New approaches and future prospects for evaluating genetic risk of thrombosis

José Manuel Soria
Jordi Fontcuberta

Studies of the genetics of common complex human diseases have progressed enormously over the past decade with the development of powerful molecular and analytical methodologies. For example, the understanding of thrombosis, which is a complex disease under strong genetic and environmental control, has been greatly advanced using the principles and practices of these methodologies. This is important because thrombosis has a high social and economic cost in western countries. Thus, identification and characterization of specific loci and genes, and the associated phenotypes involved in thrombotic risk will contribute to a greater understanding of the pathogenesis of this disease and will, ultimately, lead to the development of better diagnostic, prevention and treatment strategies. Past efforts to identify thrombosis-related genes have utilized population-based association methods, but substantial progress has been made recently with the strategy of using positional cloning of genes based on linkage studies. These methods focus on measuring quantitative traits that are correlated with the risk of disease. In this article, we review the current status and future prospects of mapping and identification of genes for thrombosis, with a focus on some promising chromosomal regions containing these genes. We include a discussion of the exciting prospects of identifying genes that are involved in complex diseases.

Key words: cancer, venous thromboembolism, low-molecular-weight heparin.

Haematologica 2005; 90:1212-1222

Unitat d'Hemostàsia i Trombosi, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain

Correspondence:
José Manuel Soria, Unitat d'Hemostàsia i Trombosi Hospital de la Santa Creu i Sant Pau, Sant Antoni Mª Claret 167, Barcelona 08025, Spain.
E-mail: jsoria@hsp.santpau.es

Venous thromboembolism is a hyper-coagulable state that reflects a complex interplay between inherited and environmental factors. The overall incidence of venous thromboembolism increases with age. Approximately 1:1000 individuals per year are affected in the USA and Western Europe. Thus, it is an important health problem. Although there are few hard data on its prevalence, retrospective and prospective data[1,2] suggest a minimum lifetime prevalence of 5-10% for deep-vein thrombosis. If one includes arterial thromboses, other venous thromboses, and undiagnosed thrombotic conditions, the lifetime prevalence of thrombosis must be substantially higher than 10%.

At present, one or more genetic risk factors can be identified in 50-60% of families which exhibit thrombophilia. Since familial thrombophilia is an oligogenetic disease,[3] it is safe to conclude that unknown genetic risk factors co-segregate in the other families. By studying families of thrombophilic patients, several genetic risk factors for venous thrombosis have been identified. Some of these risk factors are antithrombin, protein C or protein S deficiencies, dysfibrinogenemia and activated protein C (APC)-resistance associated with factor V Leiden.[4] On the other hand, case/control association studies, such as the Leiden Thrombophilia Study (LETS),[5] have been useful for identifying several complex phenotypes that increase the risk of venous thrombosis. In addition, it is likely that pathways and proteins outside the coagulation system are involved in the pathogenesis of venous thrombosis. However, finding the genes that contribute to these functional phenotypes will be a major effort requiring the application of genetic studies with state-of-the-art design and methodology.

Taking into account the progress to date, we agree fully with the statement by Bertina: *"As to the identification of new genetic risk factors for thrombosis, we seem to have arrived at the end of a practicable road with the classical approach of thrombophilia, which usually starts with the study of the association of hemostatic phenotypes and thrombosis risk"*.[5]

In the genomic era, with the completion of mapping of the human DNA sequence, the search for genes involved in complex diseases, such as thrombophilia, is gaining momentum, as is the debate over gene-finding strategies.[6,7] Thus, this review is

not a typical literature overview of post research progress, in that little of this type of research has been completed and published; it is more of a description of currently available tools and future research goals.

Quantitative risk factors for thrombosis

The genetic study of quantitative traits has several advantages over simple classification studies based on the presence or absence of a disease.[8] Many diseases can be defined in terms of an underlying quantitative liability scale, with parameters such as blood pressure for hypertension or glucose tolerance for diabetes. These intermediate risk factors tend to be closer to gene action and thus provide stronger genetic signals than when a discrete clinical endpoint, such as presence or absence of disease, is used. Another advantage of studying intermediate risk factors is that a quantitative trait may be influenced by a small number of factors, whereas a disease may be influenced by many genetic and environmental factors that produce a heterogeneous phenotype. Generally, the quantitative variables will be correlated more strongly with polymorphisms of an underlying gene, and therefore provide a better phenotype both for initial mapping purposes and for identifying the variants at a locus which affect susceptibility.

The physiological cascade that produces thrombin and the pathologic endpoint of thrombosis is complex. There are many components involved in the coagulation and fibrinolytic pathways. The risk of thrombosis is influenced not only by abnormalities in these pathways but also by quantitative physiological variations in these parameters.

Numerous hemostatic factors have been implicated as possible concomitants of both venous and arterial thrombosis. For example, there is epidemiologic evidence for a positive relationship between von Willebrand factor (VWF) and factor (F) VIII levels and the risk of venous[9] and arterial thrombosis.[10] High plasma homocysteine levels have been associated with deep-vein[11] and arterial thrombosis.[12] Quantitative measurements of APC ratio are correlated with the risk of venous thrombosis even when a major genetic influence on APC ratio, the FV Leiden polymorphism, is taken into account.[13] Similarly, levels of FXII[14] and tissue plasminogen activator[15] are correlated with arterial thrombosis. Recently, high plasma levels of FIX[16] and FXI,[17] and low plasma levels of tissue factor pathway inhibitor (TFPI)[18] have been implicated as risk factors for thrombosis.

More recently, bivariate genetic analyses in the GAIT (Genetic Analysis of Idiopathic Thrombophilia) Project demonstrated significant genetic correlations between thrombosis and APC ratio, homocysteine, tissue plasminogen activator, VWF, and clotting FVII, FVIII, FIX, FXI, and FXII. These correlations suggest that there are pleiotropic influences on these quantitative measures and the risk of thrombosis.[19]

Genes that contribute to the patterns of variation in such phenotypes have been called quantitative trait loci (QTL). Intermediate risk factors can be utilized jointly with disease status to search the genome for QTL that pleiotropically affect both intermediate risk factors and disease status.[20-22]

All hemostasis-related phenotypes are under strong genetic control, and despite the relatively small number of family studies, twins and well-characterized pedigrees have established that these phenotypes are under substantial genetic control (Table 1). More importantly, using a liability threshold model,[8] we estimated that 61% of the variability in the liability to develop thrombotic disease is due to the additive effect of genes.[19] These results have been replicated recently by Heit *et al.*,[23] who documented the importance of genetic factors that influence thrombotic disease and its hemostasis-related phenotypes. For most of the traits, genes appear to be the largest determinant of quantitative variation, so it is possible to localize the genetic determinant of these phenotypes with considerable facility.

