



# Into the Wild: GWAS Exploration of Non-coding RNAs

Hector Giral<sup>1,2</sup>, Ulf Landmesser<sup>1,2,3</sup> and Adelheid Kratzer<sup>1,2\*</sup>

<sup>1</sup> Department of Cardiology, Charité - Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany, <sup>2</sup> DZHK (German Centre for Cardiovascular Research), Partner Site Berlin, Berlin, Germany, <sup>3</sup> Berlin Institute of Health (BIH), Berlin, Germany

## OPEN ACCESS

### Edited by:

Jeanette Erdmann,  
Universität zu Lübeck, Germany

### Reviewed by:

Baiba Vilne,  
Technische Universität München,  
Germany  
Thorsten Kessler,  
Deutsches Herzzentrum München,  
Germany  
Arne S. Schaefer,  
Charité Universitätsmedizin Berlin,  
Germany

### \*Correspondence:

Adelheid Kratzer  
adelheid.kratzer@charite.de

### Specialty section:

This article was submitted to  
Cardiovascular Genetics and Systems  
Medicine,  
a section of the journal  
Frontiers in Cardiovascular Medicine

**Received:** 05 July 2018

**Accepted:** 03 December 2018

**Published:** 17 December 2018

### Citation:

Giral H, Landmesser U and Kratzer A  
(2018) Into the Wild: GWAS  
Exploration of Non-coding RNAs.  
*Front. Cardiovasc. Med.* 5:181.  
doi: 10.3389/fcvm.2018.00181

Genome-wide association studies (GWAS) have proven a fundamental tool to identify common variants associated to complex traits, thus contributing to unveil the genetic components of human disease. Besides, the advent of GWAS contributed to expose unexpected findings that urged to redefine the framework of population genetics. First, loci identified by GWAS had small effect sizes and could only explain a fraction of the predicted heritability of the traits under study. Second, the majority of GWAS hits mapped within non-coding regions (such as intergenic or intronic regions) where new functional RNA species (such as lncRNAs or circRNAs) have started to emerge. Bigger cohorts, meta-analysis and technical improvements in genotyping allowed identification of an increased number of genetic variants associated to coronary artery disease (CAD) and cardiometabolic traits. The challenge remains to infer causal mechanisms by which these variants influence cardiovascular disease development. A tendency to assign potential causal variants preferentially to coding genes close to lead variants contributed to disregard the role of non-coding elements. In recent years, in parallel to an increased knowledge of the non-coding genome, new studies started to characterize disease-associated variants located within non-coding RNA regions. The upcoming of databases integrating single-nucleotide polymorphisms (SNPs) and non-coding RNAs together with novel technologies will hopefully facilitate the discovery of causal non-coding variants associated to disease. This review attempts to summarize the current knowledge of genetic variation within non-coding regions with a focus on long non-coding RNAs that have widespread impact in cardiometabolic diseases.

**Keywords:** lncRNA, genetic variant, GWAS, coronary artery disease, cardiometabolic disorders

In the dawn of the millennium, the first draft of the human genome represented a major milestone in the path to decipher the genetic component of human disease. Further refinement of the human genome by the 1,000 Genomes Project mapped over 88 million variants from 26 populations where ~20 million correspond to common (frequency >0.5%) single-nucleotide polymorphisms (SNPs), a coverage of >95% of all estimated human common SNPs (1, 2). Other consortia such as Encyclopedia of DNA Elements (ENCODE) (3, 4) and Functional Annotation of the Mammalian Genome (FANTOM) (5) contributed to the generation of a detailed atlas of DNA functional elements and transcriptional units uncovering that more than 80–90% of the human genome is transcribed and display some functionality (4). In this context, Genome-wide association studies (GWAS) emerged as a fundamental tool to define single nucleotide polymorphisms (SNPs) associated to complex human traits or diseases (6–10). With regard to cardiovascular disease, GWAS studies identified up to 161 genetic risk loci associated to coronary artery disease (CAD) (11–13).

