Hemostaseome represents an excellent test system for exploration of this concept. Should this new view of the basis of complex inheritance turn out to be correct, novel diagnostic methods will emanate from this study that could potentially benefit a very large number of patients who suffer from diseases of hemostasis and whose care would consequently be personalized.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by NIH grant numbers 1R43HG004678-01 and 5R43HG004678-02 awarded to K.F. and 3rd Millennium, Inc. We gratefully acknowledge the use of the Database of Genomic Variants (http://projects.tcag.ca/variation/?source=hg18) for the initial identification of published CNV reports spanning the genes of the Hemostaseome. Additional citations present in Online Resources 1 and 3 but not present in the text have their full citations annotated in Online Resource 4. To access links in HGMD®, register at the Web site (http://www.hgmd.org) and then log in to successfully engage embedded links within the article.

References

ACOG Committee Opinion. ACOG Committee Opinion no. 451: Von Willebrand disease in women. Obstet Gynecol. 2009; 114:1439–1443. doi:10.1097/AOG.0b013e3181c6f975. [PubMed: 20134302]

Arellano AR, Bezemer ID, Tong CH, Catanese JJ, Devlin JJ, Reitsma PH, Bare LA, Rosendaal FR. Gene variants associated with venous thrombosis: confirmation in the MEGA study. J Thromb Haemost. 2010; 8:1132–1134. doi:10.1111/j.1538-7836.2010.03782.x. [PubMed: 20128871]

Asatiani E, Kessler CM. Multiple congenital coagulopathies co-expressed with von Willebrand's disease: the experience of Hemophilia Region III Treatment Centers over 25 years and review of the literature. Haemophilia. 2007; 13:685–696. doi:10.1111/ j.1365-2516.2007.01541.x. [PubMed: 17973844]

Ashley EA, Butte AJ, Wheeler MT, Chen R, Klein TE, Dewey FE, Dudley JT, Ormond KE, Pavlovic A, Morgan AA, Pushkarev D, Neff NF, Hudgins L, Gong L, Hodges LM, Berlin DS, Thorn CF, Sangkuhl K, Hebert JM, Woon M, Sagreiya H, Whaley R, Knowles JW, Chou MF, Thakuria JV, Rosenbaum AM, Zaranek AW, Church GM, Greely HT, Quake SR, Altman RB. Clinical assessment incorporating a personal genome. Lancet. 2010; 375:1525–1535. doi:10.1016/S0140-6736(10)60452-7. [PubMed: 20435227]

Astermark J, Tengborn L, Hedner U, Berntorp E. Anti- and procoagulant activities in factor VII-deficient subjects. Thromb Res. 2001; 101:435–440. [PubMed: 11323000]

Astermark J, Lacroix-Desmazes S, Reding MT. Inhibitor development. Haemophilia. 2008; 14(Suppl 3):36–42. doi:10.1111/j.1365-2516.2008.01711.x. [PubMed: 18510520]

Austin H, De Staercke C, Lally C, Bezemer ID, Rosendaal FR, Craig Hooper W. New gene variants associated with venous thrombosis—a replication study in United States whites and blacks. J Thromb Haemost. 2011; 9:489–495. doi: 10.1111/j.1538-7836.2011.04185.x. [PubMed: 21232005]

Beckmann JS, Sharp AJ, Antonarakis SE. CNVs and genetic medicine (excitement and consequences of a rediscovery). Cytogenet Genome Res. 2008; 123:7–16. doi:10.1159/000184687. [PubMed: 19287134]

Berg LP, Varon D, Martinowitz U, Wieland K, Kakkar VV, Cooper DN. Combined factor VIII/factor XI deficiency may cause intra-familial clinical variability in haemophilia A among Ashkenazi Jews. Blood Coagul Fibrinolysis. 1994; 5:59–62. [PubMed: 8180339]

Bernardi F, Legnani C, Micheletti F, Lunghi B, Ferraresi P, Palareti G, Biagi R, Marchetti G. A heparin cofactor II mutation (HCII Rimini) combined with factor V Leiden or type I protein C

deficiency in two unrelated thrombophilic subjects. Thromb Haemost. 1996; 76:505–509. [PubMed: 8902986]

Bertina RM, Koeleman BP, Koster T, Rosendaal FR, Dirven RJ, de Ronde H, van der Velden PA, Reitsma PH. Mutation in blood coagulation factor V associated with resistance to activated protein C. Nature. 1994; 369:64–67. doi:10.1038/369064a0. [PubMed: 8164741]

Bongers TN, de Bruijne EL, Dippel DW, de Jong AJ, Deckers JW, Pol-dermans D, de Maat MP, Leebeek FW. Lower levels of ADAMTS13 are associated with cardiovascular disease in young patients. Atherosclerosis. 2009; 207:250–254. doi:10.1016/j.atherosclerosis.2009.04.013. [PubMed: 19439298]

Bowne SJ, Sullivan LS, Koboldt DC, Ding L, Fulton R, Abbott RM, Sodergren EJ, Birch DG, Wheaton DH, Heckenlively JR, Liu Q, Pierce EA, Weinstock GM, Daiger SP. Identification of disease-causing mutations in autosomal dominant retinitis pigmentosa (adRP) using next-generation DNA sequencing. Invest Ophthalmol Vis Sci. 2011; 52:494–503. doi:10.1167/iovs.10-6180. [PubMed: 20861475]

Castaman G, Faioni EM, Tosetto A, Bernardi F. The factor V HR2 haplotype and the risk of venous thrombosis: a meta-analysis. Haematologica. 2003; 88:1182–1189. [PubMed: 14555316]