Statistical genetic methodologies: association versus linkage analysis

Genetic association studies essentially look for correlations between phenotype and genotype. Thus, the aim is to demonstrate that a particular allele or genetic marker, typically a single nucleotide polymorphism (SNP) constitutes a significant risk for a phenotype of interest (e.g., disease status). An association can be found either with functional genetic variants that have biological consequences related to a disease (e.g., Factor V Leiden or G20210A mutation in the prothrombin gene), or most likely, with other variants that are in linkage disequilibrium with these variants.

Linkage disequilibrium occurs when an allele at a locus is situated on the same chromosome with a specific allele at another locus more often than would be expected by chance. Since linkage disequilibrium is the result of the genetic history of a particular mutation, factors influencing linkage disequilibrium at the population level have been discussed amply in the genetic literature.[24] Most association studies operate under a specific hypothesis; i.e., they test for a pre-determined gene or set of genes referred to as candidate genes. However, such studies do not provide evidence that the genetic variant (the SNP) is functional, and in the absence of such data, the relevance of the association remains uncertain. It is important to remember that the absence of an association between a SNP in a certain gene and a partic-

Table 1. Heritability for hemostasis-related phenotypes.

Phenotype	*Heritability (%)*	*Design/Study*	*Reference*
APC resistance	71	Extended pedigree/GAIT	77
	58	Sib-pairs/ VITA	78
FVII	52	Extended pedigree/GAIT	77
	63	Twins	79
	33	Healthy families	80
Protein C	50	Extended pedigree/GAIT	77
	24	Sib-pairs/VITA	78
	41	Extended pedigree	81
FXIIIs	48	Extended pedigree/GAIT	77
	41	Twins	79
FVIII	40	Extended pedigree/GAIT	77
	61	Twins	79
	57	Twins	82
FIX	39	Extended Pedigree/GAIT	77
	50	Extended pedigree	81
	20	Twins	82
Fibrinogen	34	Extended pedigree/GAIT	77
	44	Twins	79
	50	Twins	83
	36	Segregation analysis	84
	30	Extended pedigree	81
	28	Healthy families	80
von Willebrand	32	Extended pedigree/GAIT	77
factor	75	Twins	79
	25	Extended pedigree	81
TAFI Ag	82	Twins	85
TFPI	52	Extended pedigree/GAIT	77
Antithrombin	48	Extended pedigree/GAIT	77
	40	Sib-pairs/ VITA	78
Free protein S	46	Extended pedigree/GAIT	77
Func protein S	45	Extended pedigree/GAIT	77
FXI	45	Extended pedigree/GAIT	77
FV	44	Extended pedigree/GAIT	77
	71	Extended pedigree	81
FX	43	Extended pedigree/GAIT	77

ular disease does not rule out an association for the whole gene itself. In addition, although association studies provide important indirect evidence for genetic effects of a particular gene, it is abundantly clear from the literature that such studies have serious limitations (using case-control designs). Thus, one must use caution when drawing conclusions from association studies.[25,26] Fortunately, there are some editorials that present explicit guidelines for publishing genetic association studies of complex diseases.[27,28] Some of these limitations can be avoided by identifying groups of markers that are inherited together in families and in populations. These groups are known as *haplotypes*. It is hypothesized that only a few SNP are needed to characterize the most common haplotypes occurring within each haplotype block; these are termed tag-SNP. However, tag-SNP have been discovered for only a small number of genes. In the future, because of the application of the haplotype map[29] to delineate these tag-SNP, haplotype-based association analyses are likely to replace single SNP-based analyses.

Recently, several studies have reported that certain haplotypes in different candidate genes are associated with increased risk of thrombotic disease.[30,31] However, these results have not been replicated in other studies.[32-34]

In addition to association studies, linkage analysis using family-based designs (sib pairs, small nuclear families and large extended pedigrees) is an efficient method for identifying susceptibility loci. In linkage analysis, a number of DNA markers are evenly dispersed at a known chromosomal locations throughout the genome or at nearby candidate genes. The DNA markers are usually variable numbers of repeats (short tandem repeats or microsatellites). They need not be part of a functional gene – they are just landmarks with known location in the genome. For each DNA marker, evidence of linkage is derived using statistical procedures that trace the co-segregation of the trait and a specific variant of the DNA marker along the familial lineage in multiple generations.[35] Such methods do not require linkage disequilibrium and are not susceptible to population stratification. Another advantage is that a given DNA marker identifies a whole chromosomal region, so some generalizations about a gene or a chromosomal region can be made if no linkage is found.

Early pronouncements on the failure of the linkage approach to analyze complex diseases were premature and based on comparing sub-optimal linkage research designs with overly optimistic association-based designs.[36] It is now well-known and widely accepted that the linkage approach is enormously successful in identifying genes involved in monogenic diseases. Although some authors believe that the linkage approach is not suitable for identifying genes underlying complex diseases[6,37] recently there has been substantial progress in the positional cloning of genes based on linkage studies of human pedigrees. These studies focus on measuring quantitative genetic traits that are correlated with the risk of disease.[38,39] This represents a novel and promising approach using linkage for initial chromosomal localization and then gene identification by association with positional candidate genes.[40,41]

Linkage and association methods are not competitors, rather they are complementary.[42] The unpre-

dictable nature of linkage disequilibrium across the genome and the fact that single makers may represent entire chromosomal regions imply that linkage methods are likely to be more efficient than association methods for initial gene localization.[24] On the other hand, the limited extent of linkage disequilibrium and the large regions identified by linkage suggest that association methods are likely to be more useful than linkage for narrowing in on specific genes.

Local or global approach: candidate genes versus whole-genome scanning

Heritable diseases are analyzed better by localizing genes that confer susceptibility and by subsequent characterization of these genes. To localize genes for multifactorial traits, such as thrombotic diseases, two different strategies are followed: hypothesis-driven candidate genes (local) and hypothesis-free genome scanning (global) approaches.

The analysis of candidate genes, whose biological functions are known and tested for association with the disease phenotype, is currently in vogue. Most of our knowledge of the genetic factors involved in common thrombotic diseases has been derived from association studies that compared cases versus controls to evaluate whether the candidate gene has any effect on the risk factor (phenotype) or on the disease risk.[43] These designs became popular after identification of the factor V Leiden (FVL) mutation, which has been considered the paradigm of candidate gene association studies.[37] The FVL mutation was identified after the application of an imaginative strategy based on an abnormal result of a clotting test[44] followed by the identification of a DNA coding variant in an obvious candidate gene with a likely functional role.[45] This represents association mapping at its best, but unfortunately this is the exception and not the rule for finding genetic factors affecting complex diseases. A clearer picture of the limitations of this method has been provided by recent studies on the role of genes in thrombosis. Thus, some cautionary considerations on the significance of this strategy need to be understood.

First, although many features of the hemostatic and fibrinolytic systems have been described, we know very little about the complete function of all but a few gene products. We know even less about the detailed molecular biology of most pathways and the proteins outside the coagulation system involved in the pathogenesis of venous thrombosis. Consequently, we must look for candidate genes among the pathways that we already know. Even so, we may still overlook causative genes because we do not know what other biological systems are involved.