Despite the profound contributions of GWAS to the understanding of human disease pathophysiology, some issues forced to redefine the framework of GWAS studies. First, most significant GWAS hits could only explain a small fraction of genetic variance for a specific trait (14). In the case of CAD, all 161 genome-wide significant loci account for 15.1% of the predicted genetic contribution to the disease (15), which is strikingly similar to the percentage of gene sets (13.9%) or gene networks (14%) implicated on these 161 CAD-associated loci (12). An emerging notion, known as omnigenic model, states that cell regulatory networks are so deeply connected that basically all genes expressed in disease-relevant cell types conspire to influence the heritability of complex traits (16). Therefore, this model assumes that thousands of loci with small size effects contribute to the overall heritability of the trait or disease by affecting the expression of a smaller set of core genes (16). It seems that the common disease-common variant (CD-CV) model that drove the first decade of GWAS studies is shifting to a complex trait-complex genetics (CT-CG) scenario, where a handful of relevant variants cannot fully explain genetic variation in whole populations. The overall notion of a widespread dispersion of genetic contributions to disease due to the interconnectivity of biological systems seems to be widely accepted. On the other hand, the concept of a set of core genes driving the phenotype of complex diseases is still controversial and as a result the choice of methodology to address the future of the field (17).

Nearly 90% of all phenotype-associated SNPs identified by GWAS lied within non-coding regions (18–20), which includes a broad spectrum of locations including intronic or promoter regions, small ncRNAs such as miRNAs, long ncRNAs, antisense, and enhancer or insulator regions. Most non-coding variants are concentrated in deoxyribonuclease I (DNase I) hypersensitive sites that label regions with increased chromatin accessibility. Currently, around 2,500 miRNAs and more than 50,000 lncRNAs have been annotated in the human genome, practically doubling the number of protein coding transcripts, highlighting the important role of this part of the genome (21).

This review summarizes genetic variations within lncRNAs associated to cardiovascular disease (CAD, MI) and to various cardiometabolic risk factors for cardiovascular disease such as lipoprotein metabolism, diabetes or hypertension (Table 1).

## IMPACT OF GENETIC VARIANTS ON LNCRNAs FUNCTIONALITY

One of the longest-standing challenges in human genetics is to assign potential causality within a locus to every variant in close linkage disequilibrium (LD) with the lead variant (34). Despite the potential of lncRNAs as causal factors of disease, GWAS studies had a tendency to explore genetic variant causality preferentially in coding genes, mostly due to our limited knowledge of ncRNAs genomic structure and functionality. Additionally, lncRNAs overlapping coding genes (such as antisense and intronic lncRNAs) are harder to dissociate from neighboring coding genes when searching for potential

causal variants compared to intergenic lncRNA (lincRNA) which do not overlap coding genes. Fortunately, interactive lncRNA databases (LincSNP2.0) (35) together with established GWAS catalogs like NHGRI-EBI (36) and GWASdb.v2 (37) have started to integrate newly identified lncRNAs transcripts and disease-associated genetic variants. The latest databases mapped 371,647 disease-associated SNPs to lncRNA what accounts for approximately 45% of all disease-associated human SNPs identified (35).

Recent approaches focused on lincRNAs by further exploring loci previously associated to CAD (32, 38–41). For example, a class-level testing framework, termed Genetic Class Association Testing (GenCAT) allowed the identification of new trait-associated variants within multiple lincRNAs contributing novel insights into their role in cardiometabolic pathophysiology (42). GenCAT approach includes SNPs directly within the lincRNA but also the ones 500 kb up- or downstream of the lincRNA (38).

In a functional perspective, many lncRNAs reside in the nucleus conducting key regulatory steps in gene transcription, transcript splicing or chromatin structure. Cytoplasmic lncRNAs affect cell homeostasis by modulating translation and stability of mRNA through scaffolding multi-protein complexes that accomplish these functions (43). Several lncRNA functions depend on structural domains that generate binding sites to interact with RNA binding proteins (RBPs) acting as scaffolds for recruitment of proteins, RNA molecules and DNA elements (44–46). Some genetic variants are predicted to impact lncRNA secondary structure and thereby lncRNA-RBP interactions which can dramatically affect their functionality. Low evolutionary conservation of lncRNAs constitutes a challenge to predict structural domains and consequently how genetic variants induce functional modifications (47). Moreover, analysis of variation frequencies suggested that functional elements in lncRNAs have a much lower variation frequency almost comparable to protein-coding exons (48). Alternative splicing is an additional mechanism to generate functional diversity of lncRNAs by differential arrangement of structural domains (19).

