Received: 13 February 2024 | Revised: 4 April 2024 | Accepted: 5 April 2024

https://doi.org/10.1016/j.jtha.2024.04.004

REVIEW ARTICLE

Next-generation sequencing strategies in venous thromboembolism: in whom and for what purpose?

David-Alexandre Trégouët[1] | Pierre-Emmanuel Morange[2]

[1]University of Bordeaux, Institut National de la Santé et de la Recherche Médicale, Bordeaux Population Health Research Center, Unité Mixte de Recherche 1219, Bordeaux, France

[2]Cardiovascular and Nutrition Research Center (Centre de Recherche en CardioVasculaire et Nutrition), Institut National de la Santé et de la Recherche Médicale, Institut National de Recherche pour l'agriculture, l' Alimentation et l'Environnement, Aix-Marseille University, Marseille, France

Correspondence
David-Alexandre Trégouët, Institut National de la Santé et de la Recherche Médicale, Unité Mixte de Recherche 1219, Bordeaux Population Health Research Center, University of Bordeaux, 146 Rue Léo Saignat 11, 33076 Bordeaux, France.
Email: david-alexandre.tregouet@u-bordeaux.fr

Funding information
D.-A.T.'s and P.-E.M.'s venous thromboembolism genomics research projects were supported by the Institut National de la Santé et de la Recherche Médicale (INSERM) GenOmic variability in heaLth & Disease (GOLD) Cross-Cutting Program, the GENMED Laboratory of Excellence on Medical Genomics (ANR-10-LABX-0013), a research program managed by the National Research Agency (ANR) as part of the French Investment for the Future, and the French Investigation Network on Venous Thrombo-Embolism.

Abstract

This invited review follows the oral presentation "To Sequence or Not to Sequence, That Is Not the Question; But 'When, Who, Which and What For?' Is" given during the State of the Art session "Translational Genomics in Thrombosis: From OMICs to Clinics" of the International Society on Thrombosis and Haemostasis 2023 Congress. Emphasizing the power of next-generation sequencing technologies and the diverse strategies associated with DNA variant analysis, this review highlights the unresolved questions and challenges in their implementation both for the clinical diagnosis of venous thromboembolism and in translational research.

KEYWORDS
clinical diagnosis, next-generation sequencing, venous thromboembolism

Manuscript handled by: Kathleen Freson

Final decision: Kathleen Freson, 5 April 2024

1 | INTRODUCTION

Venous thromboembolism (VTE) is the third most common cause of cardiovascular death [1] and is associated with a strong public health burden [2]. Its incidence increases markedly with age, with a dramatic increase occurring at 60 years of age. Around 30% of patients experience a recurrence within 10 years, underscoring a high demand for therapeutic intervention to treat and prevent VTE and to identify patients at high risk [3]. Interestingly, it has been shown that the phenotype of recurrence mimics the phenotype of the initial episode of VTE: when initial VTE is an unprovoked pulmonary embolism (PE), recurrence occurs as unprovoked PE in about 80% of cases, fatal in 10%; when initial VTE is an unprovoked deep vein thrombosis (DVT), recurrence occurs as unprovoked DVT in 70% of cases, fatal in 2% [4,5]. Except for this clinical observation, no biomarker robustly predicting the recurrence of VTE is available. All these elements argue for a deep understanding of the etiologic mechanisms leading to VTE being a mandatory prerequisite before considering any therapeutic development and any personalized/precision prophylaxis.

VTE is highly heritable, and known common disease-associated variants explain less than 15% of the heritability of the disease [6–8], whose overall estimate is expected to be ~50% [9]. A significant portion of these common VTE variants, which are generally single nucleotide polymorphisms (SNPs), have been identified owing to the remarkably productive outcomes of extensive genome-wide association studies (GWASs). So far, about 150 loci have been observed to harbor common susceptibility alleles for VTE risk [6–8]. Identified SNPs are in general markers of functional variations yet to be identified and whose associated molecular mechanisms need to be fully dissected. About 70% of these VTE-associated variants are located in noncoding regions. It remains to be determined whether these variants are regulatory variants per se or whether they are tagging for functional ones. Besides, most identified VTE-associated SNPs have very weak genetic effects (allelic odds ratio [OR] from 1.03 to 1.5 for most variants) and low predictive ability to identify at-risk patients and do not yet lead to therapeutic development or improved prophylaxis; all these elements are currently limiting the clinical utility of these genetic findings.

Despite their undeniable success in uncovering new molecular mechanisms involved in various human diseases, GWASs have not, to date, had the "profound impact on the future of medicine" [10] initially hoped for. This could be partially explained by the fact that, until very recently, GWASs were exclusively conducted on common SNPs derived (and further imputed) from high-throughput genotyping arrays, with other genetic variations of more complex nature (copy number variations and structural variants, among others) being overlooked. This limitation is on the verge of being overcome with the development and practical implementation of next-generation sequencing (NGS) technologies, allowing the exploration at an unprecedented scale of millions of nucleotide sequences and offering finer resolution and exhaustive coverage of the genome for any subtle and rare genetic variations.

Overall, the cost of sequencing has significantly decreased in recent years. For instance, in 2013, the price for sequencing a standard genome at 30× depth using an Illumina instrument was around $3000. Now, in 2023, Illumina technology allows whole-genome sequencing (WGS) at approximately $250. The cost depends of course on the desired sequencing depth, and the new version of Illumina, NovaSeq 6000, will enable genome sequencing at $70 if the depth is reduced to just 4×. In parallel, other competitive sequencing technologies have emerged, such as the PacBio (Pacific Biosciences) instrument, enabling long-read genome sequencing (see below) at an average depth of 10× for ~$333.

Once we acknowledge that sequencing technology has become more affordable, but still expensive, and accessible to a larger audience, the question arises: To sequence! For what purpose? Two distinct yet interrelated contexts emerge—clinical diagnosis and research. The objectives share similarities. In the first scenario, the goal is to pinpoint the causal variant(s) responsible for a specific clinical phenotype, aiming to provide a molecular diagnosis and ideally propose a tailored therapeutic approach, including genetic counseling. In the second case, the emphasis is primarily placed on identifying new genes associated with a given phenotype, although ideally the ultimate goal is to identify the causal variants. The discovery of new genes then contributes to not only enriching the list of genes to be studied more thoroughly in a clinical context but also enhancing our knowledge on molecular pathways underlying disease etiology. Not to mention the possibility of discovering new therapeutic targets and engaging in drug repositioning.

But before embarking on a sequencing project, one must ask several questions.

2 | WHICH GENOMIC REGIONS TO SEQUENCE?

When it comes to NGS, there are generally 3 types of strategies distinguished: candidate gene sequencing of 1 or several genes, also called gene panel sequencing; sequencing of coding regions of the genome, and possibly flanking regions, often referred to as whole-exome sequencing (WES); and finally the complete sequencing of a genome (WGS). Obviously, the cost of sequencing will depend on the type of strategy, with WGS being the most expensive. Along with this cost, there is also complexity in variant analysis. As genomic regions are sequenced more extensively, the number of identified variants increases, making the functional characterization of certain variants of interest more challenging.