Second, if we have information on the candidate gene involved in the etiology of a particular phenotype, it is not unreasonable to test for association between the candidate SNP with predicted function (e.g. non-synonymous coding single nucleotide polymorphisms, cSNP) and the phenotype. However, there are 20-30 common polymorphisms per gene in our data set, many of which could be potential nucleotide sites that change the quantitative trait (quantitative trait nucleotide, QTN). To date, it is not possible to predict whether most non-coding polymorphisms are functional. As we mentioned above, a major drawback of testing candidate SNP is that the lack of association with a candidate SNP does not rule out that nearby SNP are important functionally. As a matter of fact, the lack of association does not rule out the involvement of the gene itself.

Third, it is important to note that the candidate gene approach has been limited to the analysis of only a few DNA variants. Fortunately, the availability of state-of-the-art genotyping and sequencing techniques, and statistical methodologies allow us to identify and analyze a large number of DNA variants. Despite these technical advances, knowledge of the nature of genetic variation is incomplete for the majority of genes;[46] therefore, selection of SNP for association analyses in such candidate genes is effectively random.

We would like to emphasize that when using the association approach, the magnitude of linkage disequilibrium is a critical requirement for success in localizing causative SNP. Obviously, with only a few common SNP it is unlikely that the causative variant(s) or haplotypes will be picked up, mainly because of the relative unpredictability of linkage disequilibrium and its large variance.[47] This is especially true if the analyses are restricted to frequent coding variants (coding SNP) when there is evidence that rare variants, located in non-coding regions too, might be implicated in the variability of quantitative traits.[48] These rare variants might not be in linkage disequilibrium with other variants within the gene. As a result, they might not be correlated with any haplotype[38,49] and so they would never be detected by an association method.

The initial promise that genetic risk factors for thrombotic disease would be identified has, therefore, remained largely unfulfilled. To make matters worse, the expectations raised by early reports of positive associations have been tempered by inconsistent results for almost all of the genes studied.[43] It follows that case-control association may provide indirect evidence of a genetic effect, but it seems clear that it is not an optimum strategy to search for genetic factors that underlie a complex disease.[47]

The role of candidate genes in thrombotic disease

has also been assessed by linkage analysis.[50,51] This approach differs from association studies because it allows for confirmation[21,52-54] or, more importantly, the exclusion[50,51,54,55] of the role of candidate genes.

The localization of a gene that underlies a complex disease is a difficult task and requires the identification of true causative polymorphisms in that gene. Thus, the challenge now is how to identify those risk factors for thrombotic disease that are unknown. It must be emphasized that the causative basis of complex thrombotic disease involves a dynamic interplay among multiple genes in addition to gene-gene and gene-environment interactions, thus a *genomic* (global) approach that examines the entire genome is clearly desirable.

One of the most significant advances in identifying causative genes is the use of whole genome scanning. This remarkable technology has made the identification of human quantitative trait loci[39,56] a reality. Genome scans identify chromosomal regions that show evidence of linkage with a disease/trait phenotype. Typically, these linkage signals are detected with microsatellite polymorphism DNA markers that are distributed across the genome. They are highly polymorphic markers that consist of tandem repeats of 2-5 base pairs. The repeat units are highly susceptible to mutations that increase or decrease the number of repeats. The length polymorphism that results is readily analyzed using standard polymerase chain reaction methods. The generally accepted density of information for initial linkage analyses is a linkage map of markers with average spacing less than or equal to 10 centiMorgan (cM), which corresponds to about a 10 million DNA base pairs in a physical genome map.

Genes localized by genome scans, called *positional* candidate genes, might correspond to the initial candidate genes (encoding proteins with clear implications in the biochemical pathways of the disease). However, it is more likely that these genes were not expected and unknown. In this context, although the hunt for genes influencing thrombotic risk is still in its infancy, there are several examples of progress (Table 2).

The results of the GAIT Project are testimony to the success of linkage analysis. An excellent example is that we were able to detect the effect of factor V Leiden on the APC resistance phenotype by a family-based linkage analysis.[52] It is important to note that in the same report we identified a novel quantitative trait locus on chromosome 18 that jointly influences APC resistance and thrombotic risk.[52] It is unlikely that we would have detected this quantitative trait locus if we had employed the limited candidate gene strategy, because APC resistance is a risk factor for thrombosis independent of FVL.[13] Indeed, this conclusion is further supported by the fact that there are no candidate genes in the linkage region on chromosome 18.

Table 2. Results from whole genome-scans for thrombosis-related phenotypes.

Phenotypes	*Statistical value**	*Location*	*Gene*	*Reference*
Thrombosis	<0.0001	10p12	?	60
Thrombosis	<0.0007	18p11.2	?	60
Thrombosis	<0.0003	11q23	?	60
FII-Thrombosis	4.70 (p=1.5×10^{-6})	11p11	*F2*	22
FXII-Thrombosis	10.21 (p=3.6×10^{-12})	5q35	*F12*	23
FXII	3.53 (p=3.5×10^{-5})	10p13	?	23
Free protein S	4.07 (p=7.5×10^{-6})	1q32	*C4BP*	54
APCR/FVIII-Thrombosis	4.50 (p=3.1x10^{-5})	18p11	?	48
vWF	3.46 (p=0.00003)	9p34	*ABO*	49
Protein C	3.69 (p=0.00002)	16q23	*NQO 1*	53
FVIII	4.44 (p=1.5×10^{-6})	5/11	?	57
Homocysteine	3.01 (p=0.0001)	11q23	*NNMT*	61
Fibrinogen	3.12 (p=7.5×10^{-5})	14q11	?	55

**LOD score or p value; ?candidate gene unknown.*

Another example from the GAIT Project is the identification of a major quantitative trait locus on chromosome 16 that influences the levels of protein C, whereas the *PROC* gene that encodes protein C (an obvious candidate gene) is a weak determinant of the protein C levels.[57] Also, using a genome scan, the major determinant of free protein S levels was identified as a quantitative trait locus on chromosome 1, whereas its structural gene is located on chromosome 3.[58]

Our genome scanning to identify chromosome regions containing genes that influence fibrinogen level also revealed a region showing strong evidence of linkage on chromosome 14.[59] It is worth noting that in the multipoint genetic analysis the region on chromosome 4 that contains the fibrinogen chains of the structural genes (*FGA*, *FGB* and *FGG*) showed little evidence of linkage to fibrinogen (LOD < 1). This observation agrees with previous results from the Framingham Heart Study,[60] in which no significant results were reported from a genome-scan for fib-

rinogen level. In our study, there are no hemostasis-related candidate genes in the region of the linkage signal on chromosome 14 that might influence fibrinogen levels.[59]

Even more relevant is that the factor XII level is affected by a region on chromosome 5 where the structural *F12* gene has been mapped. It is also affected by another region on chromosome 10 that contains an unknown positional candidate gene that influences factor XII plasma levels. A single DNA variant (46C/T polymorphism) in the *F12* gene seems to be a major determinant of factor XII levels and it also influences susceptibility to thrombosis.[22]

More recently, a genome scan linkage analysis in families with high factor VIII levels has been reported. Two linkage regions under an imprinting model on chromosome 5 and 11 were identified.[61]

Another example that demonstrates the utility of genome scanning is a study of a large Vermont pedigree initially described in the context of a simple Mendelian trait of type I protein C deficiency.[62] This study provided strong evidence that the protein C mutation was not the only genetic factor co-segregating in this family and involved in thrombotic risk.[63] After many *candidate genes* had been excluded by linkage analyses,[51] three genomic regions that might be involved in thrombotic risk were localized by a whole genome scan.[64] Interestingly, all of these regions represent replications already reported by the GAIT Project.[22,52,65]

The fact that these studies consistently implicate new chromosome regions with no known thrombosis-related genes, supported by the fact that these regions have been identified in other studies, argues that the global whole genome scanning approach is exceedingly productive. Despite this progress, the important task still remaining is to identify the genes in these regions and their causative alleles (quantitative trait nucleotides).