Furthermore, SNPs may affect lncRNA transcriptional expression by altering its promoter region but also may influence expression of proximal or distal protein coding genes through the action of enhancers (19). Modulation of distant genes by trans-regulation is mediated by lncRNAs-enhancers but the effect of induced chromatin structural changes must be also considered. Chromatin structural loops link regulatory enhancer elements to distant gene promoters and variants disrupting this process broadly influence gene expression (49). Distal regulatory elements (DRE) can regulate the transcription of lincRNA through chromatin interactions, which can be influenced by GWAS-identified SNPs and define disease association (50).

## LONG NON-CODING RNAs ASSOCIATED TO CARDIOMETABOLIC TRAITS

The first examples of SNP variants associated to increased risk of CAD located within a lncRNA were identified in the locus

**TABLE 1** | List of representative lncRNA and variants identified by GWAS associated to cardiometabolic disease traits.

lncRNA genecode ID (Aliases) <sup>a</sup>	Cytogenetic region	Genes in locus	SNP variant <sup>b</sup>	Disease or phenotype traits	NHGRI Catalog <sup>c</sup>	Other sources <sup>d</sup>	GTEx eQTLs <sup>e</sup>
<b>CDKN2B-AS1</b> ENSG00000240498 (ANRIL)	9p21.3 3.4kb 3' CDKN2B	CDKN2A, CDKN2B, DMRTA1	rs2891168 (intronic)	myocardial infarction  CAD	GCST003117 (13) 5 × 10 <sup>-75</sup> GCST005194 (12) 5 × 10 <sup>-204</sup>		none
<b>MIAT</b> ENSG00000225783 (lnc-CRYBA4)	9p21.3p 13kb 3' CDKN2B	CDKN2A, CDKN2B, DMRTA1	rs4977574 (intronic)  rs10811661 (intergenic)	CAD  Type 2 Diabetes	GCST005195 (12) 9 × 10 <sup>-223</sup> GCST004894 (22) 2 × 10 <sup>-68</sup>	GWASdb v2 (23) 5 × 10 <sup>-7</sup>	none  MIAT (10hits)
<b>H19</b> ENSG00000130600 (LINC00008)	22q12.1	HOTS, MRPL23	rs2331291 (intronic)  rs217727 (exonic)	myocardial infarction  Systolic blood pressure  CAD	GCST006188 (24) 7 × 10 <sup>-15</sup>	Meta-analysis (25) 1 × 10 <sup>-7</sup> Case-control genotyping (26) 1 × 10 <sup>-3</sup>	
<b>LOC157273</b> ENSG00000248538 (AC022784.1; RP11-T0A14.4)	8p23.1	PPP1RB	rs9987289 (intronic)  rs4841132 (exonic)	HDL cholesterol  Coronary artery calcification Fasting blood insulin	GCST002223 (27) 2 × 10 <sup>-44</sup> GCST001639 (28) 2 × 10 <sup>-9</sup> GCST005185 (29) 6 × 10 <sup>-15</sup>		none  none
<b>KCNQ1OT1</b> ENSG00000269821 (KCNQ1-AS2)	11p15.5	KCNQ1	rs231362 (exonic)	Type 2 Diabetes	GCST000712 (30) 3 × 10 <sup>-13</sup>		none
<b>LINC00243</b> ENST00000419357 (lnc-IER3)	6p21.33		rs886424 (exonic)	Type 1 Diabetes Total cholesterol Total TG	GCST002876 (31) 3 × 10 <sup>-14</sup>	Meta-analysis (32) 2 × 10 <sup>-9</sup> 8 × 10 <sup>-11</sup>	(IER3, HLAs...) 153 hits
<b>LINC02137</b> ENSG00000260186 (RP11-481J2.2)	16q21		rs4784934 (intronic)	QT Interval	GCST002500 (33) 6 × 10 <sup>-9</sup>		LINC02137 (4 hits) NDRG4 (1 hit) ZNF507 (1hit) none
<b>LOC400684</b> ENSG00000267213	19q13.11	ZNF507	rs12976411 (intronic)	CAD	GCST003116 (13) 1 × 10 <sup>-14</sup>		
<b>LINC00310</b> ENSG00000227456	21q22.11	MRPS6, SLC5A3	rs28451064 (intronic)	CAD  myocardial infarction	GCST005195 (12) 1 × 10 <sup>-33</sup> GCST003117 (13) 6 × 10 <sup>-12</sup>		

A profile of the well-studied lncRNAs such as ANRIL as well as other candidate lncRNAs that carry variants which have been either associated to CAD, MI or cardiometabolic traits which are high risk factors for the disease such as T2D, hypertension, obesity.