The question that arises when implementing a panel sequencing strategy, mainly in the clinical domain, is: which gene or genes do we want to study? When the clinical phenotype is coupled with abnormal levels of a well-established VTE-associated protein, it is sensible to sequence the associated structural gene. Success examples in identifying rare variants causing thrombotic events through such target gene sequencing include, but are not limited to, the *F5* Ala2086Asp mutation [11], the factor VIII Padua duplication variant [12], and the Arg338Leu mutation in the *F9* gene [13]. Unfortunately, it often happens that

patients for whom (panel) sequencing is proposed do not present specific biological phenotypes beyond clinical phenotypes. In that case, the adoption of sequencing for a panel of candidate genes can be considered. Candidate genes could for example be those known to harbor rare VTE pathogenic variants, and this list can be retrieved from ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/) or ClinGen (https://clinicalgenome.org/affiliation/50064/). This candidate gene approach has seen numerous successes, including the identification of the Arg596Leu [14], Arg596Gln [15], and Arg596Trp [16] mutations in the *F2* gene and the *F5* Trp1920Arg mutation [17]. A wider-panel gene sequencing approach can be deployed [18–21] by sequencing (all) genes from a specific pathway (eg, coagulation cascade) or even to all loci identified in GWASs for the disease under study. While the gene panel sequencing strategy has so far focused mainly on the analysis of coding or flanking regions, studies demonstrating the existence of functional variants in noncoding gene regions have become increasingly numerous. Beyond the historical examples of the well-characterized *F2* C20209T [22] and G20210A [23] and *FGG* C10034T [24] variants, both located in 3′UTR regions, more recent examples include the c.-39C>T mutation in the 5′UTR of the *PROS1* gene responsible for inherited protein S deficiency [25] and the c.151G>T variant in the 5′UTR of the *THBD* gene influencing thrombomodulin levels [26]. Although more challenging to characterize, deep intronic variants [27,28] should not be neglected and deserve much more attention.

When panel sequencing is inconclusive and/or when more substantial funds are available, a more agnostic strategy without any a priori about the underlying causal gene can then be employed with the ultimate goal of discovering novel disease genes. This could be a WES approach with the assumption that the underlying disease variant(s) is located in exonic regions or a more agnostic WGS approach that offers the possibility to explore all types of genetic variations over all genomic regions while avoiding some technical issues due to the capture of coding regions required in WES. In cases of rare diseases, it is currently estimated that the diagnostic rate of a WES strategy is ~30% and that the use of a completely agnostic approach like WGS provides so far only a slight additional efficiency, around 20% [29].

3 | WHICH SAMPLES TO SEQUENCE?

One of the first main justifications for NGS approaches is the identification of very rare variants not covered by genotyping technologies based on the study of SNPs with overall allelic frequencies greater than 1%. It is therefore natural that these NGS approaches find a particularly well-suited application in the analysis of extended pedigrees with a strong familial aggregation compatible with the Mendelian transmission of rare genetic mutations. In that case, the optimal design consists of sequencing, as much as possible, the most distant related affected individuals together with healthy individuals closely genetically related to affected relatives. Under the assumption of a single extremely rare variant following a dominant or recessive transmission, sequencing analysis in large pedigrees can be extremely fruitful, as demonstrated by numerous publications, particularly in the field of hemostasis. For example, let us mention the discovery of the *ETV6* [30] and *RASGRP2* [31] genes in familial platelet disorders. Full success stories that used WES/WGS strategies in VTE are very scarce. To our knowledge, only 1 novel VTE gene has been identified using WES/WGS in extended pedigrees. This is the *MAST2* gene where the Arg89Gln variant was found to cosegregate with inherited thrombophilia in a pedigree from the West of France by interfering with hemostatic balance of endothelial cells [32]. The quest for characterizing disease variant(s) may encounter challenges stemming from several factors including the complexity of the underlying genetic models and the large number of candidate variants that can be detected by WES/WGS, even in informative extended pedigrees. For instance, the implementation of a WES strategy in 10 related individuals affected with VTE from 2 independent Dutch studies [33] was not able to identify the culprit variant(s). Nevertheless, this work identified 3 candidate loci on chromosomes 2p21, 5q35.2, and 17q25.1, with several candidate variants in multiple genes at a given locus such as *PLEKHH2*, *LRPPRC*, and *SRBD1* at the 2p21 locus.

NGS approaches for detecting pathogenic rare variants can also be applied in very few affected individuals pending careful selection of individuals to be sequenced (eg, patients with unprovoked VTE at young age) aiming to enrich the sample with "genetic cases." Even in absence of DNA samples from extended families, selection of independent cases with familial history can be a good strategy, as exemplified by the identification of *FLI1* gene involved in platelet disorders from the analysis of only 13 unrelated cases [34]. The selection of patients with early-onset disease and/or recurrent events is another way to increase the chance to identify rare pathogenic variants, a strategy adopted by de la Morena-Barrio et al. [35] to identify 2 novel variants in *SERPINC1*, Glu227Lys, and Asn224His, causing severe thrombophilia in patients with normal antithrombin activity. The likelihood of identifying very rare deleterious variants could also be increased by sequencing individuals whose phenotypes deviate from predictions based on biological or clinical parameters. This includes sequencing disease patients presenting with an extremely low burden of common susceptibility alleles known to increase the risk of a given disease [36], a strategy that has not yet been deployed in the context of VTE. Along the same lines, by sequencing a patient with PE for whom a machine learning algorithm based on plasma proteomics data had inaccurately predicted DVT, we identified the presence of the extremely rare *PROC* pR272C variant. This variation has a minor allele frequency (MAF) lower than 0.00006 in public database and was previously reported to be associated with moderate protein C deficiency [37]. In the sequenced patient, protein C levels were 63%, slightly lower than the recommended threshold to declare moderate protein C deficiency [38].

Nothing, of course, prevents deploying an NGS strategy within the context of large case-control and population-based samples, regardless of the cost of this sequencing. In such contexts, the goal is not necessarily to discover novel variants but rather to provide statistical support for the involvement of (new) genes in the etiology of a disease. Instead of testing the association of single variants with a studied phenotype, the overall idea is to combine the information on a set of several preselected variants and to assess whether a given gene is enriched in such variants

associated with the phenotype of interest. Various statistical methods for this so-called gene-burden analysis have been proposed, and readers are encouraged to refer to the study by Nicolae [39] for a comprehensive introduction to these methodologies. This strategy was used in the Malmö Diet and Cancer Study to provide genetic support for the role of 2 candidate genes, *SERPINA1* [40] and *TFPI* [41], in VTE risk. By conducting a gene-based approach on coding variants with MAF lower than 0.1%, which are either loss-of-function or nonbenign (as defined by the PolyPhen bioinformatics tool [http://genetics.bwh.harvard.edu/pph2/]) missense variants and studying them in ∼29 000 population-based samples, it was observed that carrying at least 1 such "qualifying variant" was marginally ($P \sim .02$) associated with a 3-fold increased risk of VTE [41]. Using the same methodology, this group observed that being compound heterozygote for qualifying variants in *SERPINA1* or being homozygous for the rare allele of the Glu366Lys variant was associated with a ∼4-fold increased risk of VTE ($P = 3.4 \times 10^{-4}$).