From localization to identification: the role of bioinformatics and biostatistics

Once linkage has been found, the search for the causative gene within the linked region can begin. Figure 1 shows a simple algorithm in the study of human complex diseases. Notably, localization of the gene is harder than its detection. The first option is to saturate the chromosomal candidate region with additional genetic markers and simultaneously exploit the information on both linkage and linkage disequilibrium to fine-map the QTL.[66] Using this strategy, quantitative trait nucleotides involved in thrombotic risk have been identified, such as the 46C/T polymorphism in the *F12* gene[22] and the G20210A mutation in the *F2* gene,[21] thanks to the presence of an obvious positional candidate gene in the region of linkage. However, the majority of the localized QTL remain to be identified (Table 2).

For most disorders, the linkage region is generally large; thus, we may be looking for a gene in a region of 10-20 million base pairs. A region that size is likely to contain 50-100 genes. It is, therefore, a major undertaking to identify the pertinent gene. One might characterize this task as the proverbial finding a needle in a haystack.

In such a situation, efficient positional cloning must accommodate a systematic, preliminary evaluation of all genes within the region so that the most promising candidates are given priority for detailed study.

To this end, a new discipline – *bioinformatics* - has emerged. Bioinformatics can guide and attribute priority to positional candidates genes by an *in silico* (computer analyses) investigation of the linkage region through the integration of different types of information, such as nucleic acid (DNA/RNA) and protein sequences, structures, functions, pathways and interactions.[67] Bioinformatics uses computer software for creating databases, data management, data warehousing, data mining and global communication networks.[68] In addition, functional genomics, biomolecular structure, transcriptome and proteome analyses, cell metabolism and biodiversity are some of the areas in which bioinformatics is an integral component.[69] Microarray data represent a new type of information that can provide important insights into the interactions of genes and thus can complement the statistical approaches to gene mapping.[70]

Integration of all this knowledge will provide a resource of an unprecedented power to decipher the intricacies of complex disease.[71] Although some of these areas are currently being developed, access to biological sequence and structure databases via the world wide web (Table 3) and powerful bioinformatics tools are already available for public use.[72]

Once genes have been selected *in silico* for study, they are screened for SNP, moving from the localization of a gene to the identification of its causative alleles (quantitative trait nucleotide). Current projects can sequence whole genes and it is usual to find dozens of polymorphisms within each gene. When studying a small set of candidate genes at one locus having one putative quantitative trait locus, it may easily be necessary to consider more than 100 polymorphisms. However, almost all genes contain SNP and complexity increases at the protein level since one human gene may produce up to five different proteins. As a consequence of this complexity, it is difficult, if not impossible, to predict the biological or clinical effect of a SNP in a given gene. Fortunately, new biostatistical and genetic epidemiology tools, which can manage multiple testing on the one hand

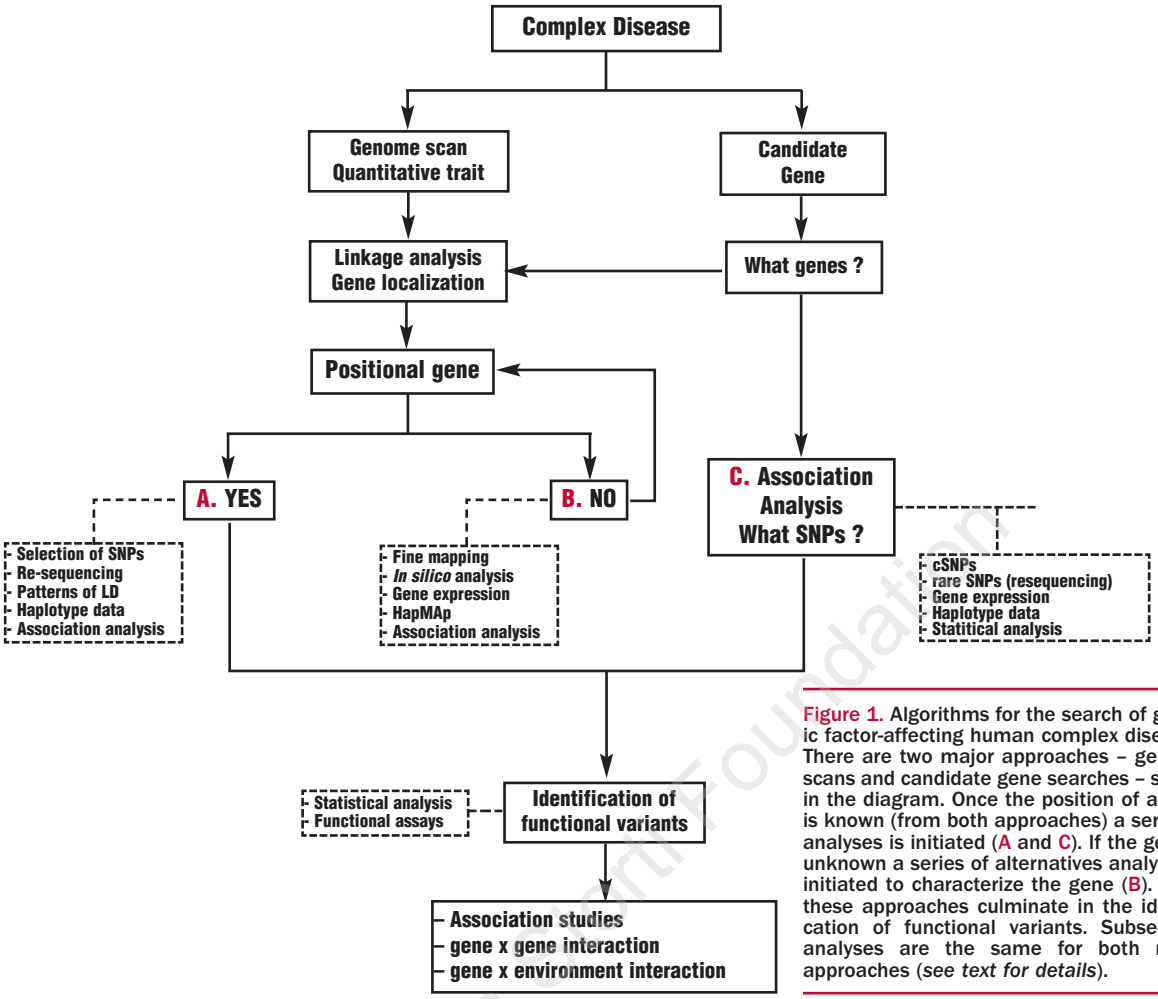

Figure 1. Algorithms for the search of genetic factor-affecting human complex diseases. There are two major approaches – genome scans and candidate gene searches – shown in the diagram. Once the position of a gene is known (from both approaches) a series of analyses is initiated (A and C). If the gene is unknown a series of alternatives analyses is initiated to characterize the gene (B). All of these approaches culminate in the identification of functional variants. Subsequent analyses are the same for both major approaches (*see text for details*).

