Genetic determinants of venous thrombosis

Haan, H.G. de

CHAPTER 8

Discussion

DISCUSSION

The aim of the research conducted for this thesis was to identify novel genetic risk factors for a first and recurrent venous thrombosis. In addition, we investigated whether previously identified genetic risk variants can be used to improve risk stratification for venous thrombosis and we discussed the potential value of using genetic variation to aid causal inferences in observational research. In this chapter, we discuss the main findings and some methodological considerations, and we provide directions for biological and clinical interpretations.

Main findings

So far, variation in seventeen genes, almost all encoding proteins related to hemostasis, have consistently been identified as genetic risk factors for a first VT.[1,2] Evidence from previous GWAS and family studies suggests that additional genetic risk variants are yet to be discovered.[3-6] In addition, the extent to which the identified risk variants contribute to recurrence risk is not clear, nor whether different genetic risk factors play a role in recurrence pathophysiology than those involved in a first event.[7-10] In chapters 2 to 5, we used various strategies to identify variants across the allele frequency spectrum that are associated with the risk of a first or recurrent VT.

In **chapter 2**, we studied the association between a first DVT and genetic variation in the coding regions of 734 genes related to hemostasis. More than 3,500 common variants, identified by next-generation DNA sequencing, were assessed in approximately 900 DVT patients and 600 controls. We confirmed, as expected, the association between DVT and variation in the *F11* region, *FGA-FGG*, *ABO*, and *F5*, which are all established risk loci for VT. At *F5* and the *F11* region we also found evidence for secondary association signals, suggesting that these risk loci contain multiple conditionally independent risk factors for DVT. Remarkably, we found only two suggestive association signals mapping to genes not previously implicated in VT pathophysiology, although these were not replicated in data from the INVENT consortium. In addition, an assessment of over 16,000 rare variants mapping to 647 genes did not reveal a burden of rare variants in DVT patients compared with controls. However, it is possible that associations of both common and rare variants conferring small effects on DVT risk were missed, as our study did not include sufficient patients and controls to identify such variants.

Instead of focusing on variation in candidate genes, we followed an agnostic approach in **chapter 3,** as for recurrence it is unknown whether the same or different genetic risk factors than those identified for a first VT play a role. We conducted a GWAS in which we studied the association between about 8 million common autosomal variants and recurrent VT, followed by a replication study. In addition to confirming the association between FV Leiden and recurrence risk, we identified a novel risk locus at 18q22.1, which was associated with recurrent VT with an odds ratio of 1.7 per minor allele copy in the replication analysis. This intergenic locus may affect recurrence risk by influencing the expression of nearby or distant genes, though further research is needed to unravel the underlying molecular mechanism. We found limited support for previously identified variant associations with recurrence, emphasizing the importance of replication in genetic association analyses.

A first investigation of variation in the Y chromosome and its effect on first and recurrent VT risk was reported in **chapter 4**. As men have an intrinsically higher risk of VT than women[11-16], we postulated that variation in the Y chromosome may increase the risk of VT in subgroups of men. We therefore explored the association between 13 common European Y chromosome haplogroups and the risk of a first and recurrent VT in over 3,700 men. Compared with the most common haplogroup R1b, none of the haplogroups were associated with the risk of a first VT. Specifically, no evidence for an association between haplogroup I, which was previously identified as a risk factor for coronary artery disease[17], and VT risk was observed, even though the analysis was powered to detect a similar association. In addition, we observed some suggestive evidence that carriers of haplogroup R1a had a decreased risk of recurrence compared with R1b-carriers. However, this cannot explain the difference in risk between men and women, as we observed a higher recurrence rate for R1a-carriers than for women.

We used a candidate gene approach in **chapter 5** to study common variation in *CADM1* and the association with the risk of a first VT. An earlier study in a protein C deficient family identified *CADM1*, encoding a cell adhesion molecule involved in endothelial cell migration, as a risk gene for VT.[18,19] To assess whether a joint effect of *CADM1* variation and protein C on VT risk also exists in the general population, we studied the association between over 300 variants in *CADM1* and VT risk in 962 individuals with an abnormality in the protein C pathway and 4004 controls. For six variants we observed a large joint effect on VT risk, of which one variant also showed evidence of an association with

VT in the overall study population of 3496 VT patients and 4004 controls. Due to the high number of statistical tests and low number of individuals with protein C pathway abnormalities, caution is needed when interpreting these results.

In the two remaining chapters, we discussed two of the main applications of genetic risk factors in research, that is risk stratification and Mendelian randomization. Using a panel of 31 previously reported VT risk variants, we constructed genetic risk scores and compared the discriminative values with a model based on clinical risk factors and a combined model (**chapter 6**). We showed that a score containing five risk variants (FV Leiden, PT G20210A, *ABO* non-O, *FGG*-rs2066865, and *F11*-rs2036914) added significant discriminative power to a clinical risk model for venous thrombosis in the general population. As genetic risk profiling is not (yet) cost-effective in the general population, we also explored risk discrimination in clinically relevant subgroups. Except among cancer patients, the genetic risk score performed similarly in the subgroups as in the general population. Replication of our findings in an independent study showed the robustness of our genetic risk score, although the genetic risk score may perform less well in populations with a different ethnic background.