By defining qualifying variants as any coding variant with MAF <0.05% that is a loss-of-function, splice, or nonbenign variant (still as defined by PolyPhen) and applying a WES approach in a case-control sample of 393 VTE patients and ∼6100 health individuals, Desch et al. [42] identified *STAB2* as a novel disease gene for VTE with the presence of carrying qualifying variants being ∼3 times more frequent in cases than in healthy individuals. By contrast, the WGS analysis of 3793 VTE cases and 7834 controls as part of the Trans-Omics for Precision Medicine (TOPMed) initiative [43] using similar statistical methodologies did not reveal novel VTE genes. Two points to note, however: first, in this work, the definition of qualifying variants was much broader as it included missense variants predicted to be deleterious by at least 1 of the 5 prediction tools used and synonymous variants predicted to be deleterious by FATHMM-XF [44]; second, a novel gene, *MS4A1*, reached the prespecified statistical threshold of $P < 2.3 \times 10^{-6}$ (corrected for the number of genes) for association with VTE in the subgroup of patients with unprovoked VTE once all missense variants were considered and not only those bioinformatically predicted to be deleterious. This association deserves further replication works.

These 4 examples highlight several limitations due to the definition of qualifying variants that may warrant harmonization, at the very least through discussions within the International Society on Thrombosis and Haemostasis (ISTH) community. Indeed, the statistical results provided by gene-burden analyses are strongly influenced by the chosen MAF thresholds, the type of included variants (missense, nonsense, frameshift, synonymous, and splice, among others), and the prediction tools used to call for a variant's deleteriousness. Furthermore, until now, these analyses do not consider other noncoding gene variants (deep intronic, 5′UTR, and 3′UTR), as mentioned earlier, despite their rightful inclusion in the spectrum of potentially pathogenic variants. It is crucial to underscore that the statistical association of a set of qualifying variants with a specific phenotype does not necessarily imply that all the included variants have a genuine functional role in the biology of that phenotype. Experimental investigations are essential to precisely identify which variants have a true functional impact, particularly in the context of their potential application in clinical practice.

As their name suggests, these gene-burden methods have primarily been applied to the analysis of variants within gene regions. Some extensions have been suggested to facilitate the analysis of variants located outside gene regions, primarily by employing sliding window approaches [43,45]. While such a strategy did not succeed in identifying novel VTE-associated loci in the TOPMed initiative [43], it did enable the suggestion of the existence of a locus associated with early-onset VTE through the analysis of a moderate sample comprising 200 VTE patients [45]. This locus on chromosome 18q22.2 maps to an enhancer region adjacent to *CD226*, a glycoprotein involved in platelet and vascular endothelial cell biology. Its precise role in thrombosis, if any, warrants further validation. These strategies suffer from the same limitations as those mentioned above for gene-burden methods, with additional challenges such as, for example, the size of the windows to be analyzed.

4 | WHICH SEQUENCING DEPTH?

The choice of sequencing depth to adopt hinges on the study's context, be it clinical or basic research.

In a clinical context focused on obtaining molecular diagnoses, there is no universal consensus on the minimum coverage depth, and each clinical laboratory may establish its own threshold. Nevertheless, a highly profound depth (eg, ∼500×) is imperative to guarantee high accuracy in identifying rare variants and their corresponding genotypes.

In the early days of NGS applied to basic research, it was recommended to undertake sequencing projects with average depths of 30× to 40× to identify, with a high probability, all extremely rare variants. Nowadays, the trend is shifting toward low-pass (or low-coverage) sequencing (∼4×) when the goals are to estimate allele frequencies and to perform GWASs in large population-based or case-control samples, aiming to detect novel disease genes [46,47]. It has been suggested to further reduce the sequencing depth when the sequenced data are complemented with high-throughput genotype data and imputation [48,49]. However, a sequencing depth of higher magnitude (>40X) remains indispensable when the focus is on detecting mosaicism or mutations not present in all cell types, as observed in clonal hematopoiesis of indeterminate potential (CHIP). CHIP is defined as the acquisition of somatic mutations that drive clonal expansion in the absence of cytopenia and dysplastic hematopoiesis. CHIP is the subject of intensive research in the cardiovascular field [50,51], and thrombosis does not escape the rule. Using whole exome data from ∼200 000 participants of the UK Biobank, Dikilitas et al. [52] reported a modest increased risk of VTE (OR, 1.6; $P = .03$) in CHIP carriers. This association was slightly higher when the analysis was restricted to PE patients (OR, 1.80; $P = .02$). In this work, the presence of CHIP was sought in 3 driver genes, *DNMT3A*, *TET2*, and *ASXL1*, and a CHIP was called when the somatic mutation presented with a variant allele frequency (VAF) ≥10%. In a small retrospective study involving 61 unprovoked PE patients, the search for mutations with a VAF of >1% in these 3 genes, along with *SF3B1* and *TP53*, identified a consistent pattern of association between CHIP and PE [53]. Once more, harmonization of the tested genes and the VAF

threshold would be welcome as different studies are using different criteria to characterize the presence of a CHIP [50,51,54,55]. Once a CHIP is detected in a patient, the question that arises concerns its true pathogenicity in relation to the patient's disease. During the ISTH 2023 State of the Art session on "Translational Genomics in Thrombosis: From OMICs to Clinics," we reported the case of a nonsmoker woman with DVT at 19 years, with normal body mass index, not using oral contraceptives and free of thrombophilia. A 30× WGS was performed in this patient, revealing as the sole identified candidate a rare variation: the previously unreported c.1668-1G>C variant in exon 15 of *DNMT3A* (NM_022552), predicted to impact splicing. This prediction was subsequently validated through a minigene splicing assay. Using WGS data, its VAF was estimated to be 16%, a value further reevaluated by digital polymerase chain reaction at a VAF of 5%. The investigation into whether this CHIP is responsible for the observed early-onset thrombotic event, and if so, through which mechanisms, or if it serves as an indicator of covert cancer is still ongoing.

5 | SHORT- OR LONG-READ SEQUENCING?

