Common and Rare 5′UTR Variants Altering Upstream Open Reading Frames in Cardiovascular Genomics

Omar Soukarieh[1]*, Caroline Meguerditchian[1], Carole Proust[1], Dylan Aïssi[1], Mélanie Eyries[2], Aurélie Goyenvalle[3] and David-Alexandre Trégouët[1]

[1] INSERM, Bordeaux Population Health, U1219, Molecular Epidemiology of Vascular and Brain Disorders, University of Bordeaux, Bordeaux, France, [2] Department of Genetics, Pitié-Salpêtrière Hospital, Assistance Publique-Hôpitaux de Paris, Sorbonne Université, Paris, France, [3] UVSQ, Inserm, END-ICAP, Université Paris-Saclay, Versailles, France

High-throughput sequencing (HTS) technologies are revolutionizing the research and molecular diagnosis landscape by allowing the exploration of millions of nucleotide sequences at an unprecedented scale. These technologies are of particular interest in the identification of genetic variations contributing to the risk of rare (Mendelian) and common (multifactorial) human diseases. So far, they have led to numerous successes in identifying rare disease-causing mutations in coding regions, but few in non-coding regions that include introns, untranslated (UTR), and intergenic regions. One class of neglected non-coding variations is that of 5′UTR variants that alter upstream open reading frames (upORFs) of the coding sequence (CDS) of a natural protein coding transcript. Following a brief summary of the molecular bases of the origin and functions of upORFs, we will first review known 5′UTR variations altering upORFs and causing rare cardiovascular disorders (CVDs). We will then investigate whether upORF-affecting single nucleotide polymorphisms could be good candidates for explaining association signals detected in the context of genome-wide association studies for common complex CVDs.

Keywords: open reading frame (ORF), genome wide association analysis (GWAS), Mendelian disease, non-coding mutations, polymorphism

INTRODUCTION

Upstream open reading frames (upORFs) are key regulatory elements located in the 5′untranslated (UTR) region of coding transcripts. UpORFs result from the presence of an upstream translation initiation site (uTIS) located within the 5′UTR and associated with an in-frame stop codon (uStop) located within the 5′UTR or the coding sequence (CDS). Different types of upORFs can be distinguished according to the position of the uStop with respect to the CDS (**Figure 1**). More precisely, when the uStop (i) is located within the 5′UTR, this results in a fully upstream ORF (uORF), (ii) is located within the CDS and is distinct from the main stop codon of the CDS, the uTIS is at the origin of an overlapping uORF (uoORF), and (iii) is the main stop codon of the CDS, this leads to an elongated CDS (eCDS). Approximately, half of the human transcripts naturally contain upORFs in their 5′UTR (1, 2) and these upORFs can contribute to modulate the production of the main protein encoded by the CDS by disturbing the translation initiation step and then the recognition of the main TIS by the ribosomes (3–5). The functional effect of a given upORF is highly variable and could be influenced by elements including the number of upORFs in the 5′UTR, their length, and the nucleotide context of the upORF as extensively discussed previously (6).

Often, the presence of upORF in general, and uoORF in particular, leads to a decrease of the expression of the main transcript (1). That could happen *via* the alteration of the translation mechanism (i.e., ribosome dissociation and ribosome stalling) or *via* transcript degradation by the non-sense-mediated decay process that recognizes the uStop as a premature stop codon (1, 7, 8). Nevertheless, under some conditions (i.e., hypoxia or cell stress), the presence of upORFs in a given transcript could be associated with an increase of the translation efficiency (9, 10). Indeed, upORF can modulate the activity of coexisting internal ribosome entry site (IRES) located on the same 5′UTR (11), thus regulating the IRES-dependent translation initiation in a context dependent manner. For instance, Chen et al., showed that the increase of fibroblst growth factor 9 (FGF9) protein levels under hypoxia happens *via* an IRES-dependent translation, regulated by the presence of a small upORF upstream to the IRES (12). In normal conditions (i.e., normaxia), FGF9 is present in low levels in human cells, thanks to the upORF-mediated translation inhibition of the CDS. Under hypoxia conditions, ribosomes probably switch from the upORF to IRES, thus activating the IRES-dependent translation and leading to efficient translation of FGF9 (12). That explains the increase of FGF9 under hypoxia conditions in cancer cells. In addition, upORFs could be translated into small-encoded peptides (SEPs) and play a regulatory role in health and disease contexts (13, 14).

High-throughput genomic studies have identified an increasingly number of single nucleotide variations (SNVs) located in 5′UTR and possibly altering upORFs by creating new ones or deleting/modifying existing ones suggesting that this kind of variants has been underestimated (15). Many of these variants have been characterized as disease causing by creating upORF and, thus, altering the production of the canonical protein, but surprisingly this has still not been investigated systematically. In fact, among the ~4,000 disease-associated 5′UTR variants reported in different databases, the most deleterious ones are those creating or deleting uTIS or uStop, responsible of the creation or the disruption of upORFs (15). Whiffin et al. have recently shown that, among all the SNVs reported in the genome aggregation (GnomAD) database to locate in 5′UTR of 18,593 canonical transcripts, on an average of 30 SNVs per gene are variations creating a uAUG canonical initiation codon (15). They also showed that only 39 uAUG-creating and four stop-removing extremely rare variants were reported in Human Gene Mutation Database (HGMD) or likely pathogenic in ClinVar (15). Very interestingly, among these rare variants, nine uAUG-creating variants are located in genes implicated in cystic fibrosis, familial hypercholesterolemia, and hematologic diseases (16–22). Moreover, recent studies have also shown that upORF could be initiated by non-AUG codons and be disease causing (23, 24). Given the diversity of the functional implication of existing upORFs in the regulation of protein expression, the possible functional impacts of upORF-altering variants, hereafter called upSNVs, on protein expression could be highly variable. Up to very recently (25, 26), this type of genetic variants was not easily predicted by available bioinformatics tools. In addition, their functional characterization requires dedicated experimental strategies that have not yet been harmonized in

