Received: 25 February 2022 | Accepted: 2 May 2022

DOI: 10.1111/jth.15754

REVIEW ARTICLE

Non-coding genetic variation in regulatory elements determines thrombosis and hemostasis phenotypes

Luca Stefanucci[1,2,3] | Mattia Frontini[1,2,3,4]

[1]Department of Haematology, University of Cambridge, Cambridge Biomedical Campus, Cambridge, UK

[2]National Health Service (NHS) Blood and Transplant, Cambridge Biomedical Campus, Cambridge, UK

[3]British Heart Foundation, Cambridge Centre for Research Excellence, University of Cambridge, Cambridge Biomedical Campus, Cambridge, UK

[4]Institute of Biomedical & Clinical Science, College of Medicine and Health, University of Exeter Medical School, Exeter, UK

Correspondence
Mattia Frontini, Institute of Biomedical & Clinical Science, College of Medicine and Health, University of Exeter Medical School, RILD Building, Barrack Road, Exeter EX2 5DW, UK.
Email: m.frontini@exeter.ac.uk

[Correction added on June 3, 2022, after first online publication: The title of the article is corrected to 'Non-coding genetic variation in regulatory elements determines thrombosis and hemostasis phenotypes.']

Funding information
British Heart Foundation, Grant/Award Number: FS/18/53/33863 and RE/18/1/34212

Abstract

Since the early inception of genome-wide association studies (GWAS), it became clear that, in all diseases or traits studied, most genetic variants are likely to exert their effect on gene expression mainly by altering the function of regulatory elements. At the same time, the regulation of the gene expression field broadened its boundaries, from the univocal relationship between regulatory elements and genes to include genome organization, long-range DNA interactions, and epigenetics. Next-generation sequencing has introduced genome-wide approaches that have greatly improved our understanding of the general principles of gene expression. However, elucidating how these apply in every single genomic locus still requires painstaking experimental work, in which several independent lines of evidence are required, and often this is helped by rare genetic variants in individuals with rare diseases. This review will focus on the non-coding features of the genome involved in transcriptional regulation, that when altered, leads to known cases of inherited (familial) thrombotic and hemostatic phenotypes, emphasizing the role of enhancers and super-enhancers.

KEYWORDS
endothelial cells, gene regulation, hemostasis, megakaryocytes, super-enhancer, thrombosis

1 | INTRODUCTION

How genes and genetic polymorphisms influence human traits, and consequently cause diseases, has been a central question in biology and medicine since genetic inception.[1–3] The technological developments that occurred over the last three decades have profoundly impacted the understanding of this topic.[4] Genome-wide association studies (GWAS) and whole, or targeted, genome sequencing have identified thousands of common and rare variants that influence human traits and diseases.[5–7] The majority of these trait-modifying polymorphisms are located in the 98% of the genome that does not encode for a protein (i.e., non-coding genome), implying that they do not alter a protein amino acid sequence. Instead, these variants are thought to be of regulatory nature for the causal genes. Gene expression quantitative traits loci (eQTL) studies have confirmed the regulatory nature of some of those, where enough statistical power (sample size and effect size) was available.[8]

Colocalization analysis overlays information from independent sources and traits (e.g., GWAS and eQTL) and tests them for signals that

Manuscript handled by: Jill Johnsen

Final decision: Jill Johnsen, 02 May 2022

TABLE 1 Summary of the non-coding regulatory variants discussed in this review

Variant	Gene	Phenotype	Possible mechanism	PMID
rs1175170	*RGS18*	Platelet aggregation	Alteration of GATA1 and NFE2 binding site	34131117
rs10886430	*GRK5*	Platelet activation	Alteration of GATA1 and MEIS1 binding site	34581777
GenBank: GQ246945	*PLAU*	Gain-of-function platelet dependent fibrinolysis	Gene duplication leads to enhancer hijacking	20007542, 32663239
GRCh37: CTCF3 4:155539849_155540258del	*FGA FGB FGG*	Reduction in fibrinogen levels	Loss of a CTCF binding site and consequent loss of local chromatin interactions	30039577
GRCh37: CTCF4 4:155543772_155544212del	*FGA FGB FGG*	Reduction in fibrinogen levels	Loss of a CTCF binding site and consequent loss of local chromatin interactions	30039577
GRCh37: X: 154230198–154252817	*F8*	Elevated FVIII levels and familial thrombophilia.	Duplication of *F8* gene promoter leads to the increased level of *F8* transcript	33275657
GRCh38: 10:27042550_28567796_dup-inv-dup	*WAC-ANKRD26 fusion*	Familial thrombocytopenia	Gain-of-function and cryptic *ANKRD26* TSS	33857290
rs9349379	*EDN1*	Increased risk of coronary artery disease, migraine headache, cervical artery dissection, fibromuscular dysplasia, and hypertension	Increase expression of *EDN1* via the alteration of the enhancer within the third intron of PHACTR1.	28753427
GRCh37: 1:145399075_145594214del	*RBM8A*	Thrombocytopenia and absent radii (TAR) syndrome	This variant reduces the function of the *RBM8A* promoter	22366785
rs12041331	*PEAR1*	Lower platelet function on aspirin and risk factor for cardiovascular events	Minor allele leads to a loss of the methylation and reductions of *PEAR1* expression	27313330