In **chapter 7**, we discussed the possibilities of using genetic variation as an instrument for an exposure of interest to aid causal inference in observational studies. In this educational chapter, we explained that, if none of the Mendelian randomization (MR) assumptions are violated, a genetic instrument can be used to estimate the causal effect of the exposure on the outcome of interest, while minimizing confounding and reverse causation. Although not all assumptions are falsifiable, and a large study population is required, MR studies are increasingly successful applied in observational research, especially when randomized trials are not possible. Outside the scope of this chapter, where we merely described the concepts of MR in general, are the different analytical methods that have recently been developed, including those dealing with pleiotropy.[20]

Methodological considerations

Venous thrombosis is a common complex trait, driven by a multitude of genetic and environmental factors. The first genetic risk factor for VT was suspected over 60 years ago[21], and ever since, studies have aimed to unravel the genetic architecture underlying VT. At first, studies used linkage analyses in families and candidate gene approaches to identify risk genes as the genetic component of common complex traits was thought

to be based on a single gene or few genes each following Mendel's law of inheritance. Technological advances and large collaboratives such as the Human Genome Project[22] paved the way for systematic analysis of millions of (common) variants across the genome. These GWASs fitted the then popular 'common disease – common variant' hypothesis, which claimed that common traits such as VT would be the result of common variants each having a low penetrance.[23,24] Although GWASs identified many risk loci for common complex traits, including several for VT[4,25,26], these loci only explained part of the heritability of each trait.[27,28] For venous thrombosis, Germain *et al*.[3], estimated that common variants could explain around 35% of the genetic variance, of which only 3% could be attributed to the four most well-known risk variants (in *F5, ABO, FGG,* and *F11*). These observations fueled the 'common disease – rare variant' hypothesis, which argued that rare variants with high penetrance contribute substantially to complex trait genetics.[28,29] The advent of high-throughput exome and whole genome sequencing now allows large-scale investigations of rare and even 'private' variants using single-variant and aggregate association tests, though the effect sizes conferred by rare variants seem to be smaller than initially thought.[30-32] For VT risk, most studies have so far focused on rare variants associated with thrombophilia. Lotta *et al*.[33], observed a burden of rare coding variants in *ADAMTS13* associated with a 4.8-fold increased risk of DVT, but we did not replicate this finding in our sequencing data (**chapter 2**). Based on recent genetic studies on other common complex traits, the genetic architecture of VT is most likely characterized by a polygenic signature of common and rare variants conferring modest-to-small effects on disease risk.[28,30,34,35] The causal variants map most likely to both coding and noncoding sequence across the genome.[36-38] This has several important methodological consequences for studies aiming to identify novel risk factors for (recurrent) VT, which are discussed below.

First, sample size is of utmost importance when conducting large genetic association studies due to the small effect sizes that need to be detected with precision and the large number of statistical tests performed thereby requiring a stringent threshold to attain statistical significance. The number of tests conducted depends on the approach taken: a few to 500 (tagging) variants in a candidate gene study compared with several millions in a GWAS study imputed to a dense reference panel. As evidenced from the two largest GWAS studies on VT so far, the effects conveyed by low-frequency and common variants (MAF ≥ 1%) on VT risk are generally small, with odds ratios ranging between 1.1 and 1.8.[25,26] Exceptions are FV Leiden (MAF 3.0% in Europeans) and PT

G20210A (MAF 1.0% in Europeans) which are associated with a 3.5-fold and 2-fold increased risk of a first VT per copy of the minor allele, respectively.[25,39,40] As part of the INVENT consortium, we have previously meta-analyzed GWAS data from 7,507 VT patients and 52,632 controls, resulting in sufficient statistical power to detect odds ratios of >1.2 for common, but not low-frequency, variants.[25] We expect that with increasing sample sizes more genetic risk factors for VT will be identified, as has been the case for other common complex traits such as height and obesity.[41,42] Recent estimates suggest that for common complex traits sample sizes ranging from a few hundred thousand to multiple millions are required to identify variants that explain most heritability found in GWASs.[35,41] Sequencing studies focusing on rare variants across the exome or the entire genome require an even larger sample size to discover novel risk variants. Achieving these large sample sizes is a major bottleneck, as venous thrombosis occurs in only 1-2 per 1000 persons per year.[43,44] As such, several analyses conducted for this thesis were underpowered, and we may have missed relevant associations with venous thrombosis. To maximize statistical power, alternative strategies can be employed, such as we did in the sequencing study (**chapter 2**), where we specifically focused on DVT risk instead of DVT or PE in order to study a homogenous phenotype. In addition, we excluded individuals with major clinical risk factors for VT in order to study a population which is more likely to carry genetic risk variants. Further strategies to maximize power include studying population isolates, conducting transethnic analyses, or by using advanced statistical models such as Bayesian models that do not require Bonferroni correction for multiple testing.[45-47] Of note, sample size is not just critical for discovery analyses, but also for replication analyses in which the top candidates per locus, usually the variants with the lowest P-values, are tested in an independent sample. This P-value driven selection can lead to the so-called 'winner's curse', which is a bias away from the null similar to regression-to-the-mean.[48,49] Genetic variants passing the threshold for statistical significance are more likely to have overestimated effect sizes in the discovery sample due to chance. Therefore, if possible, replication analyses should be powered to detect effect sizes smaller than those reported in the initial discovery analysis.

Second, the genetic ancestry of the study population should be considered before and during genetic analyses. Genetic association studies in admixed populations may be hampered by confounding due to population structure.[50] As both allele frequencies and the incidence of VT vary according to genetic ancestry, the independence assumption is violated in studies of admixed populations resulting in potentially spurious associations.

To avoid this, studies should appropriately account for population structure. Therefore, most of our analyses, were limited to individuals of self-reported European origin. In the GWAS discussed in **chapter 3**, we used principal component analysis[51] to control for population structure and calculated the genomic inflation factor[52] to assess the presence of any remaining population substructure. Recent studies suggest that confounding by population structure may be more of a concern when studying rare variants, as these may show different stratification patterns compared with common variants due to selection pressure, founder effects, and as these are more likely to have arisen recently.[53,54] Of note, the downside of studying genetic risk variants in an ethnically homogenous population is that the results are only generalizable to that population. For example, the genetic risk score in **chapter 6** was constructed and validated in individuals of European origin and, therefore, performs less well in individuals of non-European ancestry as the included variants are less informative in non-European populations. For example, FV Leiden reaches a MAF of 3% in Europeans but is virtually absent in Africans and East Asians, thereby limiting its discriminative power in those populations.[55] While it has been shown that our genetic risk score has limited predictive value in African Americans,[56] another study reported some generalizability of VT risk variants identified in Europeans to other ancestries in a study on chronic venous disease.[57] As few and only small studies on genetic risk factors for VT have been performed in populations of non-European ancestry,[58-62] it is currently difficult to assess the generalizability of our findings.