In recent years, new technologies known as long-read sequencing technologies have emerged. In contrast to initial short-read technologies (where sequenced fragments are typically less than 150 bp), which have been successful with NGS instruments like Illumina, these long-read technologies, sometimes referred to as third-generation sequencing technologies, provide the capability to sequence fragments with lengths of up to 20 000 bp [56]. These technologies are highly efficient in detecting any kind of genetic variation, in particular structural variants [57–59]. The use of these long-read sequencing technologies in the field of hemostasis is still in its infancy, but the initial applications exemplify their full potential. For instance, these permit to characterize a structural variant extending over 5 exons in *ITGB3* in a patient with Glanzmann thrombasthenia [60] and several *SERPINC1* structural variants, including a retroposon element, in patients with antithrombin deficiency [61]. Long-read sequencing technologies also have the advantage of permitting phasing of the detected genetic variations [62], then increasing the efficiency of haplotype analyses to detect SNP × SNP interactive effects, to facilitate the identification of causal variants from sets of polymorphisms in linkage disequilibrium and to detect haplotype founder effect as recently demonstrated for the *SERPINC1* Ile386Thr mutation [63]. Lastly, using single-molecule real-time sequencing, such technology also offers the possibility to look for differentially methylated DNA regions with high accuracy [64].

6 | WHERE, THEN, DO THE OBSTACLES LIE FOR THE WIDESPREAD ADOPTION OF NGS?

Throughout this review, we have endeavored to highlight the undeniable advantages of NGS while occasionally acknowledging some of its limitations. However, there is a "minor" aspect that we have overlooked. NGS technologies are generating a wealth of genetic variations, and as illustrated by some aforementioned examples, the path to identifying the causal variant(s) and characterizing them is far from being straightforward. In the initial phase of prioritization to streamline the list of candidate variants identified in a given NGS project, a common approach involves the utilization of bioinformatics prediction tools to identify variants with a high likelihood of functional impact. However, the abundance of available prediction tools often results in divergent predictions. With the deployment of artificial intelligence–based methods integrating several prediction tools and a large set of genomic features, one can expect that these predictions will become more and more accurate, especially for some types of variants that are still difficult to predict such as splice [65], deep intronic [66], or regulatory [67–69] variants. Unfortunately, even after meticulously selecting potential candidates, the resulting list of variants compatible with the assumed genetic model can still be extensive, rendering it impractical to experimentally test them, as illustrated with the aforementioned WES conducted in 2 extended Dutch pedigrees [33]. If a substantial number of candidate variants persist, employing massive parallel assays [70–72] can serve as a secondary prioritization strategy to discern the most plausible ones. For genetic models involving more than one variant, such high-throughput assays may prove insufficient. In such instances, gathering statistical support for specific candidate associations of interest—perhaps through the utilization of large genomic biobanks—could serve as a means to refine the list of candidates before undertaking further experimental validation. In the ISTH 2023 State of the Art session on "Translational Genomics in Thrombosis: From OMICs to Clinics," we also documented the case of a 19-year-old woman who experienced PE 4 months after initiating oral contraceptive. Through WGS, we identified in this patient 2 extremely rare variants of unknown significance, the *F5* p.Thr464Ile (rs141768227) and *F2* p.Arg608= (rs3136532). In a case-control study of ∼400 VTE cases and ∼400 controls, the latter has been proposed as a possible risk factor for VTE [73], but it did not show up in the large International Network of Venous Thrombosis GWAS conducted in more than 80 000 cases [8]. It would then be particularly interesting to study the simultaneous presence of these 2 variations in large biobanks to determine whether their joint co-occurrence is prothrombotic as opposed to their isolated presence. Such initiatives aimed at identifying digenism using large-scale genomics databases have been initiated [74], presenting promising prospects.

One final obstacle, and not the least, is the cost, which although has substantially decreased over the last years, still remains extremely high, making this technology inaccessible in low-income countries where NGS technologies are not widely implemented, as nicely illustrated by Paula Heller during the ISTH 2023 Scientific and Standardization Committee 02.3 session on "Reach the World – Perspectives From Around the Globe in Platelet Disorders."

7 | CONCLUDING REMARKS

To summarize, several unresolved questions arise during the implementation of NGS in the clinical diagnosis of thrombosis (and beyond):

(1) Can NGS panel replace the initial thrombophilia screening or should it be reserved for VTE patients in whom the classic thrombophilia assessment result is negative? (2) Which genes should be prioritized for sequencing? (3) In which patient populations should NGS implementation be considered? (4) How should the identified variants of uncertain significance be handled? (5) What bioinformatic tools and depth of experimental exploration are necessary to reclassify these variants of uncertain significance as class 4 or 5 (ie, pathogenic or likely pathogenic variant according to American College of Medical Genetics and Genomics)? International collaborative efforts are indeed necessary to address these issues [75] and should be encouraged. The GoldVariant initiative sponsored by the ISTH Scientific and Standardization Committee Subcommittee on Genomics Hemostasis to share variants identified in NGS projects [76] and the ClinGen resource [77] aimed at providing guidelines for the clinical relevance of variants serve as good examples.

While NGS has become nearly indispensable in the field of clinical diagnostics (even if Sanger sequencing is still effective [78]), this is not yet the case in basic research. Indeed, NGS is not the panacea and genotype-based research strategies still play a crucial role in identifying new genes, even for rare variant analysis [79]. In this invited review, we focused on NGS for DNA variants analysis mainly in the VTE context. There is much more to explore regarding the application of NGS for other molecular investigations such as RNA sequencing (including in clinics [80,81]), epigenetics, and functional genomics [82,83].

ACKNOWLEDGMENTS

D.-A.T.'s and M.P.-E.'s venous thromboembolism genomics research projects were supported by the INSERM GOLD Cross-Cutting Program, the GENMED Laboratory of Excellence on Medical Genomics (ANR-10-LABX-0013), a research program managed by the National Research Agency (ANR) as part of the French Investment for the Future, and the French Investigation Network on Venous Thrombo-Embolism.

AUTHOR CONTRIBUTIONS

D.-A.T. and P.-E.M. wrote this invited review.

DECLARATION OF COMPETING INTERESTS

There are no competing interests to disclose.

ORCID

David-Alexandre Trégouët https://orcid.org/0000-0001-9084-7800

X, FORMERLY KNOWN AS TWITTER

David-Alexandre Trégouët @TregouetBPH

REFERENCES

[1] Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, Abraham J, Adair T, Aggarwal R, Ahn SY, Alvarado M, Anderson HR, Anderson LM, Andrews KG, Atkinson C, Baddour LM, Barker-Collo S, Bartels DH, Bell ML, Benjamin EJ, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012;380:2095–128.

[2] Wendelboe AM, Raskob GE. Global burden of thrombosis: epidemiologic aspects. *Circ Res*. 2016;118:1340–7.

[3] Connors JM. Thrombophilia testing and venous thrombosis. *N Engl J Med*. 2017;377:2298.

[4] Couturaud F, Sanchez O, Pernod G, Mismetti P, Jego P, Duhamel E, Provost K, dit Sollier CB, Presles E, Castellant P, Parent F, Salaun PY, Bressollette L, Nonent M, Lorillon P, Girard P, Lacut K, Guégan M, Bosson JL, Laporte S, et al. Six months vs extended oral anticoagulation after a first episode of pulmonary embolism: the PADIS-PE randomized clinical trial. *JAMA*. 2015;314:31–40.