Abbreviations: CTCF, CCCTC-binding factor; FVIII, factor VIII; PMID, PubMed reference number.

are consistent with a shared causal variant.[9] This approach is used to connect variants to genes and phenotypes, to identify molecular and cellular phenotypes (e.g., transcription levels) that are relevant for more complex traits (e.g., GWAS-associated disease) and to determine the mechanism by which the GWAS variants are influencing the phenotype. Colocalization of GWAS and eQTL variants in tissues implicated in thrombosis and hemostasis has been reported in various studies.[10–13] Among others, rs1175170 was identified as a regulator of *RGS18* transcription in platelets, linking this gene to arterial thrombosis.[12] Colocalization also can be strengthened using additional chromatin features. For instance, Downes and colleagues identified rs10886430, in a *GRK5* intron, as a regulator of platelet activation through the protease-activated receptor-1 pathway. The alternative nucleotide in rs10886430 locus alters *GATA1* and *MEIS1* binding sites in a megakaryocyte-specific enhancer (Table 1) and alters *GRK5* expression level.[13]

In parallel, the understanding of the role of non-coding genomes increased exponentially.[14,15] The last decade has been crucial to untangling the structure, regulation, and function of the genome, a field of study generally referred to as functional genomics (Figure 1).[16,17] For instance, we now know that the control of gene expression in a spatio-temporal fashion results from a dynamic and unique combination of DNA topology and regulatory elements activity to the point that cell identities are more granularly defined by their chromatin features than by the gene expression patterns[18,19] and that most of the genome has some sort of regulatory function in one cell type or another.[15]

2 | GENOME ARCHITECTURE AND TRANSCRIPTION

Multicellular organisms derive all cell types, with vastly different functions, using different parts of the information contained in the genome.[20] Evolutionarily this has been achieved with the use of intergenic regions that structured and controlled gene expression.[21] The function of the higher order of the genome is 2-fold: (1) to separate active regions from inactive ones, called A and B compartments,[22] respectively; and (2) to connect the regulatory regions to the genes, and to do so in a manner that avoids spurious gene activation. This is achieved by anchoring DNA to the nuclear lamina[23,24] and/or via DNA looping (Figure 1).[25] Some loops are implicated in the tridimensional organization of the genome, while others are directly involved in transcriptional regulation by bringing together promoters and enhancers.[26] Loops, in the interphase, are mainly organized by architectural factors such as the CCCTC-binding factor (CTCF), the cohesin complex,[27,28] and other factors that bind to the DNA.[29]

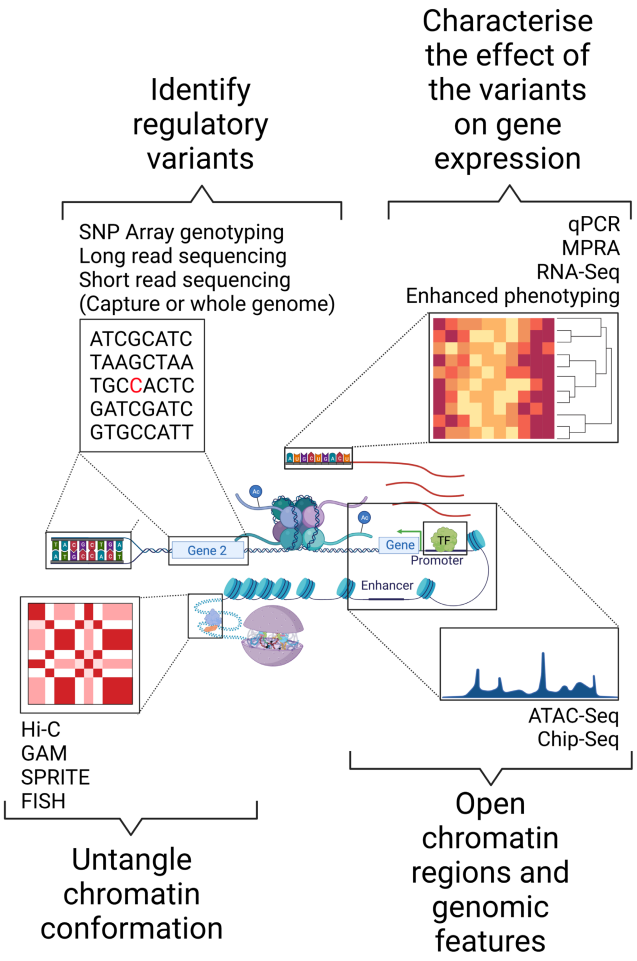

FIGURE 1 Chromatin structure, genomic features, and technologies widely adopted in functional genomics studies to characterize regulatory variants, cognate genes, and their effect on transcription. Several technologies can identify genetic variants and their location in regulatory regions.[56,58,80,81] To associate regulatory variants to their cognate genes, a series of other technologies are needed to investigate the chromatin structure in a cell-type–specific fashion.[30,42,82] To constrain the genome to regulatory regions, technologies such as ChIP-Seq and ATAC-Seq can inform us about the chromatin function via its post-translational modifications and accessibility. The effect of regulatory variants on transcription can be estimated with MPRA[83] and/or other technologies. FISH, fluorescence in situ hybridization; GAM, genome architecture mapping; MPRA, massively parallel reporter assays; qPCR, quantitative polymerase chain reaction; SNP, single nucleotide polymorphism; SPRITE, split-pool recognition of interactions by tag extension.