Table 3. Selected world wide web (WWW) bioinformatics resources.

Ensembl browser	*http://www.ensembl.org*
NCBI Entrez	*http://www.ncbi.nlm.nih.gov/gquery/gquery.fcgi*
UCSC browser	*http://genome.ucsc.edu*
OMIM	*http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM*
ExPASy	*http://us.expasy.org*
dbSNP	*http://www.ncbi.nlm.nih.gov/projects/SNP*
HapMap	*http://www.hapmap.org/index.html.en*
EMBOSS	*http://emboss.sourceforge.net/what*
mVISTA	*http://genome.lbl.gov/vista/index.shtml*

and the exponential number of possible models on the other, are being developed.[70] Although these disciplines are still young and have technical problems, a number of interesting methods have been introduced, such as: Bayesian methods, neural networks, principal component analysis, cluster analysis, support vector machines, genetic algorithms and tree-based methods.[73] All of these methods will help select which DNA variant(s) among the total number in a gene has/have the highest probability of being functional.

However, finding a gene is not enough, one must provide evidence of its biological effect, and also whether or not its specific variants will be functional when they are transferred into cellular or animal models of the disease. In addition, environmental factors can reveal or facilitate the phenotypic expression of such susceptibility genes. Indeed, in common diseases, genetic effects can be amplified considerably in the presence of triggering factors. There is now accumulating evidence that most of the susceptibility genes for common diseases do not have a primary etiological role in predisposition to disease, but rather act as response modifiers to exogenous factors such as stress, environment, disease, or drug intake.[74] A better characterization of the interactions between environmental and genetic factors constitutes a key issue in the understanding of the pathogenesis of multifactorial diseases. However, genes and environment interact with each other in extraordinarily complex ways, as well as among themselves to produce the final result of the phenotype. Thus, clinical and

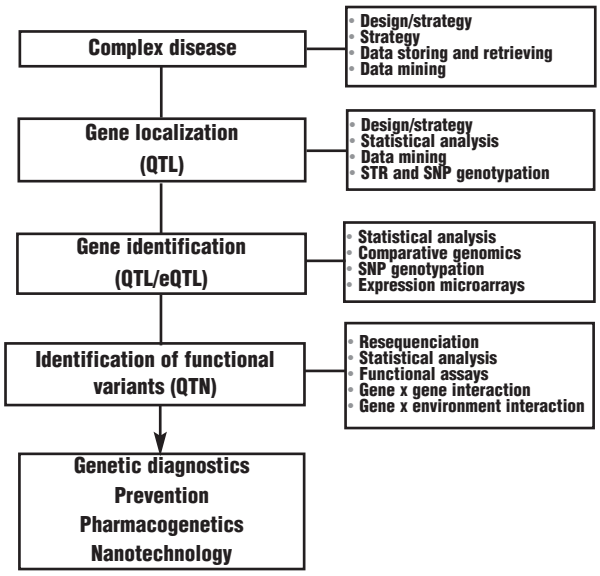

Figure 2. Flow diagram showing the pathway from the genetic analysis of a specific complex disease to the clinical application of the results. To the right of each descending step in the diagram are the strategies and resources for completing that step (*see text for details*).

epidemiological studies together with a battery of sophisticated statistical tools will be needed to understand this complexity.[75] Obviously, high-throughput genotyping technologies will also be required. Figure 2 summarizes the sequence of strategies and resources that can be used in analyses of human complex diseases.

However, finding a gene and its variants is only the end of the beginning. Ultimately, genomic research must be translated into clinical insights.[76] Therefore, identification of the genes contributing to inter-individual variation in disease risk may facilitate early identification of patients who are at elevated risk of thrombotic disease before the onset of any clinical symptoms. In the end more efficacious treatments will have to be developed by exploring previously unidentified metabolic and physiological pathways, and tailoring particular treatments to patients who are most likely to respond on the basis of their genetic constitution.

Conclusions

The field of genetics of common complex human diseases has advanced enormously over the past several years. During the course of these advances, a major paradigmatic change has been taking place that focuses on the genetic analysis of measurable quantitative traits that are correlated with disease risk. Exploiting this genetic information using genome-wide scans to detect linkage in family-based studies represents a formidable tool for localizing and identifying genes underlying susceptibility to complex diseases. We believe that joint linkage and association methods will play an important role in identifying genes. The point is that these two methods are not competing strategies, rather they are complementary. The final solution to the understanding of complex diseases will need a productive and extensive collaboration among clinicians, geneticists, biostatisticians and experts in bioinformatics. The wealth of information that molecular genetic studies, such as functional genomics, biomolecular structure, transcriptome and proteome analyses, cell metabolism, nanotechnology and biodiversity, can provide will have to be integrated to continue making major gains in the understanding of the genetics of human complex diseases. This will have an enormous impact in the field of thrombosis, cardiovascular diseases and medicine in general,[76] improving the prevention and treatment of very common and disabling disorders affecting millions of citizens.

Appendix

Glossary

Allele. Alternative form of a genetic locus; a single allele for each locus is inherited from each parent.

Association. The occurrence of a particular allele of a polymorphism in a group of patients more frequently than would be expected by chance.

Bioinformatics. The science of managing and analyzing biological data using advanced computing techniques. Especially important in analyzing genomic research data.

Candidate gene. A gene known to be located in a chromosome region suspected of being involved in a disease.

Complex disease. Both genetic and environmental factors contribute to pathogenesis. No clear Mendelian pattern of inheritance is discernible.

Complex trait. Trait that has a genetic component that does not follow strict Mendelian inheritance. May involve the interaction of two or more genes or gene-environment interactions.

DNA polymorphism. A variation in DNA sequence that occurs in at least 1% of the general population. Most polymorphisms contribute to normal human diversity rather than disease, but a small proportion will contribute to susceptibility to common disorders.

DNA mutation. A permanent structural alteration in DNA that occurs in less than 1% of the general population. In most cases, DNA changes either have no effect or cause harm.

Functional genomics. The study of genes, their resulting proteins, and the role played by the proteins in the body's biochemical processes.