Christiansen SC, Lijfering WM, Helmerhorst FM, Rosendaal FR, Cannegieter SC. Sex difference in risk of recurrent venous thrombosis and the risk profile for a second event. J Thromb Haemost. 2010; 8:2159–2168. doi:10.1111/j.1538-7836.2010.03994.x. [PubMed: 20738758]

Clarke AJ, Cooper DN. GWAS: heritability missing in action? Eur J Hum Genet. 2010; 18:859–861. doi:10.1038/ejhg.2010.35. [PubMed: 20234388]

Clarke R, Peden JF, Hopewell JC, Kyriakou T, Goel A, Heath SC, Parish S, Barlera S, Franzosi MG, Rust S, Bennett D, Silveira A, Malarstig A, Green FR, Lathrop M, Gigante B, Leander K, de Faire U, Seedorf U, Hamsten A, Collins R, Watkins H, Farrall M. Genetic variants associated with Lp(a) lipoprotein level and coronary disease. N Engl J Med. 2009; 361:2518–2528. doi:10.1056/NEJMoa0902604. [PubMed: 20032323]

Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, Aerts J, Andrews TD, Barnes C, Campbell P, Fitzgerald T, Hu M, Ihm CH, Kristiansson K, Macarthur DG, Macdonald JR, Onyiah I, Pang AW, Robson S, Stirrups K, Valsesia A, Walter K, Wei J, Tyler-Smith C, Carter NP, Lee C, Scherer SW, Hurles ME. Origins and functional impact of copy number variation in the human genome. Nature. 2010; 464:704–712. doi:10.1038/nature08516. [PubMed: 19812545]

Cooper DN, Chen JM, Ball EV, Howells K, Mort M, Phillips AD, Chuzhanova N, Krawczak M, Kehrer-Sawatzki H, Stenson PD. Genes, mutations, and human inherited disease at the dawn of the age of personalized genomics. Hum Mutat. 2010; 31:631–655. doi:10.1002/humu.21260. [PubMed: 20506564]

Corral J, Zuazu-Jausoro I, Rivera J, Gonzalez-Conejero R, Ferrer F, Vicente V. Clinical and analytical relevance of the combination of prothrombin 20210A/A and factor V Leiden: results from a large family. Br J Haematol. 1999; 105:560–563. [PubMed: 10233438]

Daly AK. Pharmacogenomics of anticoagulants: steps toward personal dosage. Genome Med. 2009; 1:10. doi:10.1186/gm10. [PubMed: 19348697]

de Smith AJ, Walters RG, Froguel P, Blakemore AI. Human genes involved in copy number variation: mechanisms of origin, functional effects and implications for disease. Cytogenet Genome Res. 2008; 123:17–26. doi:10.1159/000184688. [PubMed: 19287135]

De Stefano V, Zappacosta B, Persichilli S, Rossi E, Casorelli I, Paciaroni K, Chiusolo P, Leone AM, Giardina B, Leone G. Prevalence of mild hyperhomocysteinaemia and association with thrombophilic genotypes (factor V Leiden and prothrombin G20210A) in Italian patients with venous thromboembolic disease. Br J Haematol. 1999; 106:564–568. [PubMed: 10460623]

Douketis J, Tosetto A, Marcucci M, Baglin T, Cushman M, Eichinger S, Palareti G, Poli D, Tait RC, Iorio A. Patient-level meta-analysis: effect of measurement timing, threshold, and patient age on ability of D-dimer testing to assess recurrence risk after unprovoked venous thromboembolism. Ann Intern Med. 2010; 153:523–531. doi:10.1059/0003-4819-153-8-201010190-00009. [PubMed: 20956709]

Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, Nadeau JH. Missing heritability and strategies for finding the underlying causes of complex disease. Nat Rev Genet. 2010; 11:446–450. doi:10.1038/nrg2809. [PubMed: 20479774]

Epstein RS, Moyer TP, O'Kane DJ, Aubert RE, Xia F, Verbrugge RR, Gage BF, Teagarden JR. Warfarin genotyping reduces hospitalization rates results from the MM-WES (Medco-Mayo Warfarin Effectiveness study). J Am Coll Cardiol. 2010; 55:2804–2812. doi:10.1016/j.jacc.2010.03.009. [PubMed: 20381283]

Formstone CJ, Hallam PJ, Tuddenham EG, Voke J, Layton M, Nicolaides K, Hann IM, Cooper DN. Severe perinatal thrombosis in double and triple heterozygous offspring of a family segregating two independent protein S mutations and a protein C mutation. Blood. 1996; 87:3731–3737. [PubMed: 8611698]

Gandrille S, Greengard JS, Alhenc-Gelas M, Juhan-Vague I, Abgrall JF, Jude B, Griffin JH, Aiach M. Incidence of activated protein C resistance caused by the ARG 506 GLN mutation in factor V in 113 unrelated symptomatic protein C-deficient patients. The French Network on the behalf of INSERM. Blood. 1995; 86:219–224. [PubMed: 7795227]

Gemmati D, Serino ML, Moratelli S, Mari R, Ballerini G, Scapoli GL. Coexistence of antithrombin deficiency, factor V Leiden and hyperhomocysteinemia in a thrombotic family. Blood Coagul Fibrinolysis. 1998; 9:173–176. [PubMed: 9622215]

Ghosh K, Shetty S, Mohanty D. Milder clinical presentation of haemophilia A with severe deficiency of factor VIII as measured by one-stage assay. Haemophilia. 2001; 7:9–12. [PubMed: 11136374]