[5] Agnelli G, Prandoni P, Santamaria MG, Bagatella P, Iorio A, Bazzan M, Moia M, Guazzaloca G, Bertoldi A, Tomasi C, Scannapieco G, Ageno W, Warfarin Optimal Duration Italian Trial Investigators. Three months versus one year of oral anticoagulant therapy for idiopathic deep venous thrombosis. Warfarin Optimal Duration Italian Trial Investigators. *N Engl J Med*. 2001;345:165–9.

[6] Klarin D, Busenkell E, Judy R, Lynch J, Levin M, Haessler J, Aragam K, Chaffin M, Haas M, Lindström S, Assimes TL, Huang J, Min Lee K, Shao Q, Huffman JE, Kabrhel C, Huang Y, Sun YV, Vujkovic M, Saleheen D, et al. Genome-wide association analysis of venous thromboembolism identifies new risk loci and genetic overlap with arterial vascular disease. *Nat Genet*. 2019;51:1574–9.

[7] Ghouse J, Tragante V, Ahlberg G, Rand SA, Jespersen JB, Leinøe EB, Vissing CR, Trudsø L, Jonsdottir I, Banasik K, Brunak S, Ostrowski SR, Pedersen OB, Sørensen E, Erikstrup C, Bruun MT, Nielsen KR, Køber L, Christensen AH, Iversen K, et al. Genome-wide meta-analysis identifies 93 risk loci and enables risk prediction equivalent to monogenic forms of venous thromboembolism. *Nat Genet*. 2023;55:399–409.

[8] Thibord F, Klarin D, Brody JA, Chen MH, Levin MG, Chasman DI, Goode EL, Hveem K, Teder-Laving M, Martinez-Perez A, Aïssi D, Daian-Bacq D, Ito K, Natarajan P, Lutsey PL, Nadkarni GN, de Vries PS, Cuellar-Partida G, Wolford BN, Pattee JW, et al. Cross-ancestry investigation of venous thromboembolism genomic predictors. *Circulation*. 2022;146:1225–42.

[9] Larsen TB, Sørensen HT, Skytthe A, Johnsen SP, Vaupel JW, Christensen K. Major genetic susceptibility for venous thromboembolism in men: a study of Danish twins. *Epidemiology*. 2003;14:328–32.

[10] Harris RF. Hapmap flap. *Curr Biol*. 2002;12:R827.

[11] Castoldi E, Hézard N, Mourey G, Wichapong K, Poggi M, Ibrahim-Kosta M, Thomassen MCLGD, Fournel A, Hayward CPM, Alessi MC, Hackeng TM, Rosing J, Morange PE. Severe thrombophilia in a factor V-deficient patient homozygous for the Ala2086Asp mutation (FV Besançon). *J Thromb Haemost*. 2021;19:1186–99.

[12] Simioni P, Cagnin S, Sartorello F, Sales G, Pagani L, Bulato C, Gavasso S, Nuzzo F, Chemello F, Radu CM, Tormene D, Spiezia L, Hackeng TM, Campello E, Castoldi E. Partial F8 gene duplication (factor VIII Padua) associated with high factor VIII levels and familial thrombophilia. *Blood*. 2021;137:2383–93.

[13] Simioni P, Tormene D, Tognin G, Gavasso S, Bulato C, Iacobelli NP, Finn JD, Spiezia L, Radu C, Arruda VR. X-linked thrombophilia with a mutant factor IX (factor IX Padua). *N Engl J Med*. 2009;361:1671–5.

[14] Miyawaki Y, Suzuki A, Fujita J, Maki A, Okuyama E, Murata M, Takagi A, Murate T, Kunishima S, Sakai M, Okamoto K, Matsushita T, Naoe T, Saito H, Kojima T. Thrombosis from a prothrombin mutation conveying antithrombin resistance. *N Engl J Med*. 2012;366:2390–6.

[15] Djordjevic V, Kovac M, Miljic P, Murata M, Takagi A, Pruner I, Francuski D, Kojima T, Radojkovic D. A novel prothrombin mutation in two families with prominent thrombophilia–the first cases of antithrombin resistance in a Caucasian population. *J Thromb Haemost*. 2013;11:1936–9.

[16] Bulato C, Radu CM, Campello E, Gavasso S, Spiezia L, Tormene D, Simioni P. New prothrombin mutation (Arg596Trp, prothrombin Padua 2) associated with venous thromboembolism. *Arterioscler Thromb Vasc Biol*. 2016;36:1022–9.

[17] Nogami K, Shinozawa K, Ogiwara K, Matsumoto T, Amano K, Fukutake K, Shima M. Novel FV mutation (W1920R, FV_{Nara}) associated with serious deep vein thrombosis and more potent APC resistance relative to FV_{Leiden}. *Blood*. 2014;123:2420–8.

[18] Van Laer C, Jacquemin M, Baert S, Labarque V, Thys C, Vanassche T, Van Geet C, Verhamme P, Willekens K, Corveleyn A, Peerlinck K, Freson K. Clinical application of multigene panel testing for bleeding, thrombotic, and platelet disorders: a 3-year Belgian experience. *J Thromb Haemost*. 2023;21:887–95.

[19] Athar M, Ghita IS, Albagenny AA, Abduljaleel Z, Shadab G, Elsendiony A, Halawani SH, Alkazmi MM, Alquthami K, Alkhuzae MM, Althebyani AA, Bogari NM, Dannoun A, Al-Allaf FA. Targeted next-generation sequencing reveals novel and known variants of thrombophilia associated genes in Saudi patients with venous thromboembolism. *Clin Chim Acta*. 2021;519:247–54.

[20] Pagliari MT, Cairo A, Boscarino M, Mancini I, Pappalardo E, Bucciarelli P, Martinelli I, Rosendaal FR, Peyvandi F. Role of ADAMTS13, VWF and F8 genes in deep vein thrombosis. *PLoS One*. 2021;16:e0258675.

[21] Downes K, Megy K, Duarte D, Vries M, Gebhart J, Hofer S, Shamardina O, Deevi SVV, Stephens J, Mapeta R, Tuna S, Al Hasso N, Besser MW, Cooper N, Daugherty L, Gleadall N, Greene D, Haimel M, Martin H, Papadia S, et al. Diagnostic high-throughput sequencing of 2396 patients with bleeding, thrombotic, and platelet disorders. *Blood*. 2019;134:2082–91.

[22] Danckwardt S, Hartmann K, Katz B, Hentze MW, Levy Y, Eichele R, Deutsch V, Kulozik AE, Ben-Tal O. The prothrombin 20209 C→T mutation in Jewish-Moroccan Caucasians: molecular analysis of gain-of-function of 3′ end processing. *J Thromb Haemost*. 2006;4:1078–85.