order to demonstrate how they could affect gene expression and how the resulting dysregulations could lead to disease. Nevertheless, a first step in the assessment of the effect of upSNVs on the protein levels can be obtained using *in vitro* functional assays in which the 5′UTR and CDS of a given transcript are cloned in expression vectors followed by the expression of the produced vectors in human cells, both in the wild-type and upSNV contexts (27). upSNV-associated protein levels could then be evaluated by Western blot in comparison to the wild-type construct. Luciferase assays have also been widely used to study upSNVs. These assays are based on the cloning of the entire promoter of a given transcript before the coding sequence of a luciferase and the evaluation of the promoter activity in wild-type and mutant contexts *in vitro* by measuring the obtained luciferase luminescence normalized to a control vector. Additional methods used to characterize small ORFs and their potential translation into SEPs has been recently reviewed in (28). Altogether, upSNVs are still a neglected class of non-coding variations, and are often called as Variants of Unknown Significance when they are identified in routine clinical diagnosis, contributing then to medical wandering. In this work, with the aim of putting new light on upSNVs, we first provide a general overview of such type of variants known to cause rare cardiovascular disorders (CVDs). Then, we explore their potential role as candidates for explaining association signals detected in the context of genome-wide association studies (GWASs) for common complex CVDs.
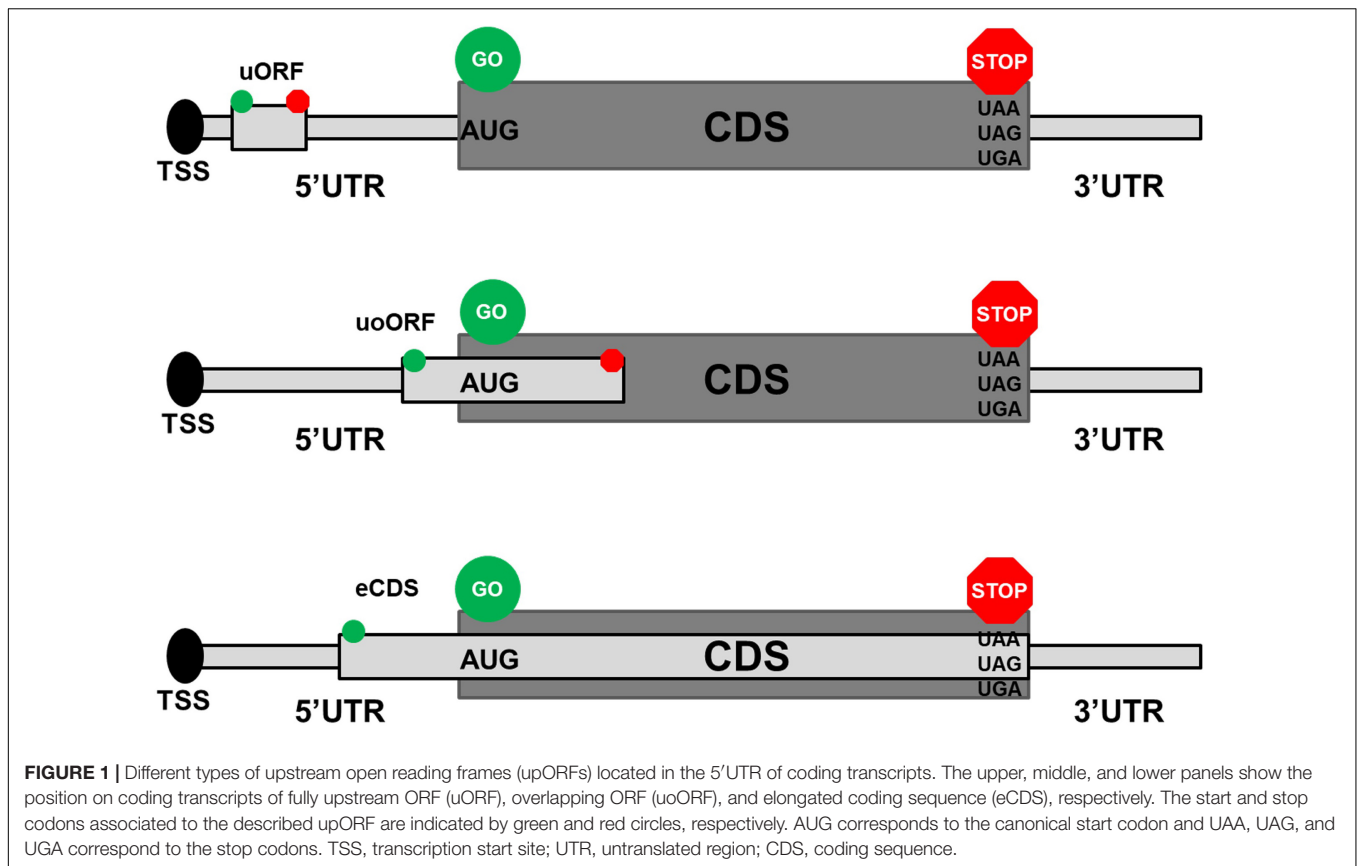
METHODS

Two complementary strategies were adopted to identify rare uAUG-creating variants in CVD genes. First, we selected variants from Supplementary Table 2 of (15) reporting upSNVs from ClinVar and HGMD. Then, we looked for additional variants in ClinVar and HGMD that were not reported in Whiffin et al., and scanned research articles in PubMed using the following keywords: "upstream ORF" and "cardio-vascular."

To investigate whether some association signals detected in GWAS for CVDs could be explained by upSNVs, we deployed MORFEE[1] on the 1,000 Genome reference dataset (phase 3-v20130502) in order to identify all the common (allele frequency > 1%) predicted upSNVs in 5′UTR regions. In a second step, we checked whether these predicted upSNVs could be in linkage disequilibrium (LD) with lead SNVs identified in GWAS studies for coronary artery disease (CAD), stroke, venous thrombosis (VT), platelets, and lipid traits. LD information was retrieved from the European populations genetic database available through the LDlink web-based tool[2] and from which we considered two SNVs to be in LD when the absolute value of their pairwise D' was greater than 0.7. For CAD, GWAS loci and lead SNVs were selected from Matsui et al. (29) and Hartiala et al. (30), while Malik et al. (31) and

---

[1]https://github.com/daissi/MORFEE

[2]https://ldlink.nci.nih.gov/?tab=home

FIGURE 1 | Different types of upstream open reading frames (upORFs) located in the 5′UTR of coding transcripts. The upper, middle, and lower panels show the position on coding transcripts of fully upstream ORF (uORF), overlapping ORF (uoORF), and elongated coding sequence (eCDS), respectively. The start and stop codons associated to the described upORF are indicated by green and red circles, respectively. AUG corresponds to the canonical start codon and UAA, UAG, and UGA correspond to the stop codons. TSS, transcription start site; UTR, untranslated region; CDS, coding sequence.

Lindström et al. (32) were used to identify GWAS loci and corresponding lead SNVs for stroke and VT, respectively. For platelets and lipids traits, we selected all the SNVs reported in the Geospatial Resource for Agriculture Species and Pests (GRASP) server[3] (33) as of September 2021 to associate at $p < 5.10^{-8}$ with any of their related quantitative traits, including mean platelet volume, platelet count, platelet aggregation or platelets' response to medication for platelet therapy, and high-density lipoprotein (HDL)-/low-density lipoprotein (LDL)/total cholesterol, triglycerides for lipids. Finally, this selection strategy led to a list of 749 CVD traits associated loci scrutinized for harboring common upSNVs.

RESULTS

Rare upSNVs Causing Cardiovascular Disorders

This section describes in detail upSNVs known to cause rare CVDs, most of which have been cataloged in (15). Information is summarized in **Table 1**.

HBB c.-29G>A appeared to be one of the first examples of uAUG-creating variants associated with tan inherited blood disorder, β-thalassemia characterized by marked reduce or absence of the beta-chain of hemoglobins (16). The created

uAUG generates a uoORF of 42 nucleotides in the NM_000518.5 transcript of the *HBB* and has been shown to be associated with an increased risk of β-Thalassemia (16). Moreover, Calvo and collaborators demonstrated that the c.-29G>A variant is associated with a decrease of the luciferase activity *in vitro*, suggesting that the presence of the uoORF could alter the levels of the main protein (1).