<sup>a</sup>Official name; GeneCode ID (GrCh38.p12); aliases

<sup>b</sup>Official variant ID; variant location in the locus

<sup>c</sup>NHGRI GWAS Catalog study ID described; reference of the study; p-value

<sup>d</sup>alternative sources such as GWASdb catalog or meta-analysis studies; reference of the study; p-value

<sup>e</sup>GTEx eQTLs; genes associated to the variant (number of hits in the database).

chr9p21.3, which resulted to be the CAD risk locus with the strongest effect found up to date. Locus chr9p21.3 contained multiple SNP variants at the antisense noncoding RNA in the INK4 locus (ANRIL), now referred to as CDKN2B-AS1 (51–53). CDKN2B-AS1 spans 126.3 kb in a gene cluster next to three tumor suppressor genes (p15/CDKN2B, p16/CDKN2A and p14/ARF), partially overlapping CDKN2B (53–55). Several CDKN2B-AS1 SNP variants also associated to other disease traits such as ischemic stroke, aortic aneurysm, atherosclerosis, specific carcinomas and type 2 diabetes (T2D) (22, 56–58).

Most SNPs in the core risk region for CAD located within CDKN2B-AS1 intronic areas (118 out of 131 variants) where several enhancers were described (59). These enhancers mediated cys-regulation of neighboring genes like CDKN2A/B or methylthioadenosine phosphorylase (MTAP) but also trans-regulation of genes such as interferon- $\alpha$ 21 (IFNA21), one million base pairs upstream (59). CDKN2B-AS1 trans-regulation of gene expression increased cell adhesion and proliferation, both atherogenic processes, in a process partially mediated by ALU elements located in CDKN2B-AS1 (60). Interestingly, CDKN2B-AS1 interacted with a component of the polycomb repressor complex (PRC) 1 and 2, which control the epigenetic repression of the CDKN2B gene (61, 62). In fact, risk variant rs10757278 located at enhancer ECAD9 inside CDKN2B-AS disrupted the binding site of STAT1 transcription factor (59). In lymphoid cells, this disruption of STAT1 binding implied a failure to recruit the repressor machinery and resulted in increased CDKN2B-AS expression, a mechanism that was confirmed by the silencing of STAT1 (Figure 1A) (59).