[23] Poort SR, Rosendaal FR, Reitsma PH, Bertina RM. A common genetic variation in the 3′-untranslated region of the prothrombin gene is associated with elevated plasma prothrombin levels and an increase in venous thrombosis. *Blood*. 1996;88:3698–703.

[24] Uitte de Willige S, de Visser MCH, 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;106:4176–83.

[25] Labrouche-Colomer S, Soukarieh O, Proust C, Mouton C, Huguenin Y, Roux M, Besse C, Boland A, Olaso R, Constans J, Deleuze JF, Morange PE, Jaspard-Vinassa B, Trégouët DA, GenMed Consortium. A novel rare c.-39C>T mutation in the PROS1 5′UTR causing PS deficiency by creating a new upstream translation initiation codon. *Clin Sci (Lond)*. 2020;134:1181–90.

[26] 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;92:177–87.

[27] Xie Z, Lu Y, Liu C, Sun C, Yu J, Ling C, Luan X, Wang W, Wang L, Liang Y, Luo Q, Meng L, Wang Z, Yuan Y. Cryptic exon activation caused by a novel deep-intronic splice-altering variant in Becker muscular dystrophy. *J Clin Lab Anal*. 2023;37:e24987.

[28] AlAbdi L, Shamseldin HE, Khouj E, Helaby R, Aljamal B, Alqahtani M, Almulhim A, Hamid H, Hashem MO, Abdulwahab F, Abouyousef O, Jaafar A, Alshidi T, Al-Owain M, Alhashem A, Al Tala S, Khan AO, Mardawi E, Alkuraya H, Faqeih E, et al. Beyond the exome: utility of long-read whole genome sequencing in exome-negative autosomal recessive diseases. *Genome Med*. 2023;15:114.

[29] Chung CCY, Hue SPY, Ng NYT, Doong PHL, , Hong Kong Genome Project, Chu ATW, Chung BHY. Meta-analysis of the diagnostic and clinical utility of exome and genome sequencing in pediatric and adult patients with rare diseases across diverse populations. *Genet Med*. 2023;25:100896.

[30] Noetzli L, Lo RW, Lee-Sherick AB, Callaghan M, Noris P, Savoia A, Rajpurkar M, Jones K, Gowan K, Balduini C, Pecci A, Gnan C, De Rocco D, Doubek M, Li L, Lu L, Leung R, Landolt-Marticorena C, Hunger S, Heller P, et al. Germline mutations in ETV6 are associated with thrombocytopenia, red cell macrocytosis and predisposition to lymphoblastic leukemia. *Nat Genet*. 2015;47:535–8.

[31] Canault M, Ghalloussi D, Grosdidier C, Guinier M, Perret C, Chelghoum N, Germain M, Raslova H, Peiretti F, Morange PE, Saut N, Pillois X, Nurden AT, Cambien F, Pierres A, van den Berg TK, Kuijpers TW, Alessi MC, Tregouet DA. Human CalDAG-GEFI gene (RASGRP2) mutation affects platelet function and causes severe bleeding. *J Exp Med*. 2014;211:1349–62.

[32] Morange PE, Peiretti F, Gourhant L, Proust C, Soukarieh O, Pulcrano-Nicolas AS, Saripella GV, Stefanucci L, Lacroix R, Ibrahim-Kosta M, Lemarié CA, Frontini M, Alessi MC, Trégouët DA, Couturaud F. A rare coding mutation in the MAST2 gene causes venous thrombosis in a French family with unexplained thrombophilia: the Breizh MAST2 Arg89Gln variant. *PLoS Genet*. 2021;17:e1009284.

[33] Cunha MLR, Meijers JCM, Rosendaal FR, van Hylckama Vlieg A, Reitsma PH, Middeldorp S. Whole exome sequencing in thrombophilic pedigrees to identify genetic risk factors for venous thromboembolism. *PLoS One*. 2017;12:e0187699.

[34] Stockley J, Morgan NV, Bem D, Lowe GC, Lordkipanidzé M, Dawood B, Simpson MA, Macfarlane K, Horner K, Leo VC, Talks K, Motwani J, Wilde JT, Collins PW, Makris M, Watson SP, Daly ME, UK Genotyping and Phenotyping of Platelets Study Group. Enrichment of FLI1 and RUNX1 mutations in families with excessive bleeding and platelet dense granule secretion defects. *Blood*. 2013;122:4090–3.

[35] de la Morena-Barrio ME, Suchon P, Jacobsen EM, Iversen N, Miñano A, de la Morena-Barrio B, Bravo-Pérez C, Padilla J, Cifuentes R, Asenjo S, Deleuze JF, Trégouët DA, Lozano ML, Vicente V, Sandset PM, Morange PE, Corral J. Two SERPINC1 variants affecting N-glycosylation of Asn224 cause severe thrombophilia not detected by functional assays. *Blood*. 2022;140:140–51.

[36] Lu T, Zhou S, Wu H, Forgetta V, Greenwood CMT, Richards JB. Individuals with common diseases but with a low polygenic risk score could be prioritized for rare variant screening. *Genet Med*. 2021;23:508–15.

[37] Reitsma PH, Poort SR, Allaart CF, Briët E, Bertina RM. The spectrum of genetic defects in a panel of 40 Dutch families with symptomatic protein C deficiency type I: heterogeneity and founder effects. *Blood*. 1991;78:890–4.

[38] Razzaq M, Iglesias MJ, Ibrahim-Kosta M, Goumidi L, Soukarieh O, Proust C, Roux M, Suchon P, Boland A, Daiain D, Olaso R, Havervall S, Thalin C, Butler L, Deleuze JF, Odeberg J, Morange PE, Trégouët DA. An artificial neural network approach integrating plasma proteomics and genetic data identifies PLXNA4 as a new susceptibility locus for pulmonary embolism. *Sci Rep*. 2021;11:14015.

[39] Nicolae DL. Association tests for rare variants. *Annu Rev Genomics Hum Genet*. 2016;17:117–30.

[40] Manderstedt E, Halldén C, Lind-Halldén C, Elf J, Svensson PJ, Engström G, Melander O, Baras A, Lotta LA, Zöller B. Regeneron Genetics Center. Thrombotic risk determined by rare and common SERPINA1 variants in a population-based cohort study. *J Thromb Haemost*. 2022;20:1421–7.

[41] Manderstedt E, Lind-Halldén C, Halldén C, Elf J, Svensson PJ, Engström G, Melander O, Baras A, Lotta LA, Zöller B, Regeneron Genetics Center. Genetic variation of the blood coagulation regulator tissue factor pathway inhibitor and venous thromboembolism

among middle-aged and older adults: a population-based cohort study. *Res Pract Thromb Haemost*. 2022;6:e12842.

[42] Desch KC, Ozel AB, Halvorsen M, Jacobi PM, Golden K, Underwood M, Germain M, Tregouet DA, Reitsma PH, Kearon C, Mokry L, Richards JB, Williams F, Li JZ, Goldstein D, Ginsburg D. Whole-exome sequencing identifies rare variants in STAB2 associated with venous thromboembolic disease. *Blood*. 2020;136:533–41.