3 | TOPOLOGICALLY ASSOCIATING DOMAINS

The sub-chromosomal regions considered, to some extent, DNA functional units, are called topologically associating domains (TADs; Figure 1).[30] TADs impose some spatial constraints on DNA's ability to move, increasing the probability of interaction between regulatory regions and cognate genes.[31] Their sizes in the human genome are variable, but on average are around one megabase (i.e., 10^6 base pairs; Mb).[32] TADs were first observed in all-versus-all chromatin conformation capture (3C) experiments and then confirmed using microscopy approaches.[32–36] These topological domains are mostly conserved across cell types[32,37] and contain, to some extent, all the genomic features that are required to allow the physiological gene expression (e.g., enhancers, promoters, and genes).[38,39] Smaller-scale structures are observed within the TADs and are often referred to as sub-TADs.[33,40,41] These are highly dynamic structures that vary quite a lot from cell type to cell type and are mainly driven by promoter-enhancer interactions.[37,42] The interactions occurring within TADs are crucial for gene expression but also to correctly structure the topology of TAD and sub-TAD domains.[43] TADs' boundaries are enriched with features like regulatory elements and genes. However, it must be noted that TAD boundaries have different abilities to insulate.[44,45] While some exert a robust insulating effect, others do not, and allow interactions between different domains.[46] Disruption of strong boundaries, either due to their deletion or by chromosomal rearrangement, as well as the formation of new ones, may result in alteration of gene expression and pathological sequelae.[47,48] For example, tandem duplication of the plasminogen activator urokinase (*PLAU*) gene and one of the enhancers for *VCL*, disrupting the sub-TAD organization of this region on chromosome 10, results in *PLAU* over-expression platelets and Quebec platelet disorder (Table 1).[49] This phenomenon is known as enhancer hijacking and, in this case, results in a dominant platelet-dependent fibrinolysis.[50] Similarly, the expression of the fibrinogen gene cluster (*FGA*, *FGB*, *FGG*) is controlled via four enhancers, CNC12, PFE2, E3, and E4, located close to it.[51] At the edge of this gene cluster, there is a CTCF binding site. Removing the *FGG*-closest CTCF binding site rearranges the TAD, resulting in a reduction of *FGB* and *FGG* expression levels and a consequent halving of the amount of fibrinogen secreted from hepatic cells (Table 1).[52]

On the other hand, enhancers and promoters directly orchestrate the transcriptional process by establishing a permissive chromatin environment and recruiting the machinery necessary for gene expression (Figure 1).[53,54]

4 | TRANSCRIPTION FACTORS

Specific DNA sequences that are recognized by transcription factors (TFs) allow this permissive status.[20] Pioneer TFs can bind to the DNA in the presence of nucleosomes and recruit remodeling complexes that displace the latter, creating open chromatin, thus allowing other TFs to bind to their motifs or binding site (TFBS).[20,55] This process occurs throughout organism development, from fertilization, through the three embryonic layers, down to the mature postmitotic cell types forming the different tissues and organs, sometimes with different TFs of the same family taking part in a relay to bind to the same site as differentiation proceeds.[18,56–58] Once TFs are bound to regulatory elements, the nearby nucleosomes are post-transcriptionally modified with marks of active chromatin while the recruitment of the transcriptional machinery begins.