Last but not least, linkage disequilibrium (LD), the non-random association of alleles at closely linked loci in a population, requires attention when conducting and interpreting genetic analyses. Specifically, LD may affect genetic association studies and Mendelian randomization studies, as associated variants may not be causal variants, but rather be in linkage with these. LD, amongst others determined by recombination rate and demographic aspects of a population, may extend for several megabases along a chromosome while sometimes interspersed with blocks of no or little LD.[63-65] As a result, causal variants may even map to different genes than the associated variants, complicating the interpretation of an association signal. Of the VT risk loci, LD blocks spanning multiple genes are, for example, observed at the *F11* locus and *FGA*-*FGG* locus.[66-68] We were therefore unable to disentangle the association between DVT and genetic variants *FGA*-rs6050 and *FGG*-rs2066865 (**chapter 2**), which have both previously been associated with VT risk[68,69] and are almost in complete LD (r^2 0.90 in

Europeans). In addition, a GWAS association signal at 11p11.2 has previously almost been misinterpreted as a novel risk locus for VT before it was tracked down to PT G20210A using LD and haplotype analyses.[70] Since LD patterns differ between genetic ancestries,[64,71] transethnic analyses could aid fine-mapping at regions with strong LD in Europeans.[72] Of note, even in regions with considerable LD, it is possible that multiple conditionally independent associations exist, either because there are multiple causal variants or the associated variants are all in moderate LD with the unmeasured causal variant(s). We and others have reported evidence for secondary associations at several of the known VT risk loci, including *ABO*, *CADM1*, *F2*, *F5,* and the *F11* locus (**chapters 2 and 5**).[4,25,58,66,67,73,74] Enlarging the sample size and extension to non-European populations will help to unravel the genetic structure at these loci.

Biological interpretation

Most of the established genetic risk factors for VT can be linked to the hemostatic system.[1,2] For some risk loci, the causal variant and the underlying biological mechanism have largely been elucidated. For example, a missense variant FV Leiden leads to loss of a cleavage site for activated protein C (APC), resulting in both APC resistance and decreased degradation of activated FVIII by APC and protein S.[39,75] PT G20210A results in increased PT plasma levels due to differential post-transcriptional regulation of PT mRNA,[40,76] whereas the *FGG*-haplotype containing rs2066865 yields lower levels of the γ'-fibrinogen and reduction of the γ'/γ ratio.[68] In addition, clearance of vWF is affected by the presence of A and B antigens of ABO on the surface of vWF.[77] The biological interpretation of other VT risk loci is more complex. VT risk variants in *F11* and *KNG1* are associated with increased FXI plasma and/or activity levels and with prolonged activated partial thromboplastin time.[25,66,67,78,79] However, it is suggested that their association with venous thrombosis cannot be completely explained by their effect on FXI levels.[79,80] Near *F11,* and part of the same LD block, lie *KLKB1* and *CYP4V2*, encoding prekallikrein and a cytochrome P450 family member, respectively. Several studies (including our sequencing study in **chapter 2**) have reported multiple conditionally independent associations between VT and variants in *KLKB1*, *CYP4V2*, and *F11*[25,66,67], but the exact causal mechanism has not been elucidated due to the extensive LD at this locus. Data from the Genotype-Tissue Expression Project[81] are also inconclusive: *F11*-rs2036914 is, for example, an expression quantitative trait locus (eQTL) for *F11* in lung tissue, whereas *F11*-rs1593 is an eQTL for *KLKB1* and *CYP4V2,* but not *F11*, in multiple tissues.

Furthermore, the link to venous thrombosis is unclear for the recently identified GWAS loci near *TSPAN15* and *SLC44A2*, which showed no evidence of an association with any of 25 hemostasis-related biomarkers.[25] It should be noted that the causal variant at these loci may also target a different gene, as many GWAS loci associated with common complex traits have shown not to impact the most nearby gene.[82,83] GWASs typically identify associations in noncoding sequence, which cannot be explained by linkage to coding variants, and are thought to impact a complex trait by affecting gene regulation, both transcriptionally and post-transcriptionally.[27,36-38,83] In order to elucidate the functional impact of such variants, integration with multiple genomics data, such as generated by ENCODE[84] and GTEx[81], is necessary. For example, colocalization analyses of GWAS hits with overlapping eQTL associations in relevant tissues can be used to pinpoint plausible causal variants and genes.[85,86] Further integration with methylation and epigenomic annotation data can help to dissect potential regulatory mechanisms, whereas chromatin interaction methods can detect long-range chromosomal interactions between variants in potential enhancers and their target genes.[87-89] These methods should also be applied to identify the causal variant and gene for the intergenic locus at 18q22.1, which was associated with recurrent VT (**chapter 3**). In addition, leveraging from data on endophenotypes, such as plasma coagulation factor levels, or metabolomics can help to dissect the biological link between the identified variants and the pathophysiology of VT.

Our lack of understanding of the biological underpinnings of GWAS loci also hampers the clinical translation of these genetic risk factors. Much effort is currently spent to increase our understanding of the role of regulatory variation in the genome. As this research field is evolving fast, with new methods and data becoming available on a regular basis, we expect that the biological mechanism underlying GWAS variants and other VT risk variants can be unraveled in the near future.