[43] Seyerle AA, Laurie CA, Coombes BJ, Jain D, Conomos MP, Brody J, Chen MH, Gogarten SM, Beutel KM, Gupta N, Heckbert SR, Jackson RD, Johnson AD, Ko D, Manson JE, McKnight B, Metcalf GA, Morrison AC, Reiner AP, Sofer T, et al. Whole genome analysis of venous thromboembolism: the Trans-Omics for Precision Medicine Program. *Circ Genom Precis Med*. 2023;16:e003532.

[44] Rogers MF, Shihab HA, Mort M, Cooper DN, Gaunt TR, Campbell C. FATHMM-XF: accurate prediction of pathogenic point mutations via extended features. *Bioinformatics*. 2018;34:511–3.

[45] Bocher O, Ludwig TE, Oglobinsky MS, Marenne G, Deleuze JF, Suryakant S, Odeberg J, Morange PE, Trégouët DA, Perdry H, Génin E. Testing for association with rare variants in the coding and non-coding genome: RAVA-FIRST, a new approach based on CADD deleteriousness score. *PLoS Genet*. 2022;18:e1009923.

[46] Gilly A, Southam L, Suveges D, Kuchenbaecker K, Moore R, Melloni GEM, Hatzikotoulas K, Farmaki AE, Ritchie G, Schwartzentruber J, Danecek P, Kilian B, Pollard MO, Ge X, Tsafantakis E, Dedoussis G, Zeggini E. Very low-depth whole-genome sequencing in complex trait association studies. *Bioinformatics*. 2019;35:2555–61.

[47] Li Y, Sidore C, Kang HM, Boehnke M, Abecasis GR. Low-coverage sequencing: implications for design of complex trait association studies. *Genome Res*. 2011;21:940–51.

[48] Li JH, Mazur CA, Berisa T, Pickrell JK. Low-pass sequencing increases the power of GWAS and decreases measurement error of polygenic risk scores compared to genotyping arrays. *Genome Res*. 2021;31:529–37.

[49] Rubinacci S, Hofmeister RJ, Sousa da Mota B, Delaneau O. Imputation of low-coverage sequencing data from 150,119 UK Biobank genomes. *Nat Genet*. 2023;55:1088–90.

[50] Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG, Lindsley RC, Mermel CH, Burtt N, Chavez A, Higgins JM, Moltchanov V, Kuo FC, Kluk MJ, Henderson B, Kinnunen L, Koistinen HA, Ladenvall C, Getz G, Correa A, et al. Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med*. 2014;371:2488–98.

[51] Kar SP, Quiros PM, Gu M, Jiang T, Mitchell J, Langdon R, Iyer V, Barcena C, Vijayabaskar MS, Fabre MA, Carter P, Petrovski S, Burgess S, Vassiliou GS. Genome-wide analyses of 200,453 individuals yield new insights into the causes and consequences of clonal hematopoiesis. *Nat Genet*. 2022;54:1155–66.

[52] Dikilitas O, Saadatagah S, Satterfield B, Kullo IJ. Clonal hematopoiesis of indeterminate potential predicts incident venous thromboembolism in the UK Biobank cohort. *Circulation*. 2021;144(Suppl 1):A12959.

[53] Soudet S, Jedraszak G, Evrard O, Marolleau JP, Garcon L, Pietri MAS. Is hematopoietic clonality of indetermined potential a risk factor for pulmonary embolism? *TH Open*. 2021;5:e338–42.

[54] Kessler MD, Damask A, O'Keeffe S, Banerjee N, Li D, Watanabe K, Marketta A, Van Meter M, Semrau S, Horowitz J, Tang J, Kosmicki JA, Rajagopal VM, Zou Y, Houvras Y, Ghosh A, Gillies C, Mbatchou J, White RR, Verweij N, et al. Common and rare variant associations with clonal haematopoiesis phenotypes. *Nature*. 2022;612:301–9.

[55] Genovese G, Kähler AK, Handsaker RE, Lindberg J, Rose SA, Bakhoum SF, Chambert K, Mick E, Neale BM, Fromer M, Purcell SM, Svantesson O, Landén M, Höglund M, Lehmann S, Gabriel SB, Moran JL, Lander ES, Sullivan PF, Sklar P, et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med*. 2014;371:2477–87.

[56] Logsdon GA, Vollger MR, Eichler EE. Long-read human genome sequencing and its applications. *Nat Rev Genet*. 2020;21:597–614.

[57] Aganezov S, Yan SM, Soto DC, Kirsche M, Zarate S, Avdeyev P, Taylor DJ, Shafin K, Shumate A, Xiao C, Wagner J, McDaniel J, Olson ND, Sauria MEG, Vollger MR, Rhie A, Meredith M, Martin S, Lee J, Koren S, et al. A complete reference genome improves analysis of human genetic variation. *Science*. 2022;376:eabl3533.

[58] Hoyt SJ, Storer JM, Hartley GA, Grady PGS, Gershman A, de Lima LG, Limouse C, Halabian R, Wojenski L, Rodriguez M, Altemose N, Rhie A, Core LJ, Gerton JL, Makalowski W, Olson D, Rosen J, Smit AFA, Straight AF, Vollger MR, et al. From telomere to telomere: the transcriptional and epigenetic state of human repeat elements. *Science*. 2022;376:eabk3112.

[59] Wagner J, Olson ND, Harris L, Khan Z, Farek J, Mahmoud M, Stankovic A, Kovacevic V, Yoo B, Miller N, Rosenfeld JA, Ni B, Zarate S, Kirsche M, Aganezov S, Schatz MC, Narzisi G, Byrska-Bishop M, Clarke W, Evani US, et al. Benchmarking challenging small variants with linked and long reads. *Cell Genom*. 2022;2:100128.

[60] Zamora-Cánovas A, de la Morena-Barrio B, Marín-Quilez A, Sierra-Aisa C, Male C, Fernández-Mosteirin N, Trapero-Marugán M, Padilla J, Garrido-Rodriguez P, Sánchez-Fuentes A, Rodríguez-Alen A, Gómez-González PL, Revilla N, de la Morena-Barrio ME, Bastida JM, Corral J, Rivera J, Lozano ML. Targeted long-read sequencing identifies and characterizes structural variants in cases of inherited platelet disorders. *J Thromb Haemost*. 2024;22:851–9.

[61] de la Morena-Barrio B, Stephens J, de la Morena-Barrio ME, Stefanucci L, Padilla J, Miñano A, Gleadall N, García JL, López-Fernández MF, Morange PE, Puurunen M, Undas A, Vidal F, Raymond FL, Vicente V, Ouwehand WH, Corral J, Sanchis-Juan A, NIHR BioResource. Long-read sequencing identifies the first retrotransposon insertion and resolves structural variants causing antithrombin deficiency. *Thromb Haemost*. 2022;122:1369–78.