Girolami A, Luzzatto G, Varvarikis C, Pellati D, Sartori R, Girolami B. Main clinical manifestations of a bleeding diathesis: an often disregarded aspect of medical and surgical history taking. Haemophilia. 2005; 11:193–202. doi: 10.1111/j.1365-2516.2005.01 100.x. [PubMed: 15876263]

Girolami A, Tezza F, Scandellari R, Vettore S, Girolami B. Associated prothrombotic conditions are probably responsible for the occurrence of thrombosis in almost all patients with congenital FVII deficiency. Critical review of the literature. J Thromb Thrombolysis. 2010; 30:172–178. doi: 10.1007/s11239-009-0435-y. [PubMed: 20044773]

Gohil R, Peck G, Sharma P. The genetics of venous thromboembolism. A meta-analysis involving approximately 120,000 cases and 180,000 controls. Thromb Haemost. 2009; 102:360–370. doi: 10.1160/TH09-01-0013. [PubMed: 19652888]

Gonzalez Ordonez AJ, Sanchez LM, Medina Rodriguez JM, Martinez Muniz MA, Menendez Caro JL, Alvarez MV. Double carriers of the factor V Leiden and prothrombin (FIIG20210A) mutations: a description of four cases. Haematologica. 1999; 84:1054–1056. [PubMed: 10553174]

Gonzalez CM, Diaz-Golpe V, Martin S, Sanchez Del Real J, Redondo MC. Combination of congenital coagulation disorders: factor II gene mutation G20210A, factor V Leiden gene mutation G1691A and protein S deficiency. A family study. Haematologica. 2003; 88:ECR20. [PubMed: 12801853]

Heinig M, Petretto E, Wallace C, Bottolo L, Rotival M, Lu H, Li Y, Sarwar R, Langley SR, Bauerfeind A, Hummel O, Lee YA, Paskas S, Rintisch C, Saar K, Cooper J, Buchan R, Gray EE, Cyster JG, Erdmann J, Hengstenberg C, Maouche S, Ouwehand WH, Rice CM, Samani NJ, Schunkert H, Goodall AH, Schulz H, Roider HG, Vingron M, Blankenberg S, Munzel T, Zeller T, Szymczak S, Ziegler A, Tiret L, Smyth DJ, Pravenec M, Aitman TJ, Cambien F, Clayton D, Todd JA, Hubner N, Cook SA. A trans-acting locus regulates an anti-viral expression network and type 1 diabetes risk. Nature. 2010; 467:460–464. doi:10.1038/nature 09386. [PubMed: 20827270]

Iorio A, Kearon C, Filippucci E, Marcucci M, Macura A, Pengo V, Siragusa S, Palareti G. Risk of recurrence after a first episode of symptomatic venous thromboembolism provoked by a transient risk factor: a systematic review. Arch Intern Med. 2010; 170:1710–1716. doi:10.1001/archinternmed.2010.367. [PubMed: 20975016]

Jayandharan GR, Nair SC, Poonnoose PM, Thomas R, John J, Keshav SK, Cherian RS, Devadarishini M, Lakshmi KM, Shaji RV, Viswabandya A, George B, Mathews V, Chandy M, Srivastava A. Polymorphism in factor VII gene modifies phenotype of severe haemophilia. Haemophilia. 2009; 15:1228–1236. doi:10.1111/ j.1365-2516.2009.02080.x. [PubMed: 19686262]

Kamstrup PR. Lipoprotein(a) and ischemic heart disease-A causal association? A review. Atherosclerosis. 2010 doi:10.1016/j.atherosclerosis.2009.12.036.

Katz R, Shani D. A case of deep vein thrombosis in patient with Hermansky-Pudlak syndrome. Ann Hematol. 2005; 84:763. doi:10.1007/s00277-005-1081-x. [PubMed: 16044314]

Koeleman BP, Reitsma PH, Allaart CF, Bertina RM. Activated protein C resistance as an additional risk factor for thrombosis in protein C-deficient families. Blood. 1994; 84:1031–1035. [PubMed: 8049422]

Koeleman BP, van Rumpt D, Hamulyak K, Reitsma PH, Bertina RM. Factor V Leiden: an additional risk factor for thrombosis in protein S deficient families? Thromb Haemost. 1995; 74:580–583. [PubMed: 8584987]

Kohane IS, Masys DR, Altman RB. The incidentalome: a threat to genomic medicine. JAMA. 2006; 296:212–215. doi:10.1001/jama.296.2.212. [PubMed: 16835427]

Koren A, Levin C, Hujirat Y, El-Hasid R, Kutai M, Lanir N, Shalev S, Brenner B. Thrombophilia in infancy: factor V Leiden and MTHFR or factor II double heterozygocity as a risk factor. Pediatr Hematol Oncol. 2003; 20:219–227. [PubMed: 12637218]

Kraft HG, Lingenhel A, Kochl S, Hoppichler F, Kronenberg F, Abe A, Muhlberger V, Schonitzer D, Utermann G. Apolipoprotein(a) kringle IV repeat number predicts risk for coronary heart disease. Arterioscler Thromb Vasc Biol. 1996; 16:713–719. [PubMed: 8640397]

Lalonde E, Albrecht S, Ha KC, Jacob K, Bolduc N, Polychronakos C, Dechelotte P, Majewski J, Jabado N. Unexpected allelic heterogeneity and spectrum of mutations in Fowler syndrome revealed by next-generation exome sequencing. Hum Mutat. 2010; 31:918–923. doi:10.1002/humu.21293. [PubMed: 20518025]