Disseminated bronchiectasis (DB) is characterized by abnormal dilation of bronchi associated with pulmonary dysfunction. A uAUG-creating variant in the 5′UTR of the *CFTR* gene (NM_000492.3:c.-34C>T) at the origin of a 108 nucleotide overlapping upORF has been described as associated with DB (19). This variant leads to a decrease of the luciferase activity in two different cell lines, in the context of two CFTR isoforms starting at positions c.-132 or c.-69. Moreover, the authors performed additional experiments *in vitro* confirming the recognition of the created uAUG by the ribosomes at the origin of a normal luciferase activity when the uAUG and its Kozak sequence were cloned in frame with the luciferase. These observations strongly support a role of the c.-34C>T variant on the reduction of the translation efficiency at the main ORF by the presence of the uoORF.

The Endoglin (*ENG*) gene is one of the main disease-causing genes for hereditary hemorrhagic telangiectasia (HHT), also known as Osler–Weber–Rendu syndrome, a rare vascular disorder causing abnormal vessel formation. *ENG* can be considered as a special gene with respect to upORFs. Indeed,

---

[3]https://grasp.nhlbi.nih.gov/Overview.aspx

four rare 5′UTR variants have been described so far in HHT patients to create uAUGs potentially at the origin of upORFs (18, 34–37, 77). These variants are NM_001114753.3: c.-142A>T, c.-127C>T, c.-10C>T, and c.-9G>A. Functional studies have been conducted for three of them (c.-142A>T, c.-127C>T, and c.-9G>A), bringing out an effect of the analyzed variants on the protein levels *in vitro* (18, 35–37, 77). Interestingly, a moderate decrease (∼20%) of the protein levels has been associated with c.-9G>A variant compared to a drastic reduction observed for c.-142A>T and c.-127C>T (∼60% and ∼75%, respectively). These studies also indicate that c.-142A>T and c.-127C>T variants are associated with severe phenotypes while patients carrying the c.-9G>A variant exhibited moderate HHT phenotype. At the molecular level, c.-142A>T, c.-127C>T, and c.-10C>T are predicted to be at the origin of uoORF (270, 255, and 138 nucleotides, respectively). The only exception holds for the c.-9G>A variant, that creates a uAUG in frame with the CDS, and generates an elongated CDS, probably at the origin of a longer form of the ENG protein carrying three additional amino acids. These molecular findings are in perfect concordance with clinical and familial data, suggesting that uoORF-creating variants in *ENG* are causative of a severe form of HHT. Among these four *ENG* variants, c.-9G>A, c.-10C>T, and c.-127C>T but not c.-142A>T are reported in public databases (ClinVar and HGMD). Interestingly, c.-10C>T and c.-127C>T are classified as likely pathogenic in ClinVar but the classification of the c.-9G>A is still conflicting. An additional uAUG-creating variant in the 5′UTR of *ENG* (c.-79C>T) at the origin of a 207 nucleotide uoORF. Of note, even though the c.-10C>T and c.-79C>T variants have not been evaluated in functional studies, one could speculate that in a similar way as for the c.-142A>T and to c.-127C>T variants, these variants would be associated with a reduction of the protein level.

Our group recently identified a disease-causing mutation in the 5′UTR of *PROS1* in an extended family affected with protein S deficiency (PSD) and familial thrombophilia (27). The identified variant was a never reported C>T substitution at c.-39 position creating a uAUG at the origin of an overlapping ORF of 156 nucleotides (NM_000313.4). Using *in vitro* assays, we demonstrated that this variant is associated with a total abolition of protein S levels. With the aim of restoring the main open reading frame in presence of the identified variant, we deleted one base pair at the new stop codon associated to the generated uoORF and, based on the detected protein weight by western blot, identified a protein probably starting at the c.-39C>T-created uAUG. This result indicated that the created uAUG could be used for translation and thus reduces or completely abolishes the translation rate at the main AUG, which explains null protein S level *in vitro* in presence of the variant.

Finally, three additional genes coding for proteins involved in CVDs have been highlighted in Whiffin et al. (15) from public databases as harboring rare uAUG-creating variants.

One is the *F8* gene coding for the coagulation factor VIII, a known susceptibility gene for venous thrombosis (38). The reported uAUG creating variant is the NM_000132.4:c.-5A>G variant that creates an overlapping upORF of 63 nucleotides

(20). Very interestingly, this variant is simultaneously predicted to modify a TAA stop codon into a TGA, in frame with two different non-canonical TIS (CTG) generating fully upstream upORFs of 39 and 123 nucleotides. upORFs ending with TGA have been shown to be associated with less translation efficiency of the main protein comparing to TAA ending ones (5). This variant was identified in a patient with mild FVIII activity, an observation compatible with an inhibitory effect on F8 expression of a variant associated with many upORFs. However, even if this variant is reported in HGMD database, its pathogenicity still needs to be validated.

The second gene is *HAMP*, coding for hepicidin whose increased plasma levels have recently been reported to associate with the risk of venous thrombosis (39). Whiffin et al., reported one rare variant in the 5′UTR of HAMP at the origin of a uAUG and catalogued in HGMD. While the *HAMP* variant has been described at the origin of an out of frame uoORF in (15) and described by Matthes and collaborators (17) as potentially generating an abnormal protein responsible for juvenile hereditary hemochromatosis, we did not find any stop codon in the transcript NM_021175.4 sequence that could be in frame with this created uAUG. Thus, this uAUG is unlikely at the origin of an ORF. Nonetheless, one cannot exclude a potential competition between the uAUG and the main TIS regarding the affinity of ribosomes. Indeed, no hepcidin was found in the urine of homozygous patient, suggesting that this variant could alter the translation of the main protein. As for *F8* c.-5A>G, experimental validation of its possible function impact on the translation of the associated protein is still needed.

The last cited gene is *LDLR* implicated in familial hypercholesterolemia associated with increased risk of cardiovascular diseases (40). The deletion of the cytosine at position c.-22 in the 5′UTR of the latest version of the *LDLR* transcript (NM_000527.5) has been identified in a homozygous form in an 8-year-old child diagnosed with familial hypercholesterolemia (22, 41). Interestingly, the c.-22delC is at the origin of an AUG generating an overlapping upORF of 174 nucleotides. This predicted effect could explain the potential pathogenicity of this variant and its association with familial hypercholesterolemia. Nonetheless, the impact of this variant on the LDLR levels still need to be evaluated.

Common upSNVs Associated With Cardiovascular Disorders and Their Quantitative Risk Factors

In this section, we report the few examples where common upSNVs were identified to be in LD with lead GWAS SNVs (**Table 2**).

F12 rs1801020 (NM_000505.4:c.-4C>T) and Venous Thrombosis

This variant is one of the most well-known and studied common upSNVs. It generates a very small overlapping ORF (nine nucleotides) and has been demonstrated in several independent studies to associate with decreased plasma levels of the clotting factor FXII (42–47). Calvo and colleagues have also demonstrated

TABLE 1 | Rare upSNVs in CVD-related diseases.