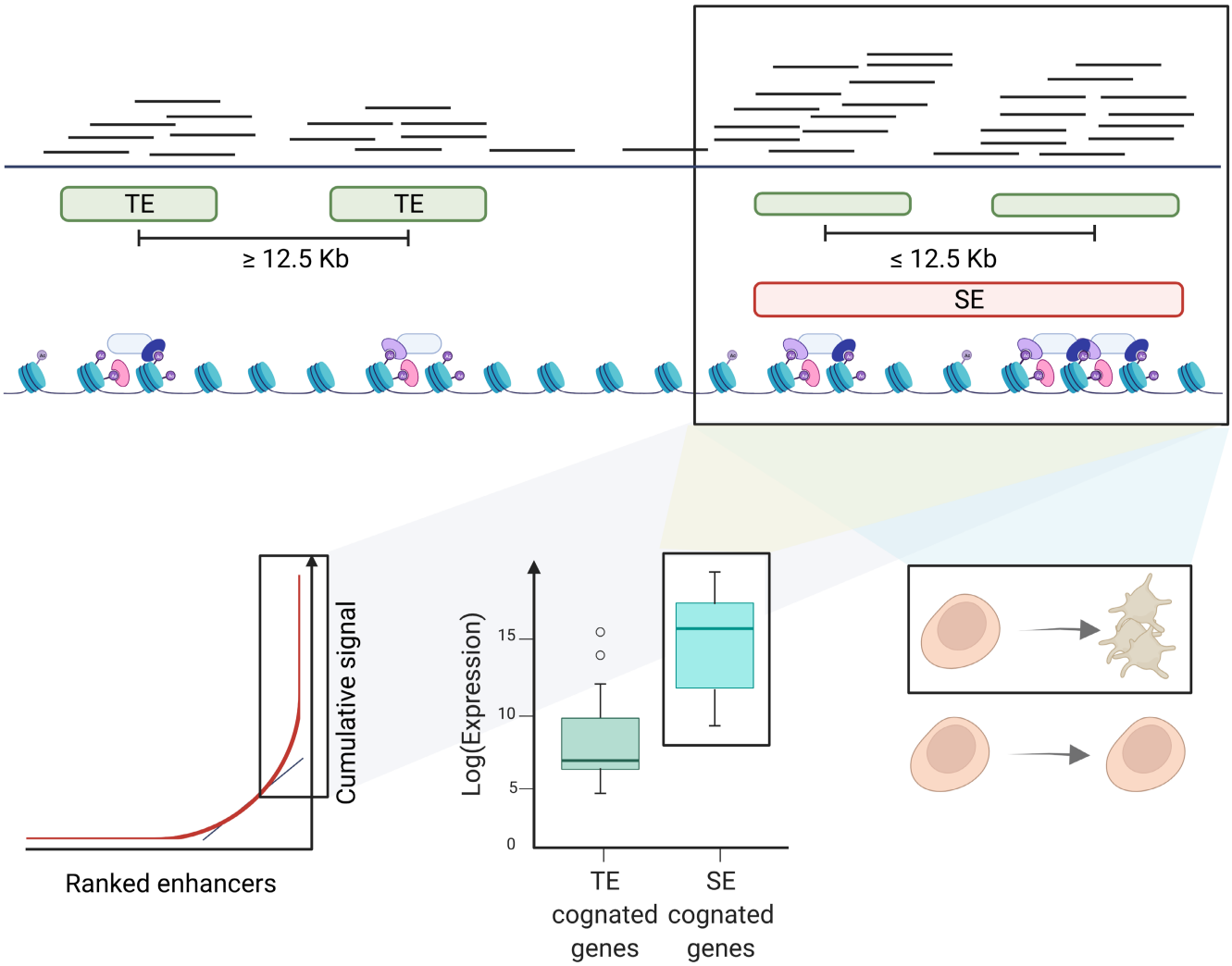

FIGURE 2 Super-enhancers (SEs; colored in red) definition via ChIP-Seq experiments and biological characteristics. Typical enhancers (TE; colored in green) are aggregated in SEs if the distance between them is less than 12.5 Kb. In ChIP-Seq experiments, SEs are characterized by having a larger amount of sequencing reads (H3K27ac, Med1, p300). SEs are defined as those that, in the ranking of the ChIP-Seq signal for H3K27ac (or Med1), are localized on the right side of the transition point (i.e., straight line of slope equals one and tangent to the curve).

The impact of DNA variants on TF binding was recognized early on with significant enrichment of variants associated with common diseases in open chromatin,[59] that is, where TFs are bound. Similar enrichments have been observed for platelet-related traits in the megakaryocyte's enhancers.[60] The consequences of genetic variation in regulatory elements span a wide range, from extremely small to very large. The former, often due to common variants, alters the observed trait by decimal points of the standard deviation percentage while the latter, usually associated with rare variants (minor allele frequency <0.1), drives the trait into the pathological spectrum.[61,62] An example of a common variant altering the phenotype of interest is rs9349379, located in the third intron of *PHACTR1*, and associated with five vascular diseases, including coronary heart disease (Table 1).[63] This variant lies within a regulatory element that controls the expression of endothelin 1 (*EDN1*) located 600 kilobases (kb) away. As an example of the latter, in the megakaryocyte/platelet axis, two rare variants critically altering TFBS and gene expression are (1) rs139428292, located in the 5′ UTR of *RBM8* and (2) a previously unknown polymorphism in the first intron of the same gene. When either is present in compound heterozygosity with a 1q21.1 deletion, the individual is affected by thrombocytopenia and absent radii syndrome (Table 1).[64]

5 | ENHANCERS AND PROMOTERS

Enhancers regulate gene expression mainly by coming in close proximity with the gene promoter and contributing to the recruitment of the necessary protein complexes, a model referred to as activity by contact.[65] There are also examples in which the enhancer needs to be located away from the regulated gene promoter.[66] In both cases, the enhancer positioning helps to reach a local conformation favoring transcription. Enhancer and promoter aberrations, either in quality or quantity, can be the etiology of human conditions. For

instance, a form of familial thrombophilia has been identified in two independent families that carry a tandem duplication of a part of the *F8* gene (exon 1 and intron 1; Table 1).[67] Simioni and colleagues[67] showed that the increased level of factor VIII (FVIII) is due to the duplication of a regulatory region present in *F8*'s first intron. The duplication of this enhancer increases the amount of relevant transcription factors that localize in the proximity of *F8* promoter and, as a consequence, inflates the amount of FVIII produced by hepatocytes. Wahlster and colleagues used long-read sequencing to identify a paired-duplication inversion of *ANKRD26-WAC* (Table 1)[68] that leads to *ANKRD26* not being silenced and consequently results in thrombocytopenia.