Clinical relevance

The ultimate goal of genetic association studies is to bring the genetic discoveries to the clinic, assuming that a better understanding of the biology underlying a disease leads to better treatments and preventive strategies. Specifically, elucidating risk genes and pathways may provide novel drug targets, for example, those that reduce thrombosis risk without (substantially) increasing the bleeding risk. Although the effect sizes of individual risk variants are small, their effect on molecular phenotypes

and the resulting drug effects can be large. A well-known example is the field of pharmacogenetics, which investigates genetic variation in metabolic pathways affecting individual responses to drugs. Variation in the vitamin K epoxide reductase (*VKORC1*) and hepatic drug-metabolizing enzyme cytochrome P450 2C9 (*CYP2C9*) genes largely determine the dose variability of coumarin anticoagulants.[90,91] As a result, patients taking these anticoagulants to prevent or treat thrombotic events have, depending on their genotypes, an increased risk of major bleeding due to over-anticoagulation. So far, several trials have investigated the use of genotype-guided dosing to reduce the number of adverse events during anticoagulant treatment, albeit with inconsistent results.[92-95] Besides guiding therapy, genetic variation may be informative in personalized risk prediction, i.e. identifying those who are at increased risk of developing VT and those who are not. In **chapter 6**, we showed that a genetic risk score of five well-known VT risk variants improved risk stratification in the general population and in clinically relevant subgroups. Our genetic risk score has been validated and extended in other studies of individuals of European ancestry, but showed limited discriminative power in African Americans.[56, 96-99] Identification of additional genetic variants, especially variants that increase VT risk in individuals of non-European ancestry, may further improve the discriminative power of such genetic risk scores. As the costs of genotyping continue to drop, the implementation of genetic risk factors into clinical prediction models may also become cost-effective. This may be most relevant for recurrence risk, as patients with a recurrent VT currently receive lifelong treatment with anticoagulants, which are associated with an increased risk of bleeding.

Two other clinically relevant applications of genetic findings are Mendelian randomization studies and studies focusing on the genetic correlation between traits. Specifically, GWAS results have shown that the same genetic variants can be associated with multiple traits, suggesting that some of the underlying causal mechanisms are shared.[100,101] This pleiotropic nature can also be exploited to quantify the genetic overlap between traits and diseases using methods such as cross-trait LD score regression.[100] As large-scale GWAS summary statistics for VT are not publicly available, a systematic analysis of genetic correlation between VT and other traits has not (yet) been published. A first study by Klarin *et al.*, based on a genetic risk score consisting of 10 VT risk variants, showed a statistically significant genetic overlap between VT and coronary artery disease risk, but not with 37 other disorders tested in data from the UK Biobank.[4] MR studies, on the other hand, can aid in unravelling the causal relationship between

clinical factors and VT risk (as explained in **chapter 7**). So far, MR studies on VT have shown that obesity and height, but not lipoprotein(a) and YKL-40, are causal risk factors for VT.[4,102-105] As more genetic variants are being identified and the analytical methods are being improved, we expect that both MR and genetic correlation analyses will become standard tools in genetic studies on VT and other common complex traits, ultimately advancing personalized medicine.

REFERENCES

1. Rosendaal FR, Reitsma PH. Genetics of venous thrombosis. *J Thromb Haemost*. 2009; 7 Suppl 1:301–304.
2. Trégouët DA, Morange PE. What is currently known about the genetics of venous thromboembolism at the dawn of next generation sequencing technologies. *Br J Haematol*. 2018;180(3):335-45.
3. Germain M, Saut N, Greliche N, Dina C, Lambert JC, Perret C, et al. Genetics of venous thrombosis: insights from a new genome wide association study. *PLoS One*. 2011;6(9):e25581.
4. Klarin D, Emdin CA, Natarajan P, Conrad MF; INVENT Consortium, Kathiresan S. Genetic Analysis of Venous Thromboembolism in UK Biobank Identifies the ZFPM2 Locus and Implicates Obesity as a Causal Risk Factor. *Circ Cardiovasc Genet*. 2017;10(2):e001643.
5. Souto J, Almasy L, Borrell M, Blanco-Vaca F, Mateo J, Soria J, Coll I, Felices R, Stone W, Fontcuberta J, Blangero J. Genetic susceptibility to thrombosis and its relationship to physiological risk factors: the GAIT study. Genetic Analysis of Idiopathic Thrombophilia. *Am J Hum Genet*. 2000;67:1452–9.
6. Heit J, Phelps M, Ward S, Slusser J, Petterson T, De Andrade M. Familial segregation of venous thromboembolism. *J Thromb Haemost*. 2004;2:731–6.
7. Ho WK, Hankey GJ, Quinlan DJ, Eikelboom JW. Risk of recurrent venous thromboembolism in patients with common thrombophilia: a systematic review. *Arch Intern Med*. 2006;166(7):729-36.
8. Marchiori A, Mosena L, Prins MH, Prandoni P. The risk of recurrent venous thromboembolism among heterozygous carriers of factor V Leiden or prothrombin G20210A mutation. A systematic review of prospective studies. *Haematologica*. 2007;92(8):1107-14.
9. van Hylckama Vlieg A, Flinterman LE, Bare LA, Cannegieter SC, Reitsma PH, Arellano AR, et al. Genetic variations associated with recurrent venous thrombosis. *Circ Cardiovasc Genet*. 2014;7(6):806-13.
10. Bruzelius M, Ljungqvist M, Bottai M, Bergendal A, Strawbridge RJ, Holmström M, et al. F11 is associated with recurrent VTE in women. A prospective cohort study. *Thromb Haemost*. 2016;115(2):406-14.
11. Kyrle PA, Minar E, Bialonczyk C, Hirschl M, Weltermann A, Eichinger S. The risk of recurrent venous thromboembolism in men and women. *N Engl J Med.* 2004;350:2558–63.
12. Christiansen SC, Cannegieter SC, Koster T, Vandenbroucke JP, Rosendaal FR. Thrombophilia, clinical factors, and recurrent venous thrombotic events. *JAMA*. 2005;293:2352-61.
13. McRae S, Tran H, Schulman S, Ginsberg J, Kearon C. Effect of patient's sex on risk of recurrent venous thromboembolism: a meta-analysis. *Lancet.* 2006;368:371-8.