Gene (orientation)	cDNA position	Predicted effect	Disease	Databases	Classification (ClinVar)	References
HBB (−1)	NM_000518.5 c.-29G>A	uoORF (42 nts)	β-Thalassaemia	ClinVar	Pathogenic	1, 16
CFTR (1)	NM_000492.3 c.-34C>T	uoORF (108 nts)	Disseminated bronchiectasis	HGMD, ClinVar	Conflicting interpretations of pathogenicity	19
ENG (−1)	NM_001114753.3 c.-142A>T	uoORF (270 nts)	Hereditary Haemorrhagic TelangiectasiaT	NA	NA	77
ENG (−1)	NM_001114753.3 c.-127C>T	uoORF (255 nts)		HGMD	Pathogenic/Likely pathogenic	18, 35, 77
ENG (−1)	NM_001114753.3 c.-10C>T	uoORF (138 nts)		HGMD	Likely pathogenic	34
ENG (−1)	NM_001114753.3 c.-9G>A	eCDS (+ 3 nts)		HGMD	Conflicting interpretations	35
ENG (−1)	NM_001114753.3 c.-79C>T	uoORF (207 nts)	NA*	ClinVar	Uncertain significance	NA
PROS1 (−1)	NM_000313.4 c.-39C>T	uoORF (156 nts)	Protein S deficiency	NA	NA	27
F8 (−1)	NM_000132.4 c.-5A>G	uoORF (63 nts)	Hemophilia A	HGMD	NA	20
HAMP (1)	NM_021175.4 c.-25G>A	uAUG**	Juvenile Hereditary Hemochromatosis	HGMD	NA	15, 17
LDLR (1)	NM_000527.5 c.-22delC	uoORF (174 nts)	Familial Hypercholesterolaemia	ClinVar	Uncertain significance	22, 41

uoORF, upstream overlapping Open Reading Frame; eCDS, elongated coding sequence; nts, nucleotides; NA, non-available.
*This variant is reported in ClinVar without any clinical annotation (https://www.ncbi.nlm.nih.gov/clinvar/variation/618621/?new_evidence=false).
**No in frame stop codon predicted.

that this polymorphism is associated with a decrease of the protein levels *in vitro* (1) and that this decrease was due to the creation of the uoORF. While this variant has also been found (48, 49) associated with activated partial thromboplastin time, a biomarker for venous thrombosis, its impact on thrombosis risk is highly debated (1, 42, 47, 50, 51), especially as it never emerged from large-scale genetic association studies on arterial, cerebral, nor venous thrombosis. However, keeping in mind that the effect of a given upORF could be dependent on the cellular environment [e.g., hypoxia (12) and stress conditions (9)], it cannot be excluded that the rs1801020 could be associated with an increased prothrombotic state under certain environmental conditions that need to be further investigated.

FRMD5 rs492571 (NM_001286491.2:c.-487A>G) and Lipids

The rs492571 located in the 5′UTR of the *FRMD5* NM_001286491.2, at c.-487 position, is in nearly complete association ($r^2 > 0.80$, $D' \sim 1$) with several intronic SNPs reported to be associated with triglycerides and HDL-cholesterol levels (52). The A>G substitution at this position is predicted to create a new start codon and could generate a uORF of 39 nucleotides. These observations suggest that the rs492571 could be a good culprit candidate for the observed associations with lipids.

PEAR1 rs75699653 (NM_001353683.2:c.-491C>T) and Platelet Aggregation

PEAR1 was identified as one of the first GWAS loci for platelet aggregation (53) with the intronic rs12566888 (or any polymorphism in strong LD with it) as lead SNV. *PEAR1* harbors one upSNV, the rs75699653, in complete negative LD ($D' = -1$) with rs12566888. Because of the difference in their allele frequencies, the minor allele frequency of the former being ~0.02, that of rs12566888 being ~0.09, their pairwise LD r^2 is close to null. However, they generate three haplotypes where the rs75699653-T allele, predicted to be at the origin of a uORF

of 63 nucleotides, is always carried by the rs12566888-G allele (Supplementary Table 1). Interestingly, the rs1256688-G allele is either positively or negatively associated with platelet aggregation depending on how platelets are stimulated (54). Haplotype association analysis of these two SNVs in relation with platelet aggregation would be mandatory to determine if the original GWAS signal could be (partially) explained by the rs75699653-T carrying haplotype.

SLC18A1 rs58852338 (NM_001135691.3:c.-276G>A) and Triglycerides

SLC18A1 is one of the numerous loci associated with triglycerides levels (55). It harbors in its 5′UTR one upSNV, rs58852338, whose minor T allele (corresponding to c.-276A on the antisense transcript) with frequency ~1% is predicted to create a uORF of 36 nucleotides. The rs58852338-T allele is always carried by the haplotype carrying the rs55682243-C allele that was observed to associate with decreased triglycerides levels (55). This case is then similar to the *PEAR1's* discussed above.

Fibroblast Growth Factor 21 (FGF21) rs2231861 (NM_019113.4:c.-173C>G) and Triglycerides

FGF21 is another locus identified by GWAS as influencing triglycerides levels in plasma (56). The lead SNV is the synonymous rs838133 that does not show strong LD with any other SNVs when one uses the pairwise r^2 threshold of 0.80. However, it is in complete negative LD ($D' = -1$) with the rs2231861 upSNV. As a consequence, these two SNVs generate 3 haplotypes. As for the two previously described examples, the rare rs2231861-G allele predicted to create a uORF of 36 nucleotides is always carried by the haplotype harboring the rs838133-G allele associated with decreased triglycerides (56). Of note, the latter has also being found associated with decreased levels of homocysteine (57), another cardiovascular biomarker.

IL1F10 rs3811050 (NM_032556.6:c.-143C>T) and Coronary Artery Disease Risk

One common upSNV is present in the 5′UTR region of the *IL1F10* gene, a susceptibility locus for myocardial infarction (MI) (30). This is rs3811050 where the rs3811050-T is predicted to create an eCDS of 603 nucleotides while the canonical CDS is of 459 nucleotides. At *IL1F10*, the rs6761276-T allele of the missense p.Ile44Thr was found to be associated with increased risk of MI (30). According to the variant effect predictor (VEP) tool (58), the predicted pathogenicity of rs6761276 could be transcript-dependent[4]. It makes then sense to hypothesize that the impact on MI of the rs6761276-T allele may be different according whether or not it is present on the eCDS. As a consequence of their LD pattern (D' = 0.74, r^2 = 0.06), the rs6761276 and rs3811050 generate 4 haplotypes among which one (frequency ∼0.015) is carrying both the rs6761276-T risk allele and the eCDS rs3811050-T creating allele. It would be interesting to determine whether this specific rare haplotype is more at risk of MI than the haplotype carrying the rs6761276-T risk allele but not the eCDS creating allele.

ANGPTL4 rs35137994 (NM_139314.3:c.-140C>T), and Cardiovascular Traits

The *ANGPTL4* gene is an interesting locus for CVD as it has been shown to associate with several cardiovascular phenotypes, including CAD risk (59), lipid-related (56, 60), and red blood cells (61, 62) traits, with lead SNV being the missense rs116843064 (p.Glu40Lys) polymorphism. The minor rs116843064-A allele, with frequency ∼1% is associated with decreases in CAD risk, in triglycerides levels, in reticulocyte

---

[4]https://gnomad.broadinstitute.org/variant/2-113832312-T-C?dataset=gnomad_r2_1

counts and with increases in HDL levels, mean corpuscular volume and red cell distribution width. *ANGPTL4* harbors in its 5′UTR a common upSNV, rs35137994, whose minor T allele with frequency ∼5% and that is predicted to generate an eCDS of 1362 nucleotides. However, due to complete negative LD (D' = −1), the rs116843064-A and rs35137994-T alleles are never present on the same haplotype, indicating that the effect of the rs116843064-A allele cannot depend on whether it is present on the elongated isoform. As a consequence, the upSNV is unlikely to explain the observed GWAS signals, especially as the missense rs116843064 is predicted to be deleterious according to several standard prediction tools such as PolyPhen (probably damaging), SIFT (damaging), and CADD (score 31.0). That said, given these current observations, one cannot completely exclude that the rs35137994-T could exert additional independent and less pronounced effects on the aforementioned CVD traits.