Haplotype. A way of denoting the collective genotype or cluster of a number of closely linked loci on a

chromosome. The groups of alleles on a chromosome which tend to consegregate together.

Human haplotype map (HapMap). The objective of the International HapMap Project is to produce a resource that describes the haplotypes in the human genome and the SNP that tag them. It is estimated that roughly 300,000 to 500,000 tag-SNP can be chosen that contain most of the information on the patterns of variation of the 10 million common SNP in the human genome.

In silico. Method to test biological models, drugs and medical interventions using sophisticated computer models rather than expensive laboratory (*in vitro*) and animal experiments (*in vivo*). Also, integrated methods of genomics and life science informatics.

Linkage disequilibrium. When alleles occur together more often than can be accounted for by chance. Indicates that the two alleles are physically close on the DNA strand (chromosome).

Linkage. Consistent co-inheritance of the genetic markers (such as DNA polymorphisms) with the disease in many families indicates that it is in close proximity to the actual disease gene, and is said to be *linked*.

Locus (pl. *loci*). The position on a chromosome of a gene or other chromosome marker; also, the DNA at that position. The use of locus is sometimes restricted to mean expressed DNA regions.

LOD score. A statistical estimate of linkage (logarithmic odds ratio) indicating whether two loci are likely to lie near each other on the same chromosome and are therefore likely to be inherited together. A LOD score of three or more is generally taken to indicate that the two loci are linked.

Genetics.The study of inheritance patterns of specific traits.

Genetic marker. A specific feature (i.e., SNP, short tandem repeat or gene) at an identified physical location on a chromosome, whose inheritance can be followed. The position of a gene implicated in a particular phenotype can be defined through its linkage to such markers.

Genomics. The study of genes and their function.

Pharmacogenomics. The identification of the genes which influence individual variation in the efficacy or toxicity of therapeutic agents, and the application of this information in clinical practice.

Quantitative trait locus (QTL). Chromosomal location of a genetic factor that affects a trait (e.g., levels of fibrinogen) that is measured on a quantitative (linear) scale.

Quantitative nucleotide trait (QTN). Nucleotide location of a polymorphism that affects a trait (e.g., levels of fibrinogen) that is measured on a quantitative (linear) scale.

Quantitative trait. Trait that is measured on a quantitative (linear) scale. These traits are typically affected by more than one gene, and also by the environment.

Single nucleotide polymorphism (SNP). A DNA polymorphism that involves a change in a single base of a DNA sequence (for example, ACGT to AGGT). The human genome is likely to contain about 10 million SNP.

Tag single nucleotide polymorphism (tag-SNP). Individual SNP that unambiguously define a particular haplotype in a block (haplotype tag SNP or ht- SNP), a much smaller number of typing assays will be sufficient to predict the pattern of alleles for all the other SNP in the same block without having to genotype them individually.

Tandem repeat sequences. Multiple copies of the same base sequence on a chromosome; used as markers in physical mapping.

Transcriptome. The full complement of activated genes, mRNA, or transcripts in a particular tissue at a particular time

Both authors contributed equally to the study. The authors declare that they have no potential conflict of interest.

The authors are also indebted to Juan Carlos Souto, Alfonso Buil and William H. Stone for their advice on and helpful discussion of this manuscript. This study was partially supported by grants HL70751 from the USA NIH; FIS 02/0375 from the Fondo Investigación Sanitaria, SAF2002-03449 partially supported by FEDER funds (Spanish Ministry of Science and Technology), RECAVA 03/01 from the Fondo Investigación Sanitaria and from Fundació "La Caixa" and Fundació d'Investigació Sant Pau. JMS was supported by FIS 99/3048 from the Fondo Investigación Sanitaria (Spanish Ministry of Health).

Manuscript received March 9, 2005. Accepted July 14, 2005.

References

1. Coon WW, Park WW, Keller JB. Venous thromboemolism and other venous disease in the Tecumseh Community Health Study. Circulation 1973;48:839-46.
2. Nordstrom M, Lindblad B, Bergqvist D, Kjellstrom T. A prospective study of the incidence of deep-vein thrombosis within a defined urban population. J Int Med 1992;232:155-60.
3. Rosendaal FR. Venous thrombosis: a multicausal disease. Lancet 1999; 353: 1167-73.
4. Lane DA, Mannucci PM, Bauer KA, Bertina RM, Bochkov NP, Boulyjenkov V, eta al. Inherited thrombophilia: part 1. Thromb Haemost 1996;76:651-62.
5. Bertina RM. Genetic approach to thrombophilia. Thromb Haemost 2001; 86:92-103.
6. Rosendaal FR. Genetic studies in complex disease: the case pro association studies. J Thromb Haemost 2003; 1: 1679-80.
7. Souto JC. Genetic studies in complex disease: the case pro linkage studies. J Thromb Haemost 2003;1:1676-8.
8. Duggirala R, Williams JT, Williams-Blangero S, Blangero J. A variance component approach to dichotomous trait linkage analysis using a threshold model. Genet Epidemiol 1997;14:987-92.
9. Koster T, Blann AD, Briet E, Vandenbroucke JP, Rosendaal FR. Role of clotting factor VIII in effect of von Willebrand factor on occurrence of deep-vein thrombosis. Lancet 1995; 345: 152-5.
10. Folsom AR, Wu KK, Rosamond WD, Sharrett AR, Chambless LE. Prospective study of hemostatic factors