Active enhancers and promoters are labeled with several post-transcriptional modifications on the histones of the nearby nucleosomes. Among these, either histone 3 lysine 27 acetylation (H3K27ac) or histone 3 lysine 122 (H3K122Ac) together with histone 3 lysine 4 mono-methylation (H3K3me1)[69,70] label enhancers and H3K27Ac with H3K4me3 label promoters.[71]

6 | SUPER ENHANCERS

Soon after chromatin modification genome-wide studies became widely available, it was noted that the distribution of H3K27Ac is not equal across all enhancers, and a number of these are localized closer to each other than by chance.[72] Enhancers located less than 12.5 Kb from each other can be grouped into super-enhancers (SEs) as their constituents, also called stretch enhancers[73] (Figure 2). SEs have some distinguishing properties. (1) They contribute to the large majority of the H3K27ac signal[74] and some other regulatory proteins (e.g., Med1[72,75] and p300[76]; Figure 2). (2) Gene expression, on average, is higher in genes connected to SEs than in genes linked to the same number of regulatory regions as the SEs' constituents but located more than 12.5 kb apart (and therefore do not qualify as SEs; Figure 2).[60] (3) SEs play a pivotal role in regulating genes that orchestrate cell fate decisions during stem cell differentiation.[72,77]

In endothelial cells, the transcription factor *ERG* plays an essential role in establishing SEs, and variants associated with cardiovascular diseases are enriched in *ERG* TFBS localized in endothelial SEs.[78] In megakaryocytes, SE constituents are physically connected and regulate genes implicated in several cellular processes. In platelet traits (i.e., mass, count, mean volume, and distribution width), genetic variants harbored in SEs influence the expression of genes implicated in the archetypical functions of these cells (response to wounding/wound healing, coagulation, hemostasis, platelet degranulation, actin cytoskeleton remodeling, regulation of body fluid levels). This evidence indicates that genetic variation in these genomic regions plays a key role in determining how each individual responds to pro-coagulant stimuli.[60]

It is also interesting to note that, while each set of SEs defines the identity of a cell type, the majority of the SEs' constituents are already specified as open chromatin early on during development.[60,79] As an example, of the 1067 megakaryocyte SEs, only 24 have a fully open chromatin profile in hematopoietic progenitors. This means that the final set of SEs is fully established by controlling the opening of about 2100 constituents in the mature cells.[60] The 1067 SEs are connected to more than 3300 genes, and while there are several linear relationships between SE and genes, more complex relationships exist reflecting the constraint in degrees of freedom dictated by the DNA itself and the organization of RNA polymerase II factories.[80] It is likely that these interactions are not happening all at the same time and/or in every cell, as different conformation supporting transcription might occur and only single-cell data could provide a definitive answer.[35] For instance, the *VWF-CD9* locus is controlled by three SEs, each contacting the promoters of both genes, which are also in contact with each other. A genetic variant, rs2363877, linked to platelet traits, lies in one of the SEs, and controls the transcription of one of the two genes. The minor allele favors *VWF* expression at the expense of *CD9*.[60] Moreover, some of these interactions might be implicated in the silencing of *VWF*, whose expression, at least in endothelial cells, is controlled by a stochastic bi-stable switch mediated by DNA methylation.[81] DNA methylation plays an important role in hematopoiesis by determining permissive cell fates by controlling accessibility to regulatory elements.[82] The same mechanism is also used to control the expression of genes implicated in platelet reactivity and cardiovascular disease like *PEAR1* (Table 1).[83]

Overall, the last decade has opened a wealth of knowledge that has established several genome-wide principles on how gene expression is organized. Unfortunately, it is less clear how these principles apply to individual genes and orthogonal lines of evidence, obtained with painstaking laboratory work, are still required to determine the effects of specific regulatory sequences. The introduction of mid- and high-throughput measurements of functional phenotypes will lead, soon, to an increase in the number of discoveries linking phenotypes, including hemostasis and thrombosis, and diseases, with genotypes, especially rare variants, and one day there will be enough data to bypass the requirement for laboratory validation.

ACKNOWLEDGMENTS

We would like to apologize to the many colleagues whose outstanding work we could not cite due to space limitations. We would like to thank Dr Carmelo Ferrai (UMG, Goettingen) for the critical reading of the manuscript. L.S. was a PhD student supported by the British Heart Foundation Cambridge Centre for Research Excellence (RE/18/1/34212). M.F. is supported by the British Heart Foundation (FS/18/53/33863) and the British Heart Foundation Cambridge Centre for Research Excellence (RE/18/1/34212).

CONFLICT OF INTEREST

The authors have no conflicts of interest to disclose.