[62] Ebert P, Audano PA, Zhu Q, Rodriguez-Martin B, Porubsky D, Bonder MJ, Sulovari A, Ebler J, Zhou W, Serra Mari R, Yilmaz F, Zhao X, Hsieh P, Lee J, Kumar S, Lin J, Rausch T, Chen Y, Ren J, Santamarina M, et al. Haplotype-resolved diverse human genomes and integrated analysis of structural variation. *Science*. 2021;372:eabf7117.

[63] Weronska A, De la Morena-Barrio B, Goldman-Mazur S, De la Morena-Barrio ME, Padilla J, Miñano A, Garrido P, Treliński J, Potaczek DP, Szczepanek A, Undas A, Corral J, Wypasek E. Functional, biochemical, molecular and clinical characterization of antithrombin c.1157T>C (p.Ile386Thr), a recurrent Polish variant with a founder effect. *Haematologica*. 2023;108:2803–7.

[64] Cheung WA, Johnson AF, Rowell WJ, Farrow E, Hall R, Cohen ASA, Means JC, Zion TN, Portik DM, Saunders CT, Koseva B, Bi C, Truong TK, Schwendinger-Schreck C, Yoo B, Johnston JJ, Gibson M, Evrony G, Rizzo WB, Thiffault I, et al. Direct haplotype-resolved 5-base HiFi sequencing for genome-wide profiling of hypermethylation outliers in a rare disease cohort. *Nat Commun*. 2023;14:3090.

[65] Sullivan PJ, Gayevskiy V, Davis RL, Wong M, Mayoh C, Mallawaarachchi A, Hort Y, McCabe MJ, Beecroft S, Jackson MR, Arts P, Dubowsky A, Laing N, Dinger ME, Scott HS, Oates E, Pinese M, Cowley MJ. Introme accurately predicts the impact of coding and noncoding variants on gene splicing, with clinical applications. *Genome Biol*. 2023;24:118.

[66] Barbosa P, Savisaar R, Carmo-Fonseca M, Fonseca A. Computational prediction of human deep intronic variation. *Gigascience*. 2022;12:giad085.

[67] Romo L, Findlay SD, Burge CB. Regulatory features aid interpretation of 3′UTR variants. *Am J Hum Genet*. 2024;111:350–63.

[68] Zhou J, Weinberger DR, Han S. Deep learning predicts DNA methylation regulatory variants in specific brain cell types and enhances fine mapping for brain disorders. Preprint. Posted online January 21, 2024. *bioRxiv*. 2024;01(18):576319. https://doi.org/10.1101/2024.01.18.576319

[69] Shukla K, Idanwekhai K, Naradikian M, Ting S, Schoenberger SP, Brunk E. Machine learning of three-dimensional protein structures to predict the functional impacts of genome variation. *J Chem Inf Model*. Published online April 18, 2024. https://doi.org/10.1021/acs.jcim.3c01967

[70] Patwardhan RP, Lee C, Litvin O, Young DL, Pe'er D, Shendure J. High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. *Nat Biotechnol*. 2009;27:1173–5.

[71] Sample PJ, Wang B, Reid DW, Presnyak V, McFadyen IJ, Morris DR, Seelig G. Human 5′ UTR design and variant effect prediction from a massively parallel translation assay. *Nat Biotechnol*. 2019;37:803–9.

[72] Patwardhan RP, Hiatt JB, Witten DM, Kim MJ, Smith RP, May D, Lee C, Andrie JM, Lee SI, Cooper GM, Ahituv N, Pennacchio LA, Shendure J. Massively parallel functional dissection of mammalian enhancers in vivo. *Nat Biotechnol*. 2012;30:265–70.

[73] Pruner I, Farm M, Tomic B, Gvozdenov M, Kovac M, Miljic P, Soutari NMH, Antovic A, Radojkovic D, Antovic J, Djordjevic V. The silence speaks, but we do not listen: synonymous c.1824C>T gene variant in the last exon of the prothrombin gene as a new prothrombotic risk factor. *Clin Chem*. 2020;66:379–89.

[74] Ryu J, Jurgens S, Choi SH, Koyama S, Ellinor P, Lubitz S, Bauer K, Bendapudi PK. Determining the risk of thrombosis in double heterozygous carriers of factor V Leiden and prothrombin G20210A mutation using population-scale genomics [abstract]. *Res Pract Thromb Haemost*. 2023;7(Suppl 2):100618.

[75] Lee AI, Connors JM. The next(gen) step in coagulation testing. *Blood*. 2019;134:2002–3.

[76] Megy K, Downes K, Morel-Kopp MC, Bastida JM, Brooks S, Bury L, Leinoe E, Gomez K, Morgan NV, Othman M, Ouwehand WH, Perez Botero J, Rivera J, Schulze H, Trégouët DA, Freson K. GoldVariants, a resource for sharing rare genetic variants detected in bleeding, thrombotic, and platelet disorders: communication from the ISTH SSC Subcommittee on Genomics in Thrombosis and Hemostasis. *J Thromb Haemost*. 2021;19:2612–7.

[77] Ross JE, Mohan S, Zhang J, Sullivan MJ, Bury L, Lee K, Futchi I, Frantz A, McDougal D, Perez Botero J, Cattaneo M, Cooper N, Downes K, Gresele P, Keenan C, Lee AI, Megy K, Morange PE, Morgan NV, Schulze H, et al. Evaluating the clinical validity of genes related to hemostasis and thrombosis using the Clinical Genome Resource gene curation framework. *J Thromb Haemost*. 2024;22:645–65.

[78] Larsen OH, Kjaergaard AD, Hvas AM, Nissen PH. Genetic variants in the protein S (PROS1) gene and protein S deficiency in a Danish population. *TH Open*. 2021;5:e479–88.

[79] Deguchi H, Shukla M, Hayat M, Torkamani A, Elias DJ, Griffin JH. Novel exomic rare variants associated with venous thrombosis. *Br J Haematol*. 2020;190:783–6.

[80] Gonorazky HD, Naumenko S, Ramani AK, Nelakuditi V, Mashouri P, Wang P, Kao D, Ohri K, Viththiyapaskaran S, Tarnopolsky MA, Mathews KD, Moore SA, Osorio AN, Villanova D, Kemaladewi DU, Cohn RD, Brudno M, Dowling JJ. Expanding the boundaries of RNA sequencing as a diagnostic tool for rare Mendelian disease. *Am J Hum Genet*. 2019;104:466–83.

[81] Ketkar S, Burrage LC, Lee B. RNA sequencing as a diagnostic tool. *JAMA*. 2023;329:85–6.

[82] Luo L, Gribskov M, Wang S. Bibliometric review of ATAC-seq and its application in gene expression. *Brief Bioinform*. 2022;23:bbac061.

[83] Jiang S, Mortazavi A. Integrating ChIP-seq with other functional genomics data. *Brief Funct Genomics*. 2018;17:104–15.