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

Luca Stefanucci1,2,3 | Mattia Frontini1,2,3,4

1Department of Haematology, University of Cambridge, Cambridge Biomedical Campus, Cambridge, UK
2National Health Service (NHS) Blood and Transplant, Cambridge Biomedical Campus, Cambridge, UK
3British Heart Foundation, Cambridge Centre for Research Excellence, University of Cambridge, Cambridge Biomedical Campus, Cambridge, UK
4Institute of Biomedical & Clinical Science, College of Medicine and Health, University of Exeter Medical School, Exeter, UK

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
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 spatiotemporal 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 patterns18,19 and that most of the genome has some sort of regulatory function in one cell type or another.15

| 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        |

**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.
The sub-chromosomal regions considered, to some extent, DNA functional units, are called topologically associating domains (TADs; Figure 1). TADs impose some spatial constraints on DNA's ability to move, increasing the probability of interaction between regulatory regions and cognate genes. Their sizes in the human genome are variable, but on average are around one megabase (i.e., \(10^6\) base pairs; Mb). TADs were first observed in all-versus-all chromatin conformation capture (3C) experiments and then confirmed using microscopy approaches. These topological domains are mostly conserved across cell types and contain, to some extent, all the genomic features that are required to allow the physiological gene expression (e.g., enhancers, promoters, and genes). Smaller-scale studies are observed within the TADs and are often referred to as sub-TADs. These are highly dynamic structures that vary quite a lot from cell type to cell type and are mainly driven by promoter-enhancer interactions. The interactions occurring within TADs are crucial for gene expression but also to correctly structure the topology of TAD and sub-TAD domains. TADs' boundaries are enriched with features like regulatory elements and genes. However, it must be noted that TAD boundaries have different abilities to insulate. While some exert a robust insulating effect, others do not, and allow interactions between different domains. 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. 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). 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). Specific DNA sequences that are recognized by transcription factors (TFs) allow this permissive status. 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). 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. 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.
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 \textit{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 (\textit{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 \textit{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

\section*{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). Simioni and colleagues 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 FVIII produced by hepatocytes. Wahlster and colleagues used long-read sequencing to identify a paired-duplication inversion of ANKR26-WAC (Table 1) that leads to ANKR26 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)\textsuperscript{69,70} label enhancers and H3K27Ac with H3K4me3 label promoters.\textsuperscript{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.\textsuperscript{72} Enhancers located less than 12.5 Kbp from each other can be grouped into super-enhancers (SEs) as their constituents, also called stretch enhancers (Figure 2). SEs have some distinguishing properties. (1) They contribute to the large majority of the H3K27ac signal and some other regulatory proteins (e.g., Med1\textsuperscript{72,75} and p300\textsuperscript{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).\textsuperscript{60} (3) SEs play a pivotal role in regulating genes that orchestrate cell fate decisions during stem cell differentiation.\textsuperscript{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. 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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{81} DNA methylation plays an important role in hematopoiesis by determining permissive cell fates by controlling accessibility to regulatory elements.\textsuperscript{82} The same mechanism is also used to control the expression of genes implicated in platelet reactivity and cardiovascular disease like PEAR1 (Table 1).\textsuperscript{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.
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, Várnaí 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-faceted 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 IncRNA 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, Jasín 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 master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*. 2013;153:307-319.

59. SCJ P, Sitzel 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.

60. Hnisz D, Abraham BJ, Lee TI, et al. Super-enhancers in the control of cell identity and disease. *Cell*. 2013;155:934-947.

61. van Geffen JP, Jiang T, Ruklisa D, Rosa IM, Tomé AR, Elding H, van Geffen JP, et al. Platelet function is modulated by genetic variants in common disease. *J Thromb Haemost*. 2017;15:268-274.

62. Yuan L, Chan GC, Beeler D, Janes L, Spokes KC, Dharaneeswaran P, Ruklisa D, Rosa IM, Tomé AR, Elding H, van Geffen JP, Jiang T, et al. Compound inheritance of a low-frequency regulatory SNP and a rare null mutation in a single nucleotide and a regulatory variant in the promoter of UBE2C reinforces PEAR1 enhancer activity. *Nat Commun*. 2016;7:10160.

63. Yuan L, Chan GC, Beeler D, Janes L, Spokes KC, Dharaneeswaran P, Ruklisa D, Rosa IM, Tomé AR, Elding H, van Geffen JP, Jiang T, et al. Compound inheritance of a low-frequency regulatory SNP and a rare null mutation in a single nucleotide and a regulatory variant in the promoter of UBE2C reinforces PEAR1 enhancer activity. *Nat Commun*. 2016;7:10160.

64. Aird WC. A role of stochastic phenotype switching in genetic variation in regulatory elements determines thrombosis and hemostasis phenotypes. *J Thromb Haemost*. 2022;20:1759-1765. doi: 10.1111/jth.15754

---

**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