AUTHOR CONTRIBUTIONS

LS and MF discussed and wrote the manuscript together.

ORCID

Luca Stefanucci https://orcid.org/0000-0002-4352-1151

Mattia Frontini https://orcid.org/0000-0001-8074-6299

REFERENCES

1. Mendel G. Versuche über plflanzenhybriden. Verhandlungen Des Naturforschenden Vereines in brünn. *Abhandlungen*. 1866;Bd. IV für das jahr 1865:3-47.
2. Garrod A. The incidence of alkaptonuria: a study in chemical individuality. *Lancet*. 1902;160:1616-1620.
3. Fisher RA. XV.—The correlation between relatives on the supposition of mendelian inheritance. *Trans R Soc Edinburgh*. 1919;52:399-433.
4. Claussnitzer M, Cho JH, Collins R, et al. A brief history of human disease genetics. *Nature*. 2020;577:179-189.
5. Ozaki K, Ohnishi Y, Iida A, et al. Functional SNPs in the lymphotoxin-α gene that are associated with susceptibility to myocardial infarction. *Nat Genet*. 2002;32:650-654.
6. WTCCC. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007;447:661-678.
7. Buniello A, MacArthur JAL, Cerezo M, et al. The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res*. 2019;47:D1005-D1012.
8. Clyde D. Disease genomics: transitioning from association to causation with eQTLs. *Nat Rev Genet*. 2017;18:271.
9. Wallace C. A more accurate method for colocalisation analysis allowing for multiple causal variants. *PLoS Genet*. 2021;17:e1009440.
10. Franceschini N, Giambartolomei C, de Vries PS, et al. GWAS and colocalization analyses implicate carotid intima-media thickness and carotid plaque loci in cardiovascular outcomes. *Nat Commun*. 2018;9:5141.
11. Kammers K, Taub MA, Rodriguez B, et al. Transcriptional profile of platelets and iPSC-derived megakaryocytes from whole-genome and RNA sequencing. *Blood*. 2021;137:959-968.
12. Keramati AR, Chen M-H, Rodriguez BAT, et al. Genome sequencing unveils a regulatory landscape of platelet reactivity. *Nat Commun*. 2021;12:3626.
13. Downes K, Zhao X, Gleadall NS, et al. G protein–coupled receptor kinase 5 regulates thrombin signaling in platelets via PAR-1. *Blood Adv*. 2022;6:2319-2330.
14. Stunnenberg HG, International Human Epigenome Consortium, Hirst M. The international human epigenome consortium: a blueprint for scientific collaboration and discovery. *Cell*. 2016;167:1145-1149.
15. ENCODE Project Consortium, Moore JE, Purcaro MJ, et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature*. 2020;583:699-710.
16. Graur D, Zheng Y, Price N, Azevedo RBR, Zufall RA, Elhaik E. On the immortality of television sets: "function" in the human genome according to the evolution-free gospel of ENCODE. *Genome Biol Evol*. 2013;5:578-590.
17. Pevsner J. *Bioinformatics and Functional Genomics*. John Wiley & Sons; 2015.
18. Song L, Zhang Z, Grasfeder LL, et al. Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. *Genome Res*. 2011;21:1757-1767.
19. Rubin AJ, Barajas BC, Furlan-Magaril M, et al. Lineage-specific dynamic and pre-established enhancer-promoter contacts cooperate in terminal differentiation. *Nat Genet*. 2017;49:1522-1528.
20. Shlyueva D, Stampfel G, Stark A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet*. 2014;15:272-286.
21. Wray GA, Hahn MW, Abouheif E, et al. The evolution of transcriptional regulation in eukaryotes. *Mol Biol Evol*. 2003;20:1377-1419.
22. Lieberman-Aiden E, van Berkum NL, Williams L, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. 2009;326:289-293.
23. Bridger JM, Foeger N, Kill IR, Herrmann H. The nuclear lamina. Both a structural framework and a platform for genome organization. *FEBS J*. 2007;274:1354-1361.
24. Peric-Hupkes D, van Steensel B. Role of the nuclear lamina in genome organization and gene expression. *Cold Spring Harb Symp Quant Biol*. 2010;75:517-524.
25. Nuebler J, Fudenberg G, Imakaev M, Abdennur N, Mirny LA. Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proc Natl Acad Sci USA*. 2018;115:E6697-E6706.
26. Bouwman BAM, de Laat W. Getting the genome in shape: the formation of loops, domains and compartments. *Genome Biol*. 2015;16:154.
27. Rubio ED, Reiss DJ, Welcsh PL, et al. CTCF physically links cohesin to chromatin. *Proc Natl Acad Sci USA*. 2008;105:8309-8314.
28. Merkenschlager M, Nora EP. CTCF and Cohesin in genome folding and transcriptional gene regulation. *Annu Rev Genomics Hum Genet*. 2016;17:17-43.
29. Dequeker BJH, Brandão HB, Scherr MJ, Gassler J, Powell S, Gaspar I, Flyamer IM, Tang W, Stocsits R, Davidson IF, Peters J-M, Duderstadt KE, Mirny LA, Tachibana K. MCM complexes are barriers that restrict cohesin-mediated loop extrusion. *bioRxiv*. 2020. p. 2020.10.15.340356.
30. Krijger PHL, de Laat W. Regulation of disease-associated gene expression in the 3D genome. *Nat Rev Mol Cell Biol*. 2016;17:771-782.
31. Schoenfelder S, Fraser P. Long-range enhancer–promoter contacts in gene expression control. *Nat Rev Genet*. 2019;20:437-455.
32. Dixon JR, Selvaraj S, Yue F, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*. 2012;485:376-380.
33. Nora EP, Lajoie BR, Schulz EG, et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*. 2012;485:381-385.
34. Giorgetti L, Galupa R, Nora EP, et al. Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell*. 2014;157:950-963.
35. Bintu B, Mateo LJ, Su J-H, et al. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science*. 2018;362:eaau1783.
36. Boettiger A, Murphy S. Advances in chromatin imaging at kilobase-scale resolution. *Trends Genet*. 2020;36:273-287.
37. Dixon JR, Gorkin DU, Ren B. Chromatin domains: the unit of chromosome organization. *Mol Cell*. 2016;62:668-680.
38. Sexton T, Cavalli G. The role of chromosome domains in shaping the functional genome. *Cell*. 2015;160:1049-1059.
39. Dekker J, Heard E. Structural and functional diversity of topologically associating domains. *FEBS Lett*. 2015;589:2877-2884.
40. Shen Y, Yue F, McCleary DF, et al. A map of the cis-regulatory sequences in the mouse genome. *Nature*. 2012;488:116-120.
41. Rao SSP, Huntley MH, Durand NC, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*. 2014;159:1665-1680.
42. Dowen JM, Fan ZP, Hnisz D, et al. Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell*. 2014;159:374-387.
43. van Steensel B, Furlong EEM. The role of transcription in shaping the spatial organization of the genome. *Nat Rev Mol Cell Biol*. 2019;20:327-337.
44. Nora EP, Goloborodko A, Valton A-L, et al. Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell*. 2017;169:930-944.e22.