PSORS1C1 (NM_014068.3: c.-199G>A and c.-94G>A) and Lipids

The last discussed GWAS locus is *PSORS1C1* that has been associated with plasma triglycerides levels (52) and hemoglobin levels (63). This locus has also been found associated with Psoriasis (64). *PSORS1C1* presents with two common upSNVs in its 5′UTR region, rs3131003 (c.-199G>A) and rs3815087 (c.-94G>A) with minor allele frequencies of ∼0.40 and ∼0.20, respectively. Individually, these two upSNVs could be at the origin of two uORFs of 183 and 78 nucleotides, respectively, both terminating at the same stop codon at c-19. However, because of complete positive LD (D' = 1), the rs3815087-A is always associated with the rs3131003-A allele, meaning that the predicted uORFs always exist together, with the 78 nucleotide length uORF always included in the longer one of 183 nucleotides. Whether this could result in one or two small

---

TABLE 2 | Common upSNVs in GWAS loci for CVDs and associated traits.

upSNV	Gene (orientation)	cDNA position	Genomic position (GRCh38.p13)	Predicted functional effect	GWAS lead SNPs	r^2/D' *	References
rs1801020	*F12* (−1)	NM_000505.4 c.-4C>T	chr5:177409531	ACG>ATG uoORF = 9 nts uSTOP = TGA	rs1801020	1.0/1	1
rs492571	*FRMD5* (−1)	NM_001286491.2 c.-487A>G	chr15:43919075	ATA>ATG uORF = 39 nts uSTOP = TAA	rs492571	1.0/1	52
rs75699653	*PEAR1* (1)	NM_001353683.2 c.-491C>T	chr1:156902203	ACG>ATG uORF = 63 nts uSTOP = TGA	rs12566888	(0.00/−1	53
rs58852338	SLC18A1 (−1)	NM_001135691.3 c.-276G (A	chr8:20181901	GTG>ATG uORF (36 nts uSTOP (TAG	rs55682243	(0.00/−1	55
rs2231861	*FGF21* (1)	NM_019113.4 c.-173C>G	chr19:48756064	ATC>ATG uORF = 36 nts uSTOP = TGA	rs838133	0.04/−1	56
rs3811050	*IL1F10* (1)	NM_032556.6 c.-143C>T	chr2:113072596	ACG>ATG eCDS = 603 nts uSTOP (TAG	rs6761276	0.06/0.74	30
rs35137994	ANGPTL4 (1)	NM_139314.3 c.-140C (T	chr19:8364182	ACG>ATG eCDS = 1362 nts uSTOP = TAG	rs116843064	∼0.00/−1	56, 59, 60
rs3131003 rs3815087	*PSORS1C1* (1)	NM_014068.3 c.-199G>A c.-94G>A	chr6:31125705 chr6:31125810	GTG>ATG uORF = 183 and 78 nts same uSTOP = TGA	rs3094205	0.61/0.97 0.03/0.26	52, 63

uoORF, upstream overlapping open reading frame, uORF, fully upstream open reading frame; nts, nucleotides; uStop, upstream stop codon, eCDS, elongated coding sequence.
*Pairwise linkage disequilibrium metrics (r^2, D') between upSNV and lead GWAS SNP.

peptides depends on the competition between the created uAUGs and remains to be elucidated. The rs3131003 is also in nearly complete positive LD ($D' = \sim 0.97$, $r^2 \sim 0.60$) with the rs3094205 lead SNV associated with triglycerides, suggesting that the former could be a good candidate for explaining the GWAS signal.

Of note, we did not observe any common upSNVs that exhibit strong LD with stroke- nor VT-associated lead SNVs and that could then explain the GWAS signals observed at their locus.

DISCUSSION/PERSPECTIVES

While there is increasingly awareness of the impact of rare upSNVs in rare Mendelian disorders, there has been so far little initiative to investigate the possible role of such variants in the susceptibility to common diseases and their quantitative risk factors. From a list of ∼700 loci identified in GWAS for CVD traits, we only identified a very minor proportion of loci (5: *FGF21, FRDM5, PEAR1, PSORS1C1,* and *SLC18A1*) where the GWAS signal could be partially explained by upSNVs. We focused here on CVDs but similar investigations merit to be conducted for other human diseases. Our results were based on *in silico* observations (bioinformatics predictions coupled to LD analyses) and deserve to be further investigated through fine-mapping association analysis and experimental molecular characterization. Several molecular techniques (gene reporter assays, toeprinting, polysome profiling, among others) are available to evaluate the effect of upSNVs on the translation machinery and/or protein expression. Here, we would like to highlight the recent advances in the antisense oligonucleotides (ASOs) strategy targeting upORFs, as it also offers therapeutic perspectives in the context of rare diseases. ASOs are very efficient molecular tools designed to modulate gene expression through Watson–Crick base pairing with specific motifs on target transcripts (65, 66). Initially, ASOs were used to downregulate gene expression or to modify RNA splicing. Recently, ASOs have been proposed to ameliorate gene expression by directly targeting uAUG (67). Liang et al., have shown that this technique depends on many factors on the RNA and on the chemical structure of the used ASOs (67). However, targeting upORF using ASOs seems to be a very innovative and efficient genetic tool to assess *in vitro* the functional impact of upSNVs on protein levels. Beyond their *in vitro* utility, effective ASOs capable of restoring protein levels could be used as a therapeutic approach to treat rare diseases caused by upSNVs. ASOs have indeed demonstrated great potential for treating rare diseases (68–71) due to coding or splice mutations. The antisense field has remarkably progressed over the last few years with the approval of several antisense drugs and with the development of even more potent compounds (72), opening promising perspectives to treat upORF-altering variants.

In this analytic review, we focused on SNVs known, or predicted, to create upORFs. We did not discuss molecular tools that are available to determine whether these upORFs could be at the origin of functional small micropeptides that could have specific physiological roles. This topic has recently been addressed in an independent review (28). Finally, we only examined in this work SNVs that could create uAUG resulting upORFs, the most known class of variants among those that affect non-canonical ORFs. Ribosome profiling data have shown the presence of small ORFs (sORFs) in coding transcripts outside the 5′UTR but also in non-coding RNAs (73, 74). Some of these sORFs have been shown to be translated into small encoded peptides and/or to have a regulatory role on gene expression (75, 76). Thus, one can easily speculate that genetic alterations in such sORFs could also have functional consequences and be involved in human diseases. The next steps would then be to characterize the spectrum of SNVs creating or deleting TIS or Stop in non-coding transcripts.