Lane DA, Auberger K, Ireland H, Roscher AA, Thein SL. Prenatal diagnosis in combined antithrombin and factor V gene mutation. Br J Haematol. 1996; 94:753–755. [PubMed: 8826905]

Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, Willer CJ, Jackson AU, Vedantam S, Raychaudhuri S, Ferreira T, Wood AR, Weyant RJ, Segre AV, Speliotes EK, Wheeler E, Soranzo N, Park JH, Yang J, Gudbjartsson D, Heard-Costa NL, Randall JC, Qi L, Vernon Smith A, Magi R, Pastinen T, Liang L, Heid IM, Luan J, Thorleifsson G, Winkler TW, Goddard ME, Sin Lo K, Palmer C, Workalemahu T, Aulchenko YS, Johansson A, Carola Zillikens M, Feitosa MF, Esko T, Johnson T, Ketkar S, Kraft P, Mangino M, Prokopenko I, Absher D, Albrecht E, Ernst F, Glazer NL, Hayward C, Hottenga JJ, Jacobs KB, Knowles JW, Kutalik Z, Monda KL, Polasek O, Preuss M, Rayner NW, Robertson NR, Steinthorsdottir V, Tyrer JP, Voight BF, Wiklund F, Xu J, Hua Zhao J, Nyholt DR, Pellikka N, Perola M, Perry JR, Surakka I, Tammesoo ML, Altmaier EL, Amin N, Aspelund T, Bhangale T, Boucher G, Chasman DI, Chen C, Coin L, Cooper MN, Dixon AL, Gibson Q, Grundberg E, Hao K, Juhani Junttila M, Kaplan LM, Kettunen J, Konig IR, Kwan T, Lawrence RW, Levinson DF, Lorentzon M, McKnight B, Morris AP, Muller M, Suh Ngwa J, Purcell S, Rafelt S, Salem RM, Salvi E, et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature. 2010; 467:832–838. doi:10.1038/nature09410. [PubMed: 20881960]

Larciprete G, Rossi F, Deaibess T, Brienza L, Barbati G, Romanini E, Gioia S, Cirese E. Double inherited thrombophilias and adverse pregnancy outcomes: fashion or science? J Obstet Gynaecol Res. 2010; 36:996–1002. doi:10.1111/j.1447-0756.2010.01262.x. [PubMed: 20868443]

Lee C, Scherer SW. The clinical context of copy number variation in the human genome. Expert Rev Mol Med. 2010; 12:e8. doi:10.1017/S1462399410001390. [PubMed: 20211047]

Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Axelrod N, Huang J, Kirkness EF, Denisov G, Lin Y, MacDonald JR, Pang AW, Shago M, Stockwell TB, Tsiamouri A, Bafna V, Bansal V, Kravitz SA, Busam DA, Beeson KY, McIntosh TC, Remington KA, Abril JF, Gill J, Borman J, Rogers YH, Frazier ME, Scherer SW, Strausberg RL, Venter JC. The diploid genome sequence of an individual human. PLoS Biol. 2007; 5:e254. doi:10.1371/ journal.pbio.0050254. [PubMed: 17803354]

Lijfering WM, Middeldorp S, Veeger NJ, Hamulyak K, Prins MH, Buller HR, van der Meer J. Risk of recurrent venous thrombosis in homozygous carriers and double heterozygous carriers of factor V Leiden and prothrombin G20210A. Circulation. 2010; 121:1706–1712. doi:10.1161/CIRCULATIONAHA.109.906347. [PubMed: 20368522]

Lockwood WW, Chari R, Coe BP, Girard L, Macaulay C, Lam S, Gazdar AF, Minna JD, Lam WL. DNA amplification is a ubiquitous mechanism of oncogene activation in lung and other cancers. Oncogene. 2008; 27:4615–4624. doi:10.1038/onc.2008.98. [PubMed: 18391978]

Lotta LA. Genome-wide association studies in atherothrombosis. Eur J Intern Med. 2010; 21:74–78. doi:10.1016/j.ejim.2009.11.003. [PubMed: 20206874]

Mannucci PM, Asselta R, Duga S, Guella I, Spreafico M, Lotta L, Angelica Merlini P, Peyvandi F, Kathiresan S, Ardissino D. The association of factor v Leiden with myocardial infarction is

replicated in 1,880 patients with premature disease. J Thromb Haemost. 2010; 8:2116–2121. doi: 10.1111/j.1538-7836.2010.03982.x. [PubMed: 20626623]

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM. Finding the missing heritability of complex diseases. Nature. 2009; 461:747–753. doi:10.1038/nature08494. [PubMed: 19812666]

Mardis ER. The $1,000 genome, the $100,000 analysis? Genome Med. 2010; 2:84. doi:10.1186/gm205. [PubMed: 21114804]

Mariani G, Herrmann FH, Schulman S, Batorova A, Wulff K, Etro D, Dolce A, Auerswald G, Astermark J, Schved JF, Ingerslev J, Bernardi F. Thrombosis in inherited factor VII deficiency. J Thromb Haemost. 2003; 1:2153–2158. [PubMed: 14521598]

Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, Beck AE, Tabor HK, Cooper GM, Mefford HC, Lee C, Turner EH, Smith JD, Rieder MJ, Yoshiura K, Matsumoto N, Ohta T, Niikawa N, Nickerson DA, Bamshad MJ, Shendure J. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nat Genet. 2010a; 42:790–793. doi:10.1038/ng.646. [PubMed: 20711175]

Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ. Exome sequencing identifies the cause of a Mendelian disorder. Nat Genet. 2010b; 42:30–35. doi:10.1038/ng.499. [PubMed: 19915526]

Nowak-Gottl U, Escuriola C, Kurnik K, Schobess R, Horneff S, Kosch A, Kreuz W, Pollmann H. Haemophilia and thrombophilia. What do we learn about combined inheritance of both genetic variations? Hamostaseologie. 2003; 23:36–40. doi: 10.1267/Hamo03010036. [PubMed: 12567198]

O'Brien LA, James PD, Othman M, Berber E, Cameron C, Notley CR, Hegadorn CA, Sutherland JJ, Hough C, Rivard GE, O'Shaunessey D, Lillicrap D. Founder von Willebrand factor haplo-type associated with type 1 von Willebrand disease. Blood. 2003; 102:549–557. doi:10.1182/blood-2002-12-3693. [PubMed: 12649144]

Parker HG, VonHoldt BM, Quignon P, Margulies EH, Shao S, Mosher DS, Spady TC, Elkahloun A, Cargill M, Jones PG, Maslen CL, Acland GM, Sutter NB, Kuroki K, Bustamante CD, Wayne RK, Ostrander EA. An expressed fgf4 retrogene is associated with breed-defining chondrodysplasia in domestic dogs. Science. 2009; 325:995–998. doi:10.1126/science.1173275. [PubMed: 19608863]

Paterson AD, Rommens JM, Bharaj B, Blavignac J, Wong I, Diamandis M, Waye JS, Rivard GE, Hayward CP. Persons with Quebec platelet disorder have a tandem duplication of PLAU, the urokinase plasminogen activator gene. Blood. 2010; 115:1264–1266. doi:10.1182/blood-2009-07-233965. [PubMed: 20007542]

Pelak K, Shianna KV, Ge D, Maia JM, Zhu M, Smith JP, Cirulli ET, Fellay J, Dickson SP, Gumbs CE, Heinzen EL, Need AC, Ruzzo EK, Singh A, Campbell CR, Hong LK, Lornsen KA, McKenzie AM, Sobreira NL, Hoover-Fong JE, Milner JD, Ottman R, Haynes BF, Goedert JJ, Goldstein DB. The characterization of twenty sequenced human genomes. PLoS Genet 6:e1001111. 2010 doi: 10.1371/journal.pgen.1001111.

Perry GH, Ben-Dor A, Tsalenko A, Sampas N, Rodriguez-Revenga L, Tran CW, Scheffer A, Steinfeld I, Tsang P, Yamada NA, Park HS, Kim JI, Seo JS, Yakhini Z, Laderman S, Bruhn L, Lee C. The fine-scale and complex architecture of human copy-number variation. Am J Hum Genet. 2008; 82:685–695. doi:10.1016/ j.ajhg.2007.12.010. [PubMed: 18304495]

Pintao MC, Garcia AA, Borgel D, Alhenc-Gelas M, Spek CA, de Visser MC, Gandrille S, Reitsma PH. Gross deletions/duplications in PROS1 are relatively common in point mutation-negative hereditary protein S deficiency. Hum Genet. 2009; 126:449–456. doi:10.1007/s00439-009-0687-9. [PubMed: 19466456]

Poort SR, Rosendaal FR, Reitsma PH, Bertina RM. A common genetic variation in the 3′-untranslated region of the prothrombin gene is associated with elevated plasma prothrombin levels and an increase in venous thrombosis. Blood. 1996; 88:3698–3703. [PubMed: 8916933]

Rogaev EI, Grigorenko AP, Faskhutdinova G, Kittler EL, Moliaka YK. Genotype analysis identifies the cause of the “royal disease”. Science. 2009; 326:817. doi:10.1126/science.1180660. [PubMed: 19815722]

Salomon O, Steinberg DM, Zucker M, Varon D, Zivelin A, Seligsohn U. Patients with severe factor XI deficiency have a reduced incidence of deep-vein thrombosis. Thromb Haemost. 2011; 105:269–273. doi:10.1160/TH10-05-0307. [PubMed: 21057700]

Samama MM, Trossaert M, Horellou MH, Elalamy I, Conard J, Deschamps A. Risk of thrombosis in patients homozygous for factor V Leiden. Blood. 1995; 86:4700–4702. [PubMed: 8541565]

Sansores-Garcia L, Majluf-Cruz A. Arterial and venous thrombosis associated to combined deficiency of protein C and antithrombin III. Am J Hematol. 1998; 57:182–183. doi:10.1002/(SICI)1096-8652(199802)57:2<182:AID-AJH20>3.0.CO;2-D. [PubMed: 9462558]

Schadt EE. Molecular networks as sensors and drivers of common human diseases. Nature. 2009; 461:218–223. doi: 10.1038/nature08454. [PubMed: 19741703]

Schork NJ, Murray SS, Frazer KA, Topol EJ. Common vs. rare allele hypotheses for complex diseases. Curr Opin Genet Dev. 2009; 19:212–219. doi:10.1016/j.gde.2009.04.010. [PubMed: 19481926]

Schuster-Bockler B, Conrad D, Bateman A. Dosage sensitivity shapes the evolution of copy-number varied regions. PLoS One. 2010; 5:e9474. doi:10.1371/journal.pone.0009474. [PubMed: 20224824]

Smith NL, Rice KM, Lumley T, Heckbert SR, Psaty BM. Discovering novel risk factors for venous thrombosis: a candidate-gene approach. Thromb Res. 2009a; 123(Suppl 4):S25–S29. doi:10.1016/S0049-3848(09)70139-2. [PubMed: 19303500]