45. Gong Y, Lazaris C, Sakellaropoulos T, et al. Stratification of TAD boundaries reveals preferential insulation of super-enhancers by strong boundaries. *Nat Commun*. 2018;9:542.
46. Javierre BM, Burren OS, Wilder SP, Kreuzhuber R, Hill SM, Sewitz S, Cairns J, Wingett SW, Várnai C, Thiecke MJ, Burden F, Farrow S, Cutler AJ, Rehnström K, Downes K, Grassi L, Kostadima M, Freire-Pritchett P, Wang F, BLUEPRINT Consortium, et al. Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell*. 2016;167:1369–1384.e19.
47. Ibrahim DM, Mundlos S. The role of 3D chromatin domains in gene regulation: a multi-facetted view on genome organization. *Curr Opin Genet Dev*. 2020;61:1-8.
48. Allou L, Balzano S, Magg A, et al. Non-coding deletions identify Maenli lncRNA as a limb-specific En1 regulator. *Nature*. 2021;592:93-98.
49. Liang M, Soomro A, Tasneem S, et al. Enhancer-gene rewiring in the pathogenesis of Quebec platelet disorder. *Blood*. 2020;136:2679-2690.
50. Kahr WH, Zheng S, Sheth PM, et al. Platelets from patients with the Quebec platelet disorder contain and secrete abnormal amounts of urokinase-type plasminogen activator. *Blood*. 2001;98:257-265.
51. Fort A, Fish RJ, Attanasio C, Dosch R, Visel A, Neerman-Arbez M. A liver enhancer in the fibrinogen gene cluster. *Blood*. 2011;117:276-282.
52. Espitia Jaimes C, Fish RJ, Neerman-Arbez M. Local chromatin interactions contribute to expression of the fibrinogen gene cluster. *J Thromb Haemost*. 2018;16:2070-2082.
53. Serfling E, Jasin M, Schaffner W. Enhancers and eukaryotic gene transcription. *Trends Genet*. 1985;1:224-230.
54. Bulger M, Groudine M. Functional and mechanistic diversity of distal transcription enhancers. *Cell*. 2011;144:327-339.
55. Zaret KS, Carroll JS. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev*. 2011;25:2227-2241.
56. Chen X, Xu H, Yuan P, et al. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell*. 2008;133:1106-1117.
57. Natoli G. Maintaining cell identity through global control of genomic organization. *Immunity*. 2010;33:12-24.
58. Ostuni R, Piccolo V, Barozzi I, et al. Latent enhancers activated by stimulation in differentiated cells. *Cell*. 2013;152:157-171.
59. Maurano MT, Humbert R, Rynes E, et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science*. 2012;337:1190-1195.
60. Petersen R, Lambourne JJ, Javierre BM, Grassi L, Kreuzhuber R, Ruklisa D, Rosa IM, Tomé AR, Elding H, van Geffen JP, Jiang T, Farrow S, Cairns J, Al-Subaie AM, Ashford S, Attwood A, Batista J, Bouman H, Burden F, Choudry FA, et al. Platelet function is modified by common sequence variation in megakaryocyte super enhancers. *Nat Commun*. 2017;8:16058.
61. Bomba L, Walter K, Soranzo N. The impact of rare and low-frequency genetic variants in common disease. *Genome Biology*. 2017;18:77.
62. Turro E, Astle WJ, Megy K, et al. Whole-genome sequencing of patients with rare diseases in a national health system. *Nature*. 2020;583:96-102.
63. Gupta RM, Hadaya J, Trehan A, et al. A genetic variant associated with five vascular diseases is a distal regulator of Endothelin-1 gene expression. *Cell*. 2017;170:522-533.e15.
64. Albers CA, Paul DS, Schulze H, Freson K, Stephens JC, Smethurst PA, Jolley JD, Cvejic A, Kostadima M, Bertone P, Breuning MH, Debili N, Deloukas P, Favier R, Fiedler J, Hobbs CM, Huang N, Hurles ME, Kiddle G, Krapels I, Nurden P, Ruivenkamp CAL, Sambrook JG, Smith K, Stemple DL, Strauss G, Thys C, van Geet C, Newbury-Ecob R, Ouwehand WH, Ghevaert C Compound inheritance of a low-frequency regulatory SNP and a rare null mutation in exon-junction complex subunit RBM8A causes TAR syndrome. *Nat Genet*. 2012;44:435-439, S1-S2.