and incidence of coronary heart disease: the atherosclerosis risk in communities (ARIC) study. Circulation 1997;96:1102-8.
11. Den Heijer M, Koster T, Blom HJ, Bos GMJ, Briet E, Reitsma PH, et al. Hyperhomocysteinemia as a risk factor for deep-vein thrombosis. N Engl J Med 1996;334:759-62.
12. Nygard O, Nordrehaug JE, Refsum H, Ueland PM, Farstad M, Vollset SE. Plasma homocysteine levels and mortality in patients with coronary artery disease. N Engl J Med 1997;337:230-6.
13. De Visser MC, Rosendaal FR, Bertina RM. A reduced sensitivity for activated protein C in the absence of factor V Leiden increases the risk of venous thrombosis. Blood 1999;93:1271-6.
14. Kohler HP, Carter AM, Strickland MH, Grant PJ. Levels of activated FXII in survivors of myocardial infarction – association with circulating risk factors and extent of coronary artery disease. Thromb Haemost 1998;79:14-8.
15. Carter AM, Catto AJ, Grant PJ. Determinants of tPA antigen and associations with coronary artery disease and acute cerebrovascular disease. Thromb Haemost 1998;80:632-6.
16. van Hylckama Vlieg A, van der Linden IK, Bertina RM, Rosendaal FR. High levels of factor IX increase the risk of venous thrombosis. Blood 2000; 95: 3678-82.
17. Meijers JC, Tekelenburg WL, Bouma BN, Bertina RM, Rosendaal FR. High levels of coagulation factor XI as a risk factor for venous thrombosis. N Engl J Med 2000;342:696-701.
18. Dahm A, Van Hylckama Vlieg A, Bendz B, Rosendaal F, Bertina RM, Sandset PM. Low levels of tissue factor pathway inhibitor (TFPI) increase the risk of venous thrombosis. Blood 2003;101:4387-92.
19. Souto JC, Almasy L, Borrell M, Blanco-Vaca F, Mateo J, Soria JM, et al. Genetic susceptibility to thrombosis and Its relationship to physiological risk factors: The GAIT study. Am J Hum Genet 2000;67:1452-1459.
20. Almasy L, Dyer TD, Blangero J. Bivariate quantitative trait linkage analysis: Pleiotropy versus coincident linkage. Genet Epidemiol 1997; 14: 953-8.
21. Soria JM, Almasy L, Souto JC, Tirado I, Borell M, Mateo J, et al. Linkage analysis demonstrated that the prothrombin G20210A mutation jointly influences plasma prothrombin levels and risk of thrombosis. Blood 2000; 95: 2780-5.
22. Soria JM, Almasy L, Souto JC, Bacq D, Buil A, Faure A, et al. A quantitative-trait locus in the human factor XII gene influences both plasma factor XII levels and susceptibility to thrombotic disease. Am J Hum Genet 2002; 70: 567-74.
23. Heit JA, Phelps MA, Ward SA, Slusser JP, Petterson TM, De Andrade M. Familial segregation of venous thromboembolism. J Thromb Haemost 2004; 2:731-6.
24. Terwilliger JD, Weiss KM. Linkage disequilibrium mapping of complex disease: fantasy or reality? Curr Opin Biotechnol 1998;9:578-94.
25. Almasy L, MacCluer JW. Association studies of vascular phenotypes: how and why? Arterioscler Thromb Vasc Biol 2002;22:1055-7.
26. Gambaro G, Anglani F, D'Angelo A. Association studies of genetic polymorphisms and complex disease. Lancet 2000;355:308-11.
27. Editorial. Freely associating. Nat Genet 1999;22:1-2.
28. Cooper DN, Nussabaum RL, Krawczak M. Proposed guidelines for papers describing DNA polymorphism-disease associations. Hum Genet 2002;110:207-8.
29. The International HapMap Consortium. The International HapMap Project. Nature 2003; 426:789-96.
30. Saposnik B, Reny JL, Gaussem P, Emmerich J, Aiach M, Gandrille S. A haplotype of the EPCR gene is associated with increased plasma levels of sEPCR and is a candidate risk factor for thrombosis. Blood 2004;103:1311-8.
31. Margaglione M, Bossone A, Coalizzo D, D'Andrea G, Brancaccio V, Ciampa A, et al. FV HR2 haplotype as additional inherited risk factor for deep vein thrombosis in individuals with a high-risk profile. Thromb Haemost 2002; 87:32-6.
32. Uitte de Willige S, Van Marion V, Rosendaal FR, Vos HL, de Visser MC, Bertina RM. Haplotypes of the EPCR gene, plasma sEPCR levels and the risk of deep venous thrombosis. J Thromb Haemost 2004;2:1305-10.
33. Faioni EM, Castaman G, Asti D, Lussana F, Rodeghiero F. Association of factor V deficiency with factor V HR2. Haematologica 2004;89:195-200.
34. de Visser MC, Guasch JF, Kamphuisen PW, Vos HL, Rosendaal FR, Bertina RM. The HR2 haplotype of factor V: effects on factor V levels, normalized activated protein C sensitivity ratios and the risk of venous thrombosis. Thromb Haemost 2000;83:577-82.
35. Vink JM, Boomsma DI. Gene finding strategies. Biol Psychol 2002;61:53-71.
36. Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science 1996;273:1516-7.
37. van Hylckama Vlieg A, Sandkuijl LA, Rosendaal FR, Bertina RM, Vos HL. Candidate gene approach in association studies: would the factor V Leiden mutation have been found by this approach? Eur J Hum Genet 2004; 12: 478-82.
38. Blangero J, Williams JT, Almasy L. Quantitative trait locus mapping using human pedigrees. Hum Biol 2000; 72: 35-62.
39. Korstanje R and Paigen B. From QTL to gene: the harvest begins. Nat Genet 2002;31:235-6.
40. Blangero J. Localization and identification of human quantitative trait loci: king harvest has surely come. Curr Opin Genet Dev 2004;14:233-40.
41. Dahlback B. Successful hunt for quantitative trait locus in thrombophilia. Arterioscler Thromb Vasc Biol 2003; 23:376-7.
42. Terwilliger JD, Goring HH. Gene mapping in the 20th and 21st centuries: statistical methods, data analysis, and experimental design. Hum Biol 2000; 72:63-132.
43. Lane DA, Grant PJ. Role of hemostatic gene polymorphisms in venous and arterial thrombotic disease. Blood 2000;95:1517-32.
44. Dahlback B, Carlsson M, Svensson PJ. Familial thrombophilia due to a previously unrecognized mechanism characterized by poor anticoagulant response to activated protein C: prediction of a cofactor to activated protein C. Proc Natl Acad Sci USA 1993; 90:1004-8.
45. Bertina RM, Koeleman BP, Koster T, Rosendaal FR, Dirven RJ, de Ronde H, et al. Mutation in blood coagulation factor V associated with resistance to activated protein C. Nature 1994; 369: 64-7.
46. Carlson CS, Eberle MA, Rieder MJ, Smith JD, Kruglyak L, Nickerson DA. Additional SNPs and linkage-disequilibrium analyses are necessary for whole-genome association studies in humans. Nat Genet 2003;33:518-21.
47. Terwilliger JD, Haghighi F, Hiekkalinna TS, Goring HH. A biased assessment of the use of SNPs in human complex traits. Curr Opin Genet Dev 2002;12:726-34.
48. Pritchard JK. Are rare variants responsible for susceptibility to complex diseases? Am J Hum Genet 2001;69:124-37.
49. Clark AG. Finding genes underlying risk of complex disease by linkage disequilibrium mapping. Curr Opin Genet Dev 2003;13:296-302.
50. Espinosa-Parrilla Y, Morell M, Souto JC, Borrell M, Heine-Suner D, Tirado I, et al. Absence of linkage between type III protein S deficiency and the PROS1 and C4BP genes in families carrying the protein S Heerlen allele. Blood 1997;89:2799-806.
51. Scott BT, Bovill EG, Callas PW, Hasstedt SJ, Leppert MF, Valliere JE, et al. Genetic screening of candidate genes for a prothrombotic interaction with type I protein C deficiency in a large kindred. Thromb Haemost 2001; 85: 82-7.
52. Soria JM, Almasy L, Souto JC, Buil A, Martinez-Sanchez E, Mateo J, et al. A new locus on chromosome 18 that influences normal variation in activated protein C resistance phenotype and factor VIII activity and its relation to thrombosis susceptibility. Blood 2003; 101:163-7.
53. Souto JC, Almasy L, Muniz-Diaz E, Soria JM, Borrell M, Bayen L, et al. Functional effects of the ABO locus polymorphism on plasma levels of von Willebrand factor, factor VIII, and activated partial thromboplastin time. Arterioscler Thromb Vasc Biol 2000; 20:2024-8.
54. De Visser MC, Sandkuijl LA, Lensen RP, Vos HL, Rosendaal FR, Bertina RM. Linkage analysis of factor VIII and von Willebrand factor loci as quantitative trait loci. J Thromb Haemost 2003; 1: 1771-6.
55. Scott BT, Hasstedt SJ, Bovill EG, Callas PW, Valliere JE, Wang L, et al. Characterization of the human prostaglandin H synthase 1 gene (PTGS1): exclusion by genetic linkage analysis as a second modifier gene in familial thrombosis. Blood Coagul Fibrinol 2002;13:519-31.
56. Blangero J, Williams JT, Almasy L. Novel family-based approaches to genetic risk in thrombosis. J Thromb Haemost 2003;1:1391-7.
57. Buil A, Soria JM, Souto JC, Almasy L, Lathrop M, Blangero J, et al. Protein C levels are regulated by a quantitative trait locus on chromosome 16: results from the Genetic Analysis of Idiopathic Thrombophilia (GAIT) Project. Artherios Thromb Vasc Biol 2004;24: 1321-5.
58. Almasy L, Soria JM, Souto JC, Coll I,