Smith NL, Wiggins KL, Reiner AP, Lange LA, Cushman M, Heckbert SR, Lumley T, Rice KM, Folsom AR, Psaty BM. Replication of findings on the association of genetic variation in 24 hemostasis genes and risk of incident venous thrombosis. J Thromb Haemost. 2009b; 7:1743–1746. doi: 10.1111/j.1538-7836.2009.03567.x. [PubMed: 19682239]

SoW F, Cesari F, Abbate R, Gensini GF, Broze GJ Jr, Fedi S. A meta-analysis of potential risks of low levels of protein Z for diseases related to vascular thrombosis. Thromb Haemost. 2010; 103:749–756. doi:10.1160/TH09-09-0645. [PubMed: 20076855]

Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, Redon R, Bird CP, de Grassi A, Lee C, Tyler-Smith C, Carter N, Scherer SW, Tavare S, Deloukas P, Hurles ME, Dermitzakis ET. Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science. 2007; 315:848–853. doi:10.1126/science.1136678. [PubMed: 17289997]

Tirado I, Mateo J, Soria JM, Oliver A, Borrell M, Coll I, Vallve C, Souto JC, Martinez-Sanchez E, Fontcuberta J. Contribution of prothrombin 20210A allele and factor V Leiden mutation to thrombosis risk in thrombophilic families with other hemostatic deficiencies. Haematologica. 2001; 86:1200–1208. [PubMed: 11694407]

U.S. Preventive Services Task Force. Aspirin for the prevention of cardiovascular disease: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med. 2009b; 150:396–404. [PubMed: 19293072]

van der Hoek YY, Wittekoek ME, Beisiegel U, Kastelein JJ, Koschinsky ML. The apolipoprotein(a) kringle IV repeats which differ from the major repeat kringle are present in variably-sized isoforms. Hum Mol Genet. 1993; 2:361–366. [PubMed: 8389224]

Wells PS, Anderson JL, Scarvelis DK, Doucette SP, Gagnon F. Factor XIII Val34Leu variant is protective against venous thromboembolism: a HuGE review and meta-analysis. Am J Epidemiol. 2006; 164:101–109. doi:10.1093/aje/kwj179. [PubMed: 16740590]

Wolff T, Miller T, Ko S. Aspirin for the primary prevention of cardiovascular events: an update of the evidence for the U.S. Preventive Services Task Force. Ann Intern Med. 2009; 150:405–410. doi: 150/6/405 [pii]. [PubMed: 19293073]

Wong KK, deLeeuw RJ, Dosanjh NS, Kimm LR, Cheng Z, Horsman DE, MacAulay C, Ng RT, Brown CJ, Eichler EE, Lam WL. A comprehensive analysis of common copy-number variations in the human genome. Am J Hum Genet. 2007; 80:91–104. doi:10.1086/510560. [PubMed: 17160897]

Yang Q, Kathiresan S, Lin JP, Tofler GH, O'Donnell CJ. Genome-wide association and linkage analyses of hemostatic factors and hematological phenotypes in the Framingham Heart Study. BMC Med Genet. 2007; 8(Suppl 1):S12. doi:10.1186/1471-2350-8-S1-S12. [PubMed: 17903294]

Ye Z, Liu EH, Higgins JP, Keavney BD, Lowe GD, Collins R, Danesh J. Seven haemostatic gene polymorphisms in coronary disease: meta-analysis of 66,155 cases and 91,307 controls. Lancet. 2006; 367:651–658. doi:10.1016/S0140-6736(06)68263-9. [PubMed: 16503463]

Young G, Albisetti M, Bonduel M, Brandao L, Chan A, Friedrichs F, Goldenberg NA, Grabowski E, Heller C, Journeycake J, Kenet G, Krumpel A, Kurnik K, Lubetsky A, Male C, Manco-Johnson M, Mathew P, Monagle P, van Ommen H, Simioni P, Svirin P, Tormene D, Nowak-Gottl U. Impact of inherited thrombophilia on venous thromboembolism in children: a systematic review and meta-analysis of observational studies. Circulation. 2008; 118:1373–1382. doi:10.1161/CIRCULATIONAHA.108.789008. [PubMed: 18779442]

Zhang AH, Skupsky J, Scott DW. Factor VIII inhibitors: risk factors and methods for prevention and immune modulation. Clin Rev Allergy Immunol. 2009; 37:114–124. doi:10.1007/s12016-009-8122-5. [PubMed: 19199081]

Table 1

Phenotypes, copy number variants and numbers of characterized mutations for the genes encoding members of the coagulation pathway

Gene symbol	MIM #	Protein level		CNV		HGMD® Public					
		Incr.	Defic.	Gain	Loss	Misnon	Reg	Indel	Del	Splicing	Total
F10	227600	Estro	Yes		CL	77	0	8	3	5	93
F11	264900	Yes	Yes			102	1	11	1	12	127
F13A1	134570	Yes	Yes			48	0	23	1	10	83
F13B	134580	Yes	Yes		CL	2	0	3	0	0	5
F2	176930	Yes	Yes			41	4	5	1	3	54
F3	134390	Yes	Mice			0	1	1	0	0	2
F5	612309	Yes	Yes		CL	38	0	21	1	6	67
F7	227500	Yes	Yes	CG	CL	115	11	19	2	22	172
F8	306700	Yes	Yes	CG	PL-1A	896	2	307	138	90	1,451
F9	300746	Yes	Yes	PG-2		634	24	178	59	93	1,001
FGA	134820	Yes	Yes			37	3	25	4	5	75
FGB	134830	Yes	Yes			33	2	5	1	5	44
FGG	134850	Yes	Yes			55	1	4	1	5	66
VWF	193400	Yes	Yes			230	5	39	14	23	313
Total						2,308	54	649	226	279	3,553