65. Fulco CP, Nasser J, Jones TR, et al. Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat Genet*. 2019;51:1664-1669.
66. Benabdallah NS, Williamson I, Illingworth RS, et al. Decreased enhancer-promoter proximity accompanying enhancer activation. *Mol Cell*. 2019;76:473-484.e7.
67. Simioni P, Cagnin S, Sartorello F, et al. Partial F8 gene duplication (factor VIII Padua) associated with high factor VIII levels and familial thrombophilia. *Blood*. 2021;137:2383-2393.
68. Wahlster L, Verboon JM, Ludwig LS, et al. Familial thrombocytopenia due to a complex structural variant resulting in a WAC-ANKRD26 fusion transcript. *J Exp Med*. 2021;218:e20210444.
69. Creyghton MP, Cheng AW, Welstead GG, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci USA*. 2010;107:21931-21936.
70. Pradeepa MM, Grimes GR, Kumar Y, et al. Histone H3 globular domain acetylation identifies a new class of enhancers. *Nat Genet*. 2016;48:681-686.
71. Liang G, Lin JCY, Wei V, et al. Distinct localization of histone H3 acetylation and H3-K4 methylation to the transcription start sites in the human genome. *Proc Natl Acad Sci U A*. 2004;101:7357-7362.
72. Whyte WA, Orlando DA, Hnisz D, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*. 2013;153:307-319.
73. SCJ P, Stitzel ML, Taylor DL, et al. Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc Natl Acad Sci U S A*. 2013;110:17921-17926.
74. Hnisz D, Abraham BJ, Lee TI, et al. Super-enhancers in the control of cell identity and disease. *Cell*. 2013;155:934-947.
75. Lovén J, Hoke HA, Lin CY, et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell*. 2013;153:320-334.
76. Lee B-K, Jang YJ, Kim M, LeBlanc L, Rhee C, Lee J, Beck S, Shen W, Kim J. Super-enhancer-guided mapping of regulatory networks controlling mouse trophoblast stem cells. *Nat Commun*. 2019;10:4749.
77. Pott S, Lieb JD. What are super-enhancers? *Nat Genet*. 2015;47:8-12.
78. Kalna V, Yang Y, Peghaire CR, et al. The transcription factor ERG regulates super-enhancers associated with an endothelial-specific gene expression program. *Circ Res*. 2019;124:1337-1349.
79. Lorzadeh A, Hammond C, Wang F, et al. Polycomb contraction differentially regulates terminal human hematopoietic differentiation programs. *bioRxiv*. 2020;647438.
80. Buckley MS, Lis JT. Imaging RNA polymerase II transcription sites in living cells. *Curr Opin Genet Dev*. 2014;25:126-130.
81. Yuan L, Chan GC, Beeler D, Janes L, Spokes KC, Dharaneeswaran H, Mojiri A, Adams WJ, Sciuto T, Garcia-Cardeña G, Molema G, Kang PM, Jahroudi N, Marsden PA, Dvorak A, Regan ER, Aird WC. A role of stochastic phenotype switching in generating mosaic endothelial cell heterogeneity. *Nat Commun*. 2016;7:10160.
82. Farlik M, Halbritter F, Müller F, et al. DNA methylation dynamics of human hematopoietic stem cell differentiation. *Cell Stem Cell*. 2016;19:808-822.
83. Izzi B, Pistoni M, Cludts K, et al. Allele-specific DNA methylation reinforces PEAR1 enhancer activity. *Blood*. 2016;128:1003-1012.

How to cite this article: Stefanucci L, Frontini M. Non-coding genetic variation in regulatory elements determines thrombosis and hemostasis phenotypes. *J Thromb Haemost*. 2022;20:1759-1765. doi: 10.1111/jth.15754