Bacq D, Faure A, et al. Genetic analysis of idiopathic thrombophilia project. A quantitative trait locus influencing free plasma protein S levels on human chromosome 1q: results from the Genetic Analysis of Idiopathic Thrombophilia (GAIT) project. Artherios Thromb Vasc Biol 2003; 23:508-11.
59. Soria JM, Almasy L, Souto JC, Buil A, Lathrop M, Blangero J, et al. A genome search for genetic determinants that influence fibrinogen plasma levels. Artherioscler Thromb Vasc Biol 2005; 25:1287-92.
60. Yang Q, Tofler GH, Cupples LA, Larson MG, Feng D, Lindpaintner K, et al. A genome-wide search for genes affecting circulating fibrinogen levels in the Framingham Heart Study. Thromb Res 2003;110:57-64.
61. Berger M, Mattheisen M, Kulle B, Schmidt H, Oldenburg J, Bickeboller H, et al. High factor VIII levels in venous thromboembolism show linkage to imprinted loci on chromosomes 5 and 11. Blood 2005;105:638-44.
62. Tomczak JA, Ando RA, Sobel HG, Bovill EG, Long GL. Genetic analysis of a large kindred exhibiting type I protein C deficiency and associated thrombosis. Thromb Res 1994;74:243-54
63. Hasstedt SJ, Bovill EG, Callas PW, Long GL. An unknown genetic defect increases venous thrombosis risk, through interaction with protein C deficiency. Am J Hum Genet 1998; 63: 569-76.
64. Hasstedt SJ, Scott BT, Callas PW, Vossen CY, Rosendaal FR, Long GL, et al. Genome scan of venous thrombosis in a pedigree with protein C deficiency. J Thromb Haemost 2004;2:868-73.
65. Souto JC, Blanco-Vaca F, Soria JM, Buil A, Ordoñez-Llanos J, Lathrop M, et al. A Genome-wide exploration suggests a new candidate gene at chromosome 11q23 as the major determinant of plasma homocysteine levels: results from the GAIT. Am J Hum Genet 2005; 76:925-33.
66. Almasy L, Williams JT, Dyer TD, Blangero J. QTL detection using combined linkage/disequilibrium analysis. Genet Epidemiol 1999;17:S31-6.
67. Searls DB. Bioinformatics tools for whole genomes. Annu Rev Genomics Hum Genet 2000;1:251-79.
68. Buckingham S. Bioinformatics: data's future shock. Nature 2004;428:774-7.
69. Roos DS. Computational biology. Bioinformatics-trying to swim in a sea of data. Science 2001;291:1260-1.
70. Ott J, Hoh J. Statistical approaches to gene mapping. Am J Hum Genet 2000; 67:289-94.
71. Debes JD, Urrutia R. Bioinformatics tools to understand human diseases. Surgery 2004;135:579-85.
72. Frishman D, Kaps A, Mewes HW. Online genomics facilities in the new millennium. Pharmacogenomics 2002; 3:265-71.
73. Bastone L, Reilly M, Rader DJ, Foulkes AS. MDR and PRP: a comparison of methods for high-order genotype-phenotype associations. Hum Hered 2004; 58:82-92.
74. Cooper A. Gene–environment interactions and the etiology of common complex disease. Ann Intern Med 2003;139:437-40.
75. Kelada SN, Eaton DL, Wang SS, Rothman NR, Khoury MJ. The role of genetic polymorphisms in environmental health. Environ Health Perspect 2003;111:1055-64.
76. Holzman NA, Marteau TM. Will genetics revolutionize Medicine. N Engl J Med 2000;343:141-4.
77. Souto JC, Almasy L, Borrell M, Gari M, Martinez E, Mateo J, et al. Genetic determinants of hemostasis phenotypes in spanish families. Circulation 2000;101:1546-51.
78. Tosetto A, Castaman G, Cappellari A, Rodeghiero F. The VITA Project: heritability of resistance to activated protein C. Vincenza Thrombophilia and Arteriosclerosis. Thromb Haemost 2000;84:811-4.
79. de Lange M, Snieder H, Ariens RA, Spector TD, Grant PJ. The genetics of haemostasis: a twin study. Lancet 2001;357:101-5.
80. Freeman MS, Mansfield MW, Barrett JH, Grant PJ. Genetic contribution to circulating levels of hemostatic factors in healthy families with effects of known genetic polymorphisms on heritability. Arterioscler Thromb Vasc Biol 2002;22:506-10.
81. Vossen CY, Hasstedt SJ, Rosendaal FR, Callas PW, Bauer KA, Broze GJ, et al. Heritability of plasma concentrations of clotting factors and measures of a prethrombotic state in a protein C-deficient family. J Thromb Haemost 2004;2:242-7.
82. Orstavik KH, Magnus P, Reisner H, Berg K, Graham JB, Nance W. Factor VIII and factor IX in a twin population. Evidence for a major effect of ABO locus on factor VIII level. Am J Hum Genet 1985;37:89-101.
83. Dunn EJ, Ariens RA, de Lange M, Snieder H, Turney JH, Spector TD, et al. Genetics of fibrin clot structure: a twin study. Blood 2004;103:1735-40.
84. Friedlander Y, Elkana Y, Sinnreich R, Kark JD. Genetic and environmental sources of fibrinogen variability in Israeli families: the Kibbutzim Family Study. Am J Hum Genet 1995;56:1194-206.
85. Ariens RA, de Lange M, Snieder H, Boothby M, Spector TD, Grant PJ. Activation markers of coagulation and fibrinolysis in twins: heritability of the prethrombotic state. Lancet 2002; 359:667-71.