Gene symbol listed according to HUGO Gene Nomenclature Committee (http://www.genenames.org/index.html)

MIM# accession number in the On-Line Mendelian Inheritance in Man (OMIM; http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim)

Protein level incr indicates a published report associating increased protein levels with hemostatic disease (see On-Line Resource 3, Supplementary Table 2 for details and citations); *estro* indicates elevated protein levels ascertained in women taking oral contraceptives; *defic* indicates that a human deficiency state has been identified, as described in the corresponding OMIM report; *mice* indicates phenotype described in murine transgenic or knockout models, not in human; *CNV* copy number variation gain or loss (*CG* complete gain, *CL* complete loss, *PG* partial gain, *PL* partial loss); *PG-2* indicates that two exons are involved; *PL-1A* indicates that one exon is involved and that this change gives rise to alternatively spliced transcripts (A); *HGMD® Public* indicates number of mutation entries listed in the public version of the Human Gene Mutation Database (http://www.hgmd.org): *misnon* missense and nonsense mutations combined; *reg* regulatory mutations; *indel* insertions, deletions and indels combined; *del* gross deletions; *splicing* mutations that impact splice sites; *total* total mutation number for each gene collated in HGMD® as of August 2010

Table 2

Phenotypes, copy number variants and numbers of characterized mutations for the genes encoding members of the fibrinolysis and kallikrein-kinin pathways (the contact system)

Gene symbol	MIM #	Protein level		CNV		HGMD® Public					
		Incr.	Defic.	Gain	Loss	Misnon	Reg	Indel	Del	Splicing	Total
F12	610619	Yes	Yes		CL	13	2	2	0	2	19
KLKB1	229000	No	Yes			6	0	0	0	0	6
KNG1	612358	Estro	Yes		CL	2	0	0	0	0	2
PLAT	173370	Yes	Yes		PL-6A	0	1	0	0	0	2
PLAU	191840	Yes	No	TD		1	0	0	0	0	1
PLG	173350	No	Yes	CG		19	0	5	0	0	24
Total						41	3	7	0	2	54

Details as given in Table 1

TD tandem duplication

Table 3

Phenotypes, copy number variants and numbers of characterized mutations for the genes encoding the proteins that regulate coagulation

Gene symbol	MIM #	Protein level		CNV		HGMD® Public					
		Incr.	Defic.	Gain	Loss	Misnon	Reg	Indel	Del	Splicing	Total
ADAMTS13	604134	No	Yes			49	0	9	2	6	66
CPB2	603101	Yes	Yes		PL-10A	2	0	0	0	0	2
HABP2	603924	Estro	No			1	0	0	0	0	1
HRG	142640	Yes	Yes		CL	1	0	0	0	0	1
PROC	612283	Yes	Yes			193	11	33	2	23	262
PROCR	600646	Yes	Mice			1	2	0	0	0	5
PROS1	176880	No	Yes	PD-8	CL	131	3	37	5	20	197
PROZ	176895	Yes	Yes	PG-1		3	1	0	0	1	5
SERPINA1	107400	No	Yes	CG		27	3	6	2	2	41
SERPINA10	605271	Yes	Yes			5	0	0	0	0	5
SERPINA5	601841	No	No			0	0	0	0	0	0
SERPINC1	107300	Yes	Yes		CL	123	0	68	12	12	218
SERPIND1	142360	No	Yes			3	0	2	0	0	5
SERPINE1	173360	Yes	Yes	CG		1	1	2	0	0	4
SERPINF2	613168	No	Yes	PG-8A	CL	2	0	4	0	1	7
SERPING1	606860	Estro	Yes	CG		83	2	59	29	19	196
SPINT2	605124	No	Yes			0	0	0	0	0	0
TFPI	152310	Yes	No	PG-7A	PL-7A	2	0	0	0	1	3
TFPI2	600033	No	No			0	0	0	0	0	0
THBD	188040	Yes	No		PL-1	7	3	3	0	0	13
THBS1	188060	No	Mice			1	0	0	0	0	1
THBS2	188061	No	Mice	PG-1	PL-1	0	0	0	0	0	0
THBS3	188062	No	No		CL	0	0	0	0	0	0
THBS4	600715	No	No			1	0	0	0	0	1
VTN	193190	Yes	Mice			0	0	0	0	0	0
Total						636	26	223	52	85	1,033

Details as given in Table 1

PD-8 partial duplication aVecting eight exons

Table 4

Phenotypes, copy number variants and numbers of characterized mutations for the genes encoding platelet cytoplasmic proteins

Gene symbol	MIM #	Protein level		CNV		HGMD® Public					
		Incr.	Defic.	Gain	Loss	Misnon	Reg	Indel	Del	Splicing	Total
AP3B1	603401	No	Yes			3	0	0	2	0	5
BLOC1S3	609762	No	Yes	CG	CL	0	0	1	0	0	1
DTNBP1	607145	No	Yes			2	0	0	0	0	2
GNA13	604406	No	No			0	0	0	0	0	0
GNAS	139320	No	Yes			48	0	38	7	11	107
HPS1	604982	No	Yes		PL-14A	6	0	14	0	2	23
HPS3	606118	No	Yes	PG-1		1	0	0	1	5	7
HPS4	606682	No	Yes			6	0	2	0	0	9
HPS5	607521	No	Yes			3	0	4	0	0	7
HPS6	607522	No	Yes			0	0	2	0	0	2
PLA2G4A	607522	Yes	Yes			0	0	0	0	0	0
PTGS1	176805	No	No			6	1	0	0	0	7
RAP1B	179530	No	Mice			0	0	0	0	0	0
TPH1	191060	No	Mice			0	0	0	0	1	1
Total						75	1	61	10	19	171