AUTHOR CONTRIBUTIONS

CM and DA developed and applied the MORFEE bioinformatics tool. OS and D-AT designed the study, conducted the systematic review, and drafted the manuscript. OS and CP performed *in silico* annotations of the predicted upORFs. AG and ME completed the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

1. Calvo SE, Pagliarini DJ, Mootha VK. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc Natl Acad Sci USA.* (2009) 106:7507–12. doi: 10.1073/pnas.0810916106

2. Ye Y, Liang Y, Yu Q, Hu L, Li H, Zhang Z, et al. Analysis of human upstream open reading frames and impact on gene expression. *Hum Genet.* (2015) 134:605–12. doi: 10.1007/s00439-015-1544-7

3. Hinnebusch AG, Ivanov IP, Sonenberg N. Translational control by 5′-untranslated regions of eukaryotic mRNAs. *Science.* (2016) 352:1413–6. doi: 10.1126/science.aad9868

4. Silva J, Fernandes R, Romão L. Translational regulation by upstream open reading frames and human diseases. *Adv Exp Med Biol.* (2019) 1157:99–116. doi: 10.1007/978-3-030-19966-1_5

5. Giess A, Torres Cleuren YN, Tjeldnes H, Krause M, Bizuayehu TT, Hiensch S, et al. Profiling of small ribosomal subunits reveals modes and regulation of translation initiation. *Cell Rep.* (2020) 31:107534. doi: 10.1016/j.celrep.2020.107534

6. Wethmar K. The regulatory potential of upstream open reading frames in eukaryotic gene expression. *Wiley Interdiscip Rev RNA.* (2014) 5:765–78. doi: 10.1002/wrna.1245

7. Mendell JT, Sharifi NA, Meyers JL, Martinez-Murillo F, Dietz HC. Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. *Nat Genet.* (2004) 36:1073–8. doi: 10.1038/ng1429

8. Barbosa C, Peixeiro I, Romão L. Gene expression regulation by upstream open reading frames and human disease. *PLoS Genet.* (2013) 9:e1003529. doi: 10.1371/journal.pgen.1003529

9. Vattem KM, Wek RC. Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. *Proc Natl Acad Sci USA.* (2004) 101:11269–74. doi: 10.1073/pnas.0400541101

10. Spriggs KA, Bushell M, Willis AE. Translational regulation of gene expression during conditions of cell stress. *Mol Cell.* (2010) 40:228–37. doi: 10.1016/j.molcel.2010.09.028

11. Lacerda R, Menezes J, Romão L. More than just scanning: the importance of cap-independent mRNA translation initiation for cellular stress response and cancer. *Cell Mol Life Sci.* (2017) 74:1659–80. doi: 10.1007/s00018-016-2428-2

12. Chen T-M, Shih Y-H, Tseng JT, Lai M-C, Wu C-H, Li Y-H, et al. Overexpression of FGF9 in colon cancer cells is mediated by hypoxia-induced translational activation. *Nucleic Acids Res.* (2014) 42:2932–44. doi: 10.1093/nar/gkt1286

13. Young SK, Willy JA, Wu C, Sachs MS, Wek RC. Ribosome reinitiation directs gene-specific translation and regulates the integrated stress response. *J Biol Chem.* (2015) 290:28257–71. doi: 10.1074/jbc.M115.693184

14. Rodriguez CM, Chun SY, Mills RE, Todd PK. Translation of upstream open reading frames in a model of neuronal differentiation. *BMC Genomics.* (2019) 20:391. doi: 10.1186/s12864-019-5775-1

15. Whiffin N, Karczewski KJ, Zhang X, Chothani S, Smith MJ, Evans DG, et al. Characterising the loss-of-function impact of 5′ untranslated region variants in 15,708 individuals. *Nat Commun.* (2020) 11:2523. doi: 10.1038/s41467-019-10717-9

16. Oner R, Agarwal S, Dimovski AJ, Efremov GD, Petkov GH, Altay C, et al. The G—-A mutation at position +22 3′ to the Cap site of the beta-globin gene as a possible cause for a beta-thalassemia. *Hemoglobin.* (1991) 15:67–76. doi: 10.3109/03630269109072485

17. Matthes T, Aguilar-Martinez P, Pizzi-Bosman L, Darbellay R, Rubbia-Brandt L, Giostra E, et al. Severe hemochromatosis in a Portuguese family associated with a new mutation in the 5′-UTR of the HAMP gene. *Blood.* (2004) 104:2181–3. doi: 10.1182/blood-2004-01-0332

18. Kim M-J, Kim S-T, Lee H-D, Lee K-Y, Seo J, Lee J-B, et al. Clinical and genetic analyses of three Korean families with hereditary hemorrhagic telangiectasia. *BMC Med Genet.* (2011) 12:130. doi: 10.1186/1471-2350-12-130

19. Lukowski SW, Bombieri C, Trezise AEO. Disrupted post-transcriptional regulation of the cystic fibrosis transmembrane conductance regulator (CFTR) by a 5′UTR mutation is associated with a CFTR-related disease. *Hum Mutat.* (2011) 32:E2266–82. doi: 10.1002/humu.21545

20. Debeljak M, Kitanovski L, Trampuš Bakija A, Benedik Dolnièar M. Spectrum of F8 gene mutations in haemophilia A patients from Slovenia. *Haemophilia.* (2012) 18:e420–3. doi: 10.1111/hae.12003

21. Tichý L, Freiberger T, Zapletalová P, Soška V, Ravèuková B, Fajkusová L. The molecular basis of familial hypercholesterolemia in the Czech Republic: spectrum of LDLR mutations and genotype-phenotype correlations. *Atherosclerosis.* (2012) 223:401–8. doi: 10.1016/j.atherosclerosis.2012.05.014

22. Khamis A, Palmen J, Lench N, Taylor A, Badmus E, Leigh S, et al. Functional analysis of four LDLR 5′UTR and promoter variants in patients with familial hypercholesterolaemia. *Eur J Hum Genet.* (2015) 23:790–5. doi: 10.1038/ejhg.2014.199

23. Sendoel A, Dunn JG, Rodriguez EH, Naik S, Gomez NC, Hurwitz B, et al. Translation from unconventional 5′ start sites drives tumour initiation. *Nature.* (2017) 541:494–9. doi: 10.1038/nature21036

24. Kearse MG, Goldman DH, Choi J, Nwaezeapu C, Liang D, Green KM, et al. Ribosome queuing enables non-AUG translation to be resistant to multiple protein synthesis inhibitors. *Genes Dev.* (2019) 33:871–85. doi: 10.1101/gad.324715.119

25. Aïssi D, Soukarieh O, Proust C, Jaspard-Vinassa B, Fautrad P, Ibrahim-Kosta M, et al. MORFEE: a new tool for detecting and annotating single nucleotide variants creating premature ATG codons from VCF files. *bioRxiv* [Preprint]. (2020). doi: 10.1101/2020.03.29.012054

26. Zhang X, Wakeling M, Ware J, Whiffin N. Annotating high-impact 5′untranslated region variants with the UTRannotator. *Bioinformatics.* (2020) 37:1171–3. doi: 10.1093/bioinformatics/btaa783