14. Rodger MA, Kahn SR, Wells PS, Anderson DA, Chagnon I, Le Gal G, Solymoss S, Crowther M, Perrier A, White R, Vickars L, Ramsay T, Betancourt MT, Kovacs MJ. Identifying unprovoked thromboembolism patients at low risk for recurrence who can discontinue anticoagulant therapy. *CMAJ.* 2008;179:417-426.
15. Douketis J, Tosetto A, Marcucci M, Baglin T, Cosmi B, Cushman M, Kyrle P, Poli D, Tait RC, Iorio A. Risk of recurrence after venous thromboembolism in men and women: patient level meta-analysis. *BMJ.* 2011;342:d813.
16. Roach RE, Lijfering WM, Rosendaal FR, Cannegieter SC, le Cessie S. Sex difference in risk of second but not of first venous thrombosis: paradox explained. *Circulation.* 2014;129:51-6.
17. Charchar FJ, Bloomer LD, Barnes TA, Cowley MJ, Nelson CP, Wang Y, Denniff M, Debiec R, Christofidou P, Nankervis S, Dominiczak AF, Bani-Mustafa A, Balmforth AJ, Hall AS, Erdmann J, Cambien F, Deloukas P, Hengstenberg C, Packard C, Schunkert H, et al. Inheritance of coronary artery disease in men: an analysis of the role of the Y chromosome. *Lancet.* 2012;379:915-22.
18. Hasstedt SJ, Bezemer ID, Callas PW, Vossen CY, Trotman W, Hebbel RP, et al. Cell adhesion molecule 1: a novel risk factor for venous thrombosis. *Blood*. 2009;114:3084–91.
19. Tatsumi K, Taatjes DJ, Wadsworth MP, Bouchard BA, Bovill EG. Cell adhesion molecule 1 (CADM1) is ubiquitously present in the endothelium and smooth muscle cells of the human macro- and micro-vasculature. *Histochem Cell Biol*. 2012;138:815–20.
20. Hemani G, Bowden J, Davey Smith G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Hum Mol Genet*. 2018;27(R2):R195-R208.
21. Jordan FL, Nandorff A. The familial tendency in thrombo-embolic disease. *Acta Med Scand.* 1956;156(4):267-75.
22. Collins FS, Morgan M, Patrinos A. The Human Genome Project: Lessons from Large-Scale Biology. *Science*. 2003;300(5617):286-90.
23. Lander ES. The new genomics: global views of biology. *Science*. 1996;274(5287):536–9.
24. Reich DE, Lander ES. On the allelic spectrum of human disease. *Trends Genet*. 2001;17(9):502–10.
25. Germain M, Chasman DI, de Haan H, Tang W, Lindström S, Weng LC, et al. Meta-analysis of 65,734 individuals identifies TSPAN15 and SLC44A2 as two susceptibility loci for venous thromboembolism. *Am J Hum Genet*. 2015;96(4):532-42.
26. Hinds DA, Buil A, Ziemek D, Martinez-Perez A, Malik R, Folkersen L, Germain M, Mälarstig A, Brown A, Soria JM, Dichgans M, Bing N, Franco-Cereceda A, Souto JC, Dermitzakis ET, Hamsten A, Worrall BB, Tung JY; METASTROKE Consortium, INVENT Consortium, Sabater-Lleal M. Genome-wide association analysis of self-reported events in 6135 individuals and 252 827 controls identifies 8 loci associated with thrombosis. *Hum Mol Genet*. 2016;25(9):1867-74.

27. Hindorff LA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci*. 2009;106:9362–9367.
28. 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(7265):747-53.
29. Pritchard JK. Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet*. 2001;69(1):124-37.
30. Park JH, Gail MH, Weinberg CR, Carroll RJ, Chung CC, Wang Z, et al. Distribution of allele frequencies and effect sizes and their interrelationships for common genetic susceptibility variants. *Proc Natl Acad Sci U S A*. 2011;108:18026–31.
31. Purcell SM, Moran JL, Fromer M, Ruderfer D, Solovieff N, Roussos P, et al. A polygenic burden of rare disruptive mutations in schizophrenia. *Nature*. 2014;506(7487):185-90.
32. UK10K Consortium, Walter K, Min JL, Huang J, Crooks L, Memari Y, et al. The UK10K project identifies rare variants in health and disease. *Nature*. 2015;526:82–90.
33. Lotta LA, Tuana G, Yu J, Martinelli I, Wang M, Yu F, Passamonti SM, Pappalardo E, Valsecchi C, Scherer SE, Hale W 4th, Muzny DM, Randi G, Rosendaal FR, Gibbs RA, Peyvandi F. Next-generation sequencing study finds an excess of rare, coding single-nucleotide variants of ADAMTS13 in patients with deep vein thrombosis. *J Thromb Haemost*. 2013;11:1228-39.
34. Marouli E, Graff M, Medina-Gomez C, Lo KS, Wood AR, Kjaer TR, et al. Rare and low-frequency coding variants alter human adult height. *Nature*. 2017;542(7640):186-190.
35. Zhang Y, Qi G, Park JH, Chatterjee N. Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits. *Nat Genet.* 2018;50(9):1318-1326.
36. Pickrell JK. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am J Hum Genet*. 2014;94(4):559-73.
37. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorff L, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res*. 2014;42(D1):D1001–D1006.
38. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh PR, Anttila V, Xu H, Zang C, Farh K, ReproGen Consortium. Schizophrenia Working Group of the Psychiatric Genomics Consortium. RACI Consortium Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet*. 2015;47(11):1228–35.
39. 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(6475):64-7.