Details as given in Table 1

Table 5

Phenotypes, copy number variants and numbers of characterized mutations for the genes encoding platelet surface proteins

Gene symbol	MIM #	Protein level		CNV		HGMD® Public					
		Incr.	Defic.	Gain	Loss	Misnon	Reg	Indel	Del	Splicing	Total
AXL	109135	No	Mice			1	0	0	0	0	1
CD36	173510	No	Yes			7	0	11	1	0	20
CD40	109535	No	Yes		CL	1	1	0	0	2	4
EPR1	603411	No	No			0	0	0	0	0	0
F2R	187930	No	Mice			0	0	1	0	2	3
FERMT3	607901	No	Yes			0	0	0	0	0	0
GP1BA	606672	No	Yes		CL	14	1	7	2	0	24
GP1BB	138720	No	Yes	CG		9	1	4	0	0	14
GP5	173511	Yes	Yes			0	0	0	0	0	0
GP6	605546	Yes	No	PG-3		1	0	0	0	0	1
GP9	173515	No	Yes			10	0	0	0	0	10
ITGA2B	607759	No	Yes			55	0	25	1	17	100
ITGB3	173470	No	Yes			45	0	16	1	6	71
P2RX1	600845	No	Yes		CL	0	0	1	0	0	1
P2RY1	601167	No	Yes			0	0	0	0	0	0
P2RY12	600515	No	Yes			4	0	1	0	0	5
PTAFR	173393	No	No			0	0	0	0	0	0
PTGIR	600022	No	Mice			0	0	0	0	0	0
SELP	173610	No	No		CL	2	0	0	0	0	2
TBXA2R	188070	No	Yes		CL	3	0	0	0	0	3
Total						152	3	66	5	27	259

Details as given in Table 1

Table 6

Phenotypes, copy number variants and numbers of characterized mutations for the genes encoding proteins that aVect platelet stability underlying thrombocytopenia

Gene symbol	MIM #	Protein level		CNV		HGMD® Public					
		Incr.	Defic.	Gain	Loss	Misnon	Reg	Indel	Del	Splicing	Total
CD59	107271	Yes	Yes			0	0	2	0	0	2
CYCS	123970	No	No			0	0	0	0	0	0
MASTL	608221	No	No		CL	1	0	0	0	0	1
WAS	301000	No	Yes			118	0	130	11	39	300
WASF2	605975	No	Mice		PL-1	0	0	0	0	0	0
Total						119	0	132	11	39	303

Details as given in Table 1

Table 7

Phenotypes, copy number variants and numbers of characterized mutations for the genes encoding metabolic proteins with coagulation phenotypes

Gene symbol	MIM #	Protein level		CNV		HGMD® Public					
		Incr.	Defic.	Gain	Loss	Misnon	Reg	Indel	Del	Splicing	Total
CBS	613381	No	Mice	D		105	0	15	14	12	147
CYP2A6	122720	No	Yes	CG	CL	7	5	0	2	0	15
CYP2C9	601130	No	No			10	0	2	0	1	13
ENTPD1	601752	Mice	Mice			0	0	0	0	0	0
GGCX	137167	No	Mice	CG		8	0	0	1	1	11
MTHFR	607093	No	Yes			30	0	2	0	2	34
NNMT	600008	No	No			0	0	0	0	0	0
PIGM	610273	No	Yes			0	1	0	0	0	1
PLA2G7	601690	No	Yes	PG-1	PL-1	5	0	1	0	0	6
SULT1E1	600043	No	Mice			1	0	0	0	0	1
VKORC1	608547	No	Yes			7	1	0	0	3	11
Total						173	7	20	17	19	239

Details as given in Table 1

D duplication

Table 8

Phenotypes, copy number variants and numbers of characterized mutations for the genes encoding miscellaneous proteins with coagulation phenotypes

Gene symbol	MIM #	Protein level		CNV		HGMD® Public					
		Incr.	Defic.	Gain	Loss	Misnon	Reg	Indel	Del	Splicing	Total
ADIPOQ	605441	Yes	Mice		CL	5	6	0	0	2	14
CTSG	116830	No	No			1	0	0	0	0	1
EPB41L2	603237	No	No			0	0	0	0	0	0
ESR1	133430	No	No	PG-5A		2	0	0	0	1	5
F2RL1	600933	No	Mice			0	0	0	0	0	0
F2RL2	601919	No	Mice			0	0	0	0	0	0
F2RL3	602779	No	Mice			0	0	0	0	0	0
GAS6	600441	Yes	Mice	PG-8A	PL-8A	0	0	0	0	0	0
IL4	147780	No	No			0	2	0	0	0	2
LMAN1	601567	No	Yes			8	0	14	1	6	29
LPA	152200	Yes	Yes	PG-43A	PL-44A	5	3	0	0	1	10
LRP1	107770	No	Mice			2	1	0	0	0	3
MBL2	154545	No	Yes		PL-4	3	3	0	0	0	6
MCFD2	607788	No	Yes			4	0	3	1	2	10
Total						30	15	17	2	12	80

Details as given in Table 1