27. Labrouche-Colomer S, Soukarieh O, Proust C, Mouton C, Huguenin Y, Roux M, et al. 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. doi: 10.1042/CS20200403

28. Kute PM, Soukarieh O, Tjeldnes H, Trégouët D-A, Valen E. Small open reading frames, how to find them and determine their function. *Front Genet.* (2021) 12:796060. doi: 10.3389/fgene.2021.796060

29. Matsui H, Tatezaki S, Tsuji H, Ochiai H. Isolation and characterization of low- and high-metastatic clones from murine RCT (radiological, Chiba, and Toyama) sarcoma. *J Cancer Res Clin Oncol.* (1989) 115:9–16. doi: 10.1007/BF00391593

30. Hartiala JA, Han Y, Jia Q, Hilser JR, Huang P, Gukasyan J, et al. Genome-wide analysis identifies novel susceptibility loci for myocardial infarction. *Eur Heart J.* (2021) 42:919–33. doi: 10.1093/eurheartj/ehaa1040

31. Malik R, Chauhan G, Traylor M, Sargurupremraj M, Okada Y, Mishra A, et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet.* (2018) 50:524–37. doi: 10.1038/s41588-018-0058-3

32. Lindström S, Wang L, Smith EN, Gordon W, van Hylckama Vlieg A, de Andrade M, et al. Genomic and transcriptomic association studies identify 16 novel susceptibility loci for venous thromboembolism. *Blood.* (2019) 134:1645–57. doi: 10.1182/blood.2019000435

33. Leslie R, O'Donnell CJ, Johnson AD. GRASP: analysis of genotype-phenotype results from 1390 genome-wide association studies and corresponding open access database. *Bioinformatics.* (2014) 30:i185–94. doi: 10.1093/bioinformatics/btu273

34. Bossler AD, Richards J, George C, Godmilow L, Ganguly A. Novel mutations in ENG and ACVRL1 identified in a series of 200 individuals undergoing clinical genetic testing for hereditary hemorrhagic telangiectasia (HHT): correlation of genotype with phenotype. *Hum Mutat.* (2006) 27:667–75. doi: 10.1002/humu.20342

35. Damjanovich K, Langa C, Blanco FJ, McDonald J, Botella LM, Bernabeu C, et al. 5′UTR mutations of ENG cause hereditary hemorrhagic telangiectasia. *Orphanet J Rare Dis.* (2011) 6:85. doi: 10.1186/1750-1172-6-85

36. Albiñana V, Zafra MP, Colau J, Zarrabeitia R, Recio-Poveda L, Olavarrieta L, et al. Mutation affecting the proximal promoter of endoglin as the origin of hereditary hemorrhagic telangiectasia type 1. *BMC Med Genet.* (2017) 18:20. doi: 10.1186/s12881-017-0380-0

37. Ruiz-Llorente L, Gallardo-Vara E, Rossi E, Smadja DM, Botella LM, Bernabeu C. Endoglin and alk1 as therapeutic targets for hereditary hemorrhagic telangiectasia. *Expert Opin Ther Targets.* (2017) 21:933–47. doi: 10.1080/14728222.2017.1365839

38. Hinds DA, Buil A, Ziemek D, Martinez-Perez A, Malik R, Folkersen L, et al. Genome-wide association analysis of self-reported events in 6135 individuals and 252 827 controls identifies 8 loci associated with thrombosis. *Hum Mol Genet.* (2016) 25:1867–74. doi: 10.1093/hmg/ddw037

39. Ellingsen TS, Lappegård J, Ueland T, Aukrust P, Brækkan SK, Hansen J-B. Plasma hepcidin is associated with future risk of venous thromboembolism. *Blood Adv.* (2018) 2:1191–7. doi: 10.1182/bloodadvances.2018018465

40. Sharifi M, Rakhit RD, Humphries SE, Nair D. Cardiovascular risk stratification in familial hypercholesterolaemia. *Heart.* (2016) 102:1003–8. doi: 10.1136/heartjnl-2015-308845

41. Sözen MM, Rakhit R, Oner C, Tokatli A, Kalkanoǧlu HS, Dursun A, et al. The molecular basis of familial hypercholesterolaemia in Turkish patients. *Atherosclerosis.* (2005) 180:63–71. doi: 10.1016/j.atherosclerosis.2004.12.042

42. Kanaji T, Okamura T, Osaki K, Kuroiwa M, Shimoda K, Hamasaki N, et al. A common genetic polymorphism (46 C to T substitution) in the 5′-untranslated region of the coagulation factor XII gene is associated with low translation efficiency and decrease in plasma factor XII level. *Blood.* (1998) 91:2010–4.

43. Endler G, Mannhalter C, Sunder-Plassmann H, Lalouschek W, Kapiotis S, Exner M, et al. Homozygosity for the C–>T polymorphism at nucleotide 46 in the 5′ untranslated region of the factor XII gene protects from development of acute coronary syndrome. *Br J Haematol.* (2001) 115:1007–9. doi: 10.1046/j.1365-2141.2001.03201.x

44. Endler G, Exner M, Mannhalter C, Meier S, Ruzicka K, Handler S, et al. A common C–>T polymorphism at nt 46 in the promoter region of coagulation factor XII is associated with decreased factor XII activity. *Thromb Res.* (2001) 101:255–60. doi: 10.1016/s0049-3848(00)00404-7

45. Tirado I, Soria JM, Mateo J, Oliver A, Souto JC, Santamaria A, et al. Association after linkage analysis indicates that homozygosity for the 46C–>T polymorphism in the F12 gene is a genetic risk factor for venous thrombosis. *Thromb Haemost.* (2004) 91:899–904. doi: 10.1160/TH03-10-0620

46. Bertina RM, Poort SR, Vos HL, Rosendaal FR. The 46C–>T polymorphism in the factor XII gene (F12) and the risk of venous thrombosis. *J Thromb Haemost.* (2005) 3:597–9. doi: 10.1111/j.1538-7836.2005.01198.x

47. Bach J, Endler G, Winkelmann BR, Boehm BO, Maerz W, Mannhalter C, et al. Coagulation factor XII (FXII) activity, activated FXII, distribution of FXII C46T gene polymorphism and coronary risk. *J Thromb Haemost.* (2008) 6:291–6. doi: 10.1111/j.1538-7836.2007.02839.x

48. Tang W, Schwienbacher C, Lopez LM, Ben-Shlomo Y, Oudot-Mellakh T, Johnson AD, et al. Genetic associations for activated partial thromboplastin time and prothrombin time, their gene expression profiles, and risk of coronary artery disease. *Am J Hum Genet.* (2012) 91:152–62. doi: 10.1016/j.ajhg.2012.05.009

49. Kanai M, Akiyama M, Takahashi A, Matoba N, Momozawa Y, Ikeda M, et al. Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat Genet.* (2018) 50:390–400. doi: 10.1038/s41588-018-0047-6

50. Bersano A, Ballabio E, Bresolin N, Candelise L. Genetic polymorphisms for the study of multifactorial stroke. *Hum Mutat.* (2008) 29:776–95. doi: 10.1002/humu.20666

51. Johnson CY, Tuite A, Morange PE, Tregouet DA, Gagnon F. The factor XII -4C>T variant and risk of common thrombotic disorders: a huge review and meta-analysis of evidence from observational studies. *Am J Epidemiol.* (2011) 173:136–44. doi: 10.1093/aje/kwq349

52. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature.* (2010) 466:707–13. doi: 10.1038/nature09270

53. Johnson AD, Yanek LR, Chen M-H, Faraday N, Larson MG, Tofler G, et al. Genome-wide meta-analyses identifies seven loci associated with platelet aggregation in response to agonists. *Nat Genet.* (2010) 42:608–13. doi: 10.1038/ng.604

54. Keramati AR, Chen MH, Rodriguez BAT, Yanek LR, Bhan A, Gaynor BJ, et al. Genome sequencing unveils a regulatory landscape of platelet reactivity. *Nat Commun.* (2021) 12:3626. doi: 10.1038/s41467-021-23470-9

55. Sinnott-Armstrong N, Tanigawa Y, Amar D, Mars N, Benner C, Aguirre M, et al. Genetics of 35 blood and urine biomarkers in the UK biobank. *Nat Genet.* (2021) 53:185–94. doi: 10.1038/s41588-020-00757-z

56. Richardson TG, Sanderson E, Palmer TM, Ala-Korpela M, Ference BA, Davey Smith G, et al. Evaluating the relationship between circulating lipoprotein lipids and apolipoproteins with risk of coronary heart disease: a multivariable mendelian randomisation analysis. *PLoS Med.* (2020) 17:e1003062. doi: 10.1371/journal.pmed.1003062

57. van Meurs JBJ, Pare G, Schwartz SM, Hazra A, Tanaka T, Vermeulen SH, et al. Common genetic loci influencing plasma homocysteine concentrations and their effect on risk of coronary artery disease. *Am J Clin Nutr.* (2013) 98:668–76. doi: 10.3945/ajcn.112.044545

58. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The ensembl variant effect predictor. *Genome Biol.* (2016) 17:122. doi: 10.1186/s13059-016-0974-4

59. Moody RP, Benoit FM, Riedel D, Ritter L. Dermal absorption of the insect repellent DEET (N,N-diethyl-m-toluamide) in rats and monkeys: effect of anatomical site and multiple exposure. *J Toxicol Environ Health.* (1989) 26:137–47. doi: 10.1080/15287398909531240

60. van Leeuwen EM, Sabo A, Bis JC, Huffman JE, Manichaikul A, Smith AV, et al. Meta-analysis of 49 549 individuals imputed with the 1000 genomes project reveals an exonic damaging variant in ANGPTL4 determining fasting TG levels. *J Med Genet.* (2016) 53:441–9. doi: 10.1136/jmedgenet-2015-103439

61. Chen M-H, Raffield LM, Mousas A, Sakaue S, Huffman JE, Moscati A, et al. Trans-ethnic and ancestry-specific blood-cell genetics in 746,667 individuals from 5 global populations. *Cell.* (2020) 182:1198–213.e14. doi: 10.1016/j.cell.2020.06.045

62. Vuckovic D, Bao EL, Akbari P, Lareau CA, Mousas A, Jiang T, et al. The polygenic and monogenic basis of blood traits and diseases. *Cell.* (2020) 182:1214–31.e11. doi: 10.1016/j.cell.2020.08.008

63. Oskarsson GR, Oddsson A, Magnusson MK, Kristjansson RP, Halldorsson GH, Ferkingstad E, et al. Predicted loss and gain of function mutations in ACO1 are associated with erythropoiesis. *Commun Biol.* (2020) 3:189. doi: 10.1038/s42003-020-0921-5

64. Lee K-Y, Leung K-S, Tang NLS, Wong M-H. Discovering genetic factors for psoriasis through exhaustively searching for significant second order SNP-SNP interactions. *Sci Rep.* (2018) 8:15186. doi: 10.1038/s41598-018-33493-w

65. Bizot F, Vulin A, Goyenvalle A. Current status of antisense oligonucleotide-based therapy in neuromuscular disorders. *Drugs.* (2020) 80:1397–415. doi: 10.1007/s40265-020-01363-3

66. Ferlini A, Goyenvalle A, Muntoni F. RNA-targeted drugs for neuromuscular diseases. *Science.* (2021) 371:29–31. doi: 10.1126/science.aba4515

67. Liang W-C, Wong C-W, Liang P-P, Shi M, Cao Y, Rao S-T, et al. Translation of the circular RNA circβ-catenin promotes liver cancer cell growth through activation of the WNT pathway. *Genome Biol.* (2019) 20:84. doi: 10.1186/s13059-019-1685-x

68. Goyenvalle A, Griffith G, Babbs A, El Andaloussi S, Ezzat K, Avril A, et al. Functional correction in mouse models of muscular dystrophy using exon-skipping tricyclo-DNA oligomers. *Nat Med.* (2015) 21:270–5. doi: 10.1038/nm.3765

69. Goyenvalle A, Leumann C, Garcia L. Therapeutic potential of tricyclo-DNA antisense oligonucleotides. *J Neuromuscul Dis.* (2016) 3:157–67. doi: 10.3233/JND-160146

70. Robin V, Griffith G, Carter J-PL, Leumann CJ, Garcia L, Goyenvalle A. Efficient SMN rescue following subcutaneous tricyclo-DNA antisense oligonucleotide treatment. *Mol Ther Nucleic Acids.* (2017) 7:81–9. doi: 10.1016/j.omtn.2017.02.009

71. Relizani K, Echevarría L, Zarrouki F, Gastaldi C, Dambrune C, Aupy P, et al. Palmitic acid conjugation enhances potency of tricyclo-DNA splice switching oligonucleotides. *Nucleic Acids Res.* (2021) 50:17–34. doi: 10.1093/nar/gkab1199

72. Roberts TC, Langer R, Wood MJA. Advances in oligonucleotide drug delivery. *Nat Rev Drug Discov.* (2020) 19:673–94. doi: 10.1038/s41573-020-0075-7

73. van Heesch S, Witte F, Schneider-Lunitz V, Schulz JF, Adami E, Faber AB, et al. The translational landscape of the human heart. *Cell.* (2019) 178:242–60.e29. doi: 10.1016/j.cell.2019.05.010

74. Yang Y, Gao H, Zhou H, Liu Q, Qi Z, Zhang Y, et al. The role of mitochondria-derived peptides in cardiovascular disease: recent updates. *Biomed Pharmacother.* (2019) 117:109075. doi: 10.1016/j.biopha.2019.109075

75. Couso J-P, Patraquim P. Classification and function of small open reading frames. *Nat Rev Mol Cell Biol.* (2017) 18:575–89. doi: 10.1038/nrm.2017.58

76. Wu Q, Wright M, Gogol MM, Bradford WD, Zhang N, Bazzini AA. Translation of small downstream ORFs enhances translation of canonical main open reading frames. *EMBO J.* (2020) 39:e104763. doi: 10.15252/embj. 2020104763

77. Ruiz-Llorente L, McDonald J, Wooderchak-Donahue W, Briggs E, Chesnutt M, Bayrak-Toydemir P, et al. Characterization of a family mutation in the 5′ untranslated region of the endoglin gene causative of hereditary hemorrhagic telangiectasia. *J Hum Genet.* (2019) 64:333–9. doi: 10.1038/s10038-019-05 64-x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.