40. 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(10):3698-703.
41. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet*. 2010;42:565–569.
42. Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, Frayling TM, Hirschhorn J, Yang J, Visscher PM1; GIANT Consortium. Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Hum Mol Genet*. 2018;27(20):3641-3649.
43. Anderson FA Jr, Wheeler HB, Goldberg RJ, Hosmer DW, Patwardhan NA, Jovanovic B, et al. A population-based perspective of the hospital incidence and case-fatality rates of deep vein thrombosis and pulmonary embolism. The Worcester DVT Study. *Arch Intern Med*. 1991;151(5):933-8.
44. Naess IA, Christiansen SC, Romundstad P, Cannegieter SC, Rosendaal FR, Hammerstrøm J. Incidence and mortality of venous thrombosis: a population-based study. *J Thromb Haemost*. 2007;5(4):692-9.
45. Mägi R, Horikoshi M, Sofer T, Mahajan A, Kitajima H, Franceschini N, McCarthy MI; COGENT-Kidney Consortium, T2D-GENES Consortium, Morris AP. Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. *Hum Mol Genet*. 2017; 26(18): 3639–3650.
46. Kenny EE, Kim M, Gusev A, Lowe JK, Salit J, Smith JG, Kovvali S, Kang HM, Newton-Cheh C, Daly MJ, Stoffel M, Altshuler DM, Friedman JM, Eskin E, Breslow JL, Pe'er I. Increased power of mixed models facilitates association mapping of 10 loci for metabolic traits in an isolated population. *Hum Mol Genet*. 2011;20(4):827-39.
47. Loh PR, Tucker G, Bulik-Sullivan BK, Vilhjálmsson BJ, Finucane HK, Salem RM, Chasman DI, Ridker PM, Neale BM, Berger B, Patterson N, Price AL. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet*. 2015;47(3):284-90.
48. Capen EC, Clapp RV, Campbell WM. Competitive bidding in high-risk situations. *J Petrol Technol*. 1971;23:641–653.
49. Göring HHH, Terwilliger JD, Blangero J. Large upward bias in estimation of locus-specific effects from genomewide scans. *Am J Hum Genet*. 2001;69:1357–1369.
50. Marchini J, Cardon LR, Phillips MS, Donnelly P. The effects of human population structure on large genetic association studies. *Nat Genet*. 2004;36(5):512-7.
51. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006;38(8):904-9.

52. Devlin B, Roeder K. Genomic control for association studies. *Biometrics*. 1999;55(4):997-1004.
53. Babron MC, de Tayrac M, Rutledge DN, Zeggini E, Génin E. Rare and low frequency variant stratification in the UK population: description and impact on association tests. *PLoS One*. 2012;7(10):e46519.
54. Mathieson I, McVean G. Differential confounding of rare and common variants in spatially structured populations. *Nat Genet*. 2012;44(3):243-6.
55. Rees DC, Cox M, Clegg JB. World distribution of factor V Leiden. *Lancet*. 1995;346(8983):1133-4.
56. Folsom AR, Tang W, Weng LC, Roetker NS, Cushman M, Basu S, Pankow JS. Replication of a genetic risk score for venous thromboembolism in whites but not in African Americans. *J Thromb Haemost*. 2016;14(1):83-8.
57. Wassel CL, Rasmussen-Torvik LJ, Callas PW, Denenberg JO, Durda JP, Reiner AP, Smith NL, Allison MA, Rosendaal FR, Criqui MH, Cushman M. A genetic risk score comprising known venous thromboembolism loci is associated with chronic venous disease in a multi-ethnic cohort. *Thromb Res*. 2015;136(5):966-73.
58. Uitte de Willige S, Pyle ME, Vos HL, de Visser MC, Lally C, Dowling NF, Hooper WC, Bertina RM, Austin H. Fibrinogen gamma gene 3'-end polymorphisms and risk of venous thromboembolism in the African-American and Caucasian population. *Thromb Haemost*. 2009;101(6):1078-84.
59. Tang L, Lu X, Yu JM, Wang QY, Yang R, Guo T, Mei H, Hu Y. PROC c.574_576del polymorphism: A common genetic risk factor for venous thrombosis in the Chinese population. *J Thromb Haemost*. 2012;10(10):2019–2026.
60. Tang L, Wang HF, Lu X, Jian XR, Jin B, Zheng H, Li YQ, Wang QY, Wu TC, Guo H, Liu H, Guo T, Yu JM, Yang R, Yang Y, Hu Y. Common genetic risk factors for venous thrombosis in the Chinese population. *Am J Hum Genet*. 2013;92(2):177-87.
61. Hernandez W, Gamazon ER, Smithberger E, O'Brien TJ, Harralson AF, Tuck M, Barbour A, Kittles RA, Cavallari LH, Perera MA. Novel genetic predictors of venous thromboembolism risk in African Americans. *Blood*. 2016;127(15):1923-9.
62. Heit JA, Armasu SM, McCauley BM, Kullo IJ, Sicotte H, Pathak J, Chute CG, Gottesman O, Bottinger EP, Denny JC, Roden DM, Li R, Ritchie MD, de Andrade M. Identification of unique venous thromboembolism-susceptibility variants in African-Americans. *Thromb Haemost*. 2017;117(4):758-768.
63. Peterson AC, Rienzo AD, Lehesjokl A-E, Chapelle Adl, Slatkin M, et al. The distribution of linkage disequilibrium over anonymous genome regions. *Hum Mol Genet*. 1995;4:887–894.
64. Ardlie KG, Kruglyak L, Seielstad M. Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet*. 2002;3(4):299-309.

65. Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES. Linkage disequilibrium in the human genome. *Nature*. 2001;411(6834):199-204.
66. Bezemer ID, Bare LA, Doggen CJ, Arellano AR, Tong C, Rowland CM, Catanese J, Young BA, Reitsma PH, Devlin JJ, Rosendaal FR. Gene variants associated with deep vein thrombosis. *JAMA*. 2008;299(11):1306-14.
67. Li Y, Bezemer ID, Rowland CM, Tong CH, Arellano AR, Catanese JJ, Devlin JJ, Reitsma PH, Bare LA, Rosendaal FR. Genetic variants associated with deep vein thrombosis: the F11 locus. *J Thromb Haemost*. 2009;7(11):1802-8.
68. Uitte de Willige S, de Visser MC, Houwing-Duistermaat JJ, Rosendaal FR, Vos HL, Bertina RM. Genetic variation in the fibrinogen gamma gene increases the risk for deep venous thrombosis by reducing plasma fibrinogen gamma' levels. *Blood*. 2005;106(13):4176-83.
69. Lotta LA, Wang M, Yu J, Martinelli I, Yu F, Passamonti SM, Consonni D, Pappalardo E, Menegatti M, Scherer SE, Lewis LL, Akbar H, Wu Y, Bainbridge MN, Muzny DM, Mannucci PM, Gibbs RA, Peyvandi F. Identification of genetic risk variants for deep vein thrombosis by multiplexed next-generation sequencing of 186 hemostatic/pro-inflammatory genes. *BMC Med Genomics*. 2012;5:7.
70. Germain M, Saut N, Oudot-Mellakh T, Letenneur L, Dupuy AM, Bertrand M, Alessi MC, Lambert JC, Zelenika D, Emmerich J, Tiret L, Cambien F, Lathrop M, Amouyel P, Morange PE, Trégouët DA. Caution in interpreting results from imputation analysis when linkage disequilibrium extends over a large distance: a case study on venous thrombosis. *PLoS One*. 2012;7(6):e38538.
71. Shifman S, Kuypers J, Kokoris M, Yakir B, Darvasi A. Linkage disequilibrium patterns of the human genome across populations. *Hum Mol Genet*. 2003;12(7):771-6.
72. Zaitlen N, Paşaniuc B, Gur T, Ziv E, Halperin E. Leveraging genetic variability across populations for the identification of causal variants. *Am J Hum Genet*. 2010;86(1):23-33.
73. Smith NL, Hindorff LA, Heckbert SR, Lemaitre RN, Marciante KD, Rice K, Lumley T, Bis JC, Wiggins KL, Rosendaal FR, Psaty BM. Association of genetic variations with nonfatal venous thrombosis in postmenopausal women. *JAMA*. 2007;297(5):489-98.
74. Heit JA, Armasu SM, Asmann YW, Cunningham JM, Matsumoto ME, Petterson TM, De Andrade M. A genome-wide association study of venous thromboembolism identifies risk variants in chromosomes 1q24.2 and 9q. *J Thromb Haemost*. 2012;10(8):1521-31.
75. Nicolaes GA, Dahlbäck B. Factor V and thrombotic disease: description of a janus-faced protein. *Arterioscler Thromb Vasc Biol*. 2002;22(4):530-8.
76. Pollak ES, Lam HS, Russell JE. The G20210A mutation does not affect the stability of prothrombin mRNA in vivo. *Blood*. 2002;100(1):359-62.

77. Gallinaro L, Cattini MG, Sztukowska M, Padrini R, Sartorello F, Pontara E, Bertomoro A, Daidone V, Pagnan A, Casonato A. A shorter von Willebrand factor survival in O blood group subjects explains how ABO determinants influence plasma von Willebrand factor. *Blood*. 2008;111(7):3540-5.
78. Sabater-Lleal M, Martinez-Perez A, Buil A, Folkersen L, Souto JC, Bruzelius M, Borrell M, Odeberg J, Silveira A, Eriksson P, Almasy L, Hamsten A and Soria JM. A genome-wide association study identifies KNG1 as a genetic determinant of plasma factor XI Level and activated partial thromboplastin time. *Arterioscler Thromb Vasc Biol*. 2012;32:2008-16.
79. Sennblad B, Basu S, Mazur J, Suchon P, Martinez-Perez A, van Hylckama Vlieg A, Truong V, Li Y, Gadin JR, Tang W, Grossman V, de Haan HG, Handin N, Silveira A, Souto JC, Franco-Cereceda A, Morange PE, Gagnon F, Soria JM, Eriksson P, Hamsten A, Maegdefessel L, Rosendaal FR, Wild P, Folsom AR, Tregouet DA and Sabater-Lleal M. Genome-wide association study with additional genetic and post-transcriptional analyses reveals novel regulators of plasma factor XI levels. *Hum Mol Genet*. 2017.
80. Rohmann JL, de Haan HG, Algra A, Vossen CY, Rosendaal FR, Siegerink B. Genetic determinants of activity and antigen levels of contact system factors. *J Thromb Haemost*. 2019;17(1):157-168.
81. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*. 2015;348(6235):648-60.
82. Brodie A, Azaria JR, Ofran Y: How far from the SNP may the causative genes be? *Nucleic Acids Res*. 2016;44:6046–54.
83. Maurano MT, et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science*. 2012;337:1190–1195.
84. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489:57–74.
85. Hormozdiari F, van de Bunt M, Segrè AV, Li X, Joo JWJ, Bilow M, Sul JH, Sankararaman S, Pasaniuc B, Eskin E. Colocalization of GWAS and eQTL Signals Detects Target Genes. *Am J Hum Genet*. 2016;99(6):1245-1260.
86. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BW, Jansen R, de Geus EJ, Boomsma DI, Wright FA. Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet*. 2016;48:245–52.
87. Hannon E, Weedon M, Bray N, O'Donovan M, Mill J. Pleiotropic Effects of Trait-Associated Genetic Variation on DNA Methylation: Utility for Refining GWAS Loci. *Am J Hum Genet*. 2017;100(6):954-959.

88. Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, Kellis M, Marra MA, Beaudet AL, Ecker JR, Farnham PJ, Hirst M, Lander ES, Mikkelsen TS, Thomson JA. The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol*. 2010;28(10):1045-8.
89. Schmitt AD, Hu M, Jung I, Xu Z, Qiu Y, Tan CL, Li Y, Lin S, Lin Y, Barr CL, Ren B. A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Rep.* 2016;17:2042–59.
90. Bodin L, Verstuyft C, Tregouet DA, Robert A, Dubert L, Funck-Brentano C, Jaillon P, Beaune P, Laurent-Puig P, Becquemont L, Loriot MA. Cytochrome P450 2C9 (CYP2C9) and vitamin K epoxide reductase (VKORC1) genotypes as determinants of acenocoumarol sensitivity. *Blood* 2005;106:135-140.
91. Wadelius M, Chen LY, Lindh JD, Eriksson N, Ghori MJ, Bumpstead S, Holm L, McGinnis R, Rane A, Deloukas P. The largest prospective warfarin-treated cohort supports genetic forecasting. *Blood*. 2009;113:784-792.
92. Verhoef TI, Ragia G, de Boer A, Barallon R, Kolovou G, Kolovou V, Konstantinides S, Le Cessie S, Maltezos E, van der Meer FJ, Redekop WK, Remkes M, Rosendaal FR, van Schie RM, Tavridou A, Tziakas D, Wadelius M, Manolopoulos VG, Maitland-van der Zee AH; EU-PACT Group. A randomized trial of genotype-guided dosing of acenocoumarol and phenprocoumon. *N Engl J Med*. 2013;369(24):2304-12.
93. Pirmohamed M, Burnside G, Eriksson N, Jorgensen AL, Toh CH, Nicholson T, Kesteven P, Christersson C, Wahlstrom B, Stafberg C, Zhang JE, Leathart JB, Kohnke H, Maitland-van der Zee AH, Williamson PR, Daly AK, Avery P, Kamali F, Wadelius M; EU-PACT Group. A randomized trial of genotype-guided dosing of warfarin. *N Engl J Med.* 2013;369:2294–303.
94. Kimmel SE, French B, Kasner SE, Johnson JA, Anderson JL, Gage BF, Rosenberg YD, Eby CS, Madigan RA, McBane RB, Abdel-Rahman SZ, Stevens SM, Yale S, Mohler ER 3rd, Fang MC, Shah V, Horenstein RB, Limdi NA, Muldowney JA 3rd, Gujral J, *et al.* A pharmacogenetic versus a clinical algorithm for warfarin dosing. *N Engl J Med.* 2013;369:2283–93.
95. Gage BF, Bass AR, Lin H, Woller SC, Stevens SM, Al-Hammadi N, Li J, Rodríguez T Jr, Miller JP, McMillin GA, Pendleton RC, Jaffer AK, King CR, Whipple BD, Porche-Sorbet R, Napoli L, Merritt K, Thompson AM, Hyun G, Anderson JL, Hollomon W, Barrack RL, Nunley RM, Moskowitz G, Dávila-Román V, Eby CS. Effect of Genotype-Guided Warfarin Dosing on Clinical Events and Anticoagulation Control Among Patients Undergoing Hip or Knee Arthroplasty: The GIFT Randomized Clinical Trial. *JAMA*. 2017;318(12):1115-24.
96. van Hylckama Vlieg A, Flinterman LE, Bare LA, Cannegieter SC, Reitsma PH, Arellano AR, Tong CH, Devlin JJ, Rosendaal FR. Genetic variations associated with recurrent venous thrombosis. *Circ Cardiovasc Genet*. 2014;7(6):806-13.

97. Soria JM, Morange PE, Vila J, Souto JC, Moyano M, Trégouët DA, Mateo J, Saut N, Salas E, Elosua R. Multilocus genetic risk scores for venous thromboembolism risk assessment. *J Am Heart Assoc*. 2014;3(5):e001060.
98. Bruzelius M, Bottai M, Sabater-Lleal M, Strawbridge RJ, Bergendal A, Silveira A, Sundström A, Kieler H, Hamsten A, Odeberg J. Predicting venous thrombosis in women using a combination of genetic markers and clinical risk factors. *J Thromb Haemost*. 2015;13(2):219-27.
99. McDaid A, Logette E, Buchillier V, Muriset M, Suchon P, Pache TD, Tanackovic G, Kutalik Z, Michaud J. Risk prediction of developing venous thrombosis in combined oral contraceptive users. *PLoS One*. 2017;12(7):e0182041.
100. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, Duncan L, Perry JR, Patterson N, Robinson EB, ReproGen Consortium. Psychiatric Genomics Consortium. Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3 An atlas of genetic correlations across human diseases and traits. *Nat Genet*. 2015;47:1236–41.
101. Pickrell JK, Berisa T, Liu JZ, Ségurel L, Tung JY, Hinds DA. Detection and interpretation of shared genetic influences on 42 human traits. *Nat Genet*. 2016;48:709–717.
102. Kamstrup PR, Tybjærg-Hansen A, Nordestgaard BG. Genetic evidence that lipoprotein(a) associates with atherosclerotic stenosis rather than venous thrombosis. *Arterioscler Thromb Vasc Biol*. 2012;32(7):1732-41.
103. Kjaergaard AD, Johansen JS, Bojesen SE, Nordestgaard BG. Observationally and Genetically High YKL-40 and Risk of Venous Thromboembolism in the General Population: Cohort and Mendelian Randomization Studies. *Arterioscler Thromb Vasc Biol*. 2016;36(5):1030-6.
104. Lindström S, Germain M, Crous-Bou M, Smith EN, Morange PE, van Hylckama Vlieg A, de Haan HG, Chasman D, Ridker P, Brody J, de Andrade M, Heit JA, Tang W, DeVivo I, Grodstein F, Smith NL, Tregouet D, Kabrhel C; INVENT Consortium. Assessing the causal relationship between obesity and venous thromboembolism through a Mendelian Randomization study. *Hum Genet*. 2017;136(7):897-902.
105. Roetker NS, Armasu SM, Pankow JS, Lutsey PL, Tang W, Rosenberg MA, Palmer TM, MacLehose RF, Heckbert SR, Cushman M, de Andrade M, Folsom AR. Taller height as a risk factor for venous thromboembolism: a Mendelian randomization meta-analysis. *J Thromb Haemost*. 2017;15(7):1334-43.