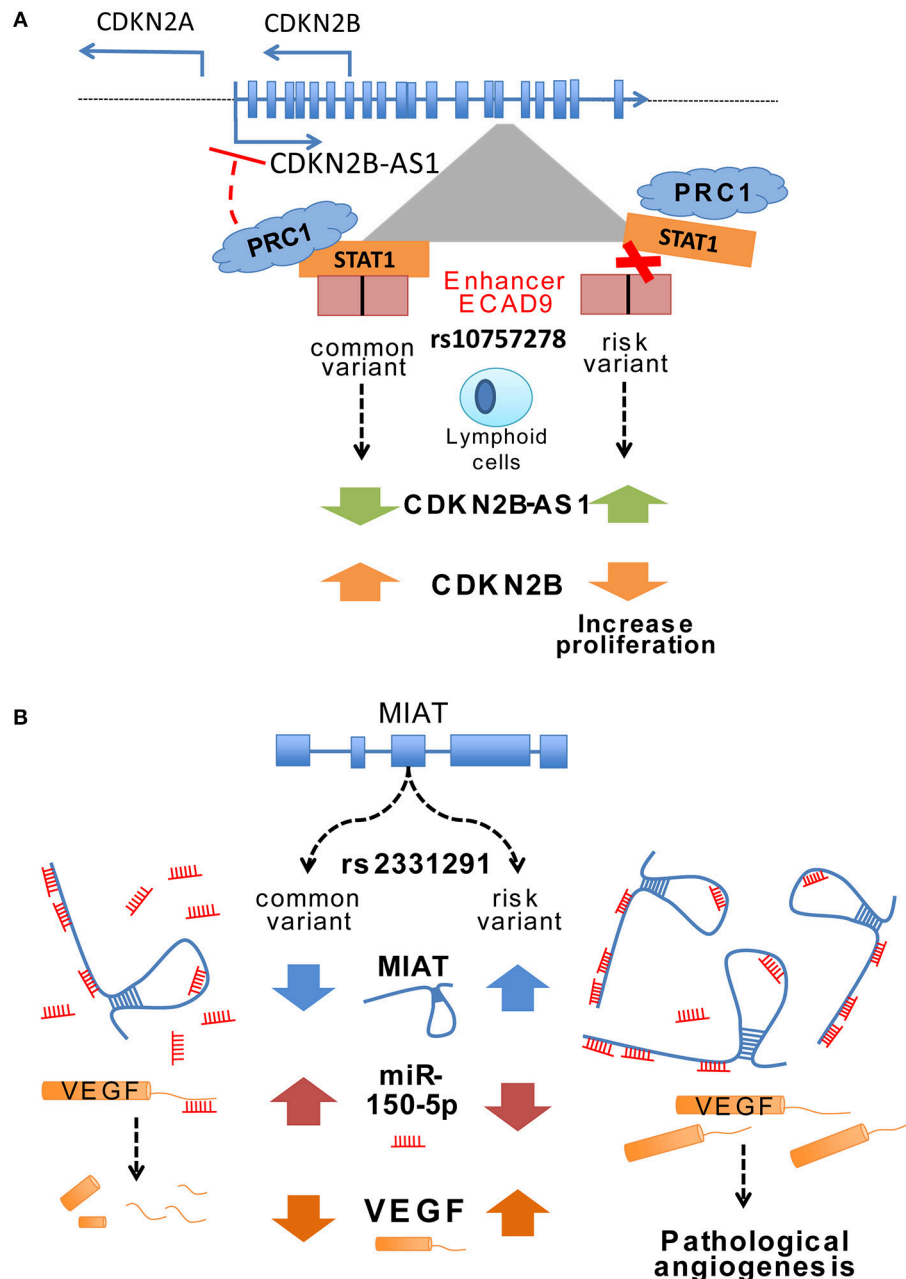
Only five of the CAD candidate variants are located in exons of CDKN2B-AS1 but none of them are located in conserved elements, questioning the likeliness to affect functional domains (59). However, numerous splice isoforms have been identified for CDKN2B-AS1 (14 isoforms, Genbank; 21 isoforms, GENCODE) highlighting a complex alternative splicing regulation that potentially affects the structural domain organization of the lncRNA leading to modulation of its functionality (64). Carriers of risk haplotype presented increased expression of CDKN2B-AS1 splice-isoforms EU741058 (short form) and NR\_003529 (long form) but not DQ485454 (short form) which directly correlated with the severity of atherosclerosis, suggesting distinct roles for CDKN2B-AS1 splicing variants (65). Additionally, splicing isoforms defined by their polyadenylation site in proximal (exon 13) or distal (exon 19) showed trans-regulation of different set of genes. Proximal CDKN2B-AS1 isoforms modulated expression of glucose and lipid metabolism genes (66) while distal isoforms regulated RBMS1 (RNA Binding Motif Single Stranded Interacting Protein 1), a cell cycle suppressor (67). Conversely, circularized CDKN2B-AS1, another form of alternative splicing, showed an atheroprotective role via interaction with pescadillo homolog 1 (PES1) which leads to impaired ribosomal biogenesis (68). An SNP located in the 3' region of CDKN2B-AS1 associated with reduced expression of CDKN2A, CDKN2B and CDKN2B-AS1 but also with increased VSMC proliferation (69). Other CDKN2B-AS1 variants confer increased myocardial infarction (MI) risk (70), supporting

previous findings, where the level of CDKN2B-AS1 significantly increased in peripheral blood mononuclear cells after MI (71). Despite great efforts, causal mechanisms of CDKN2B-AS1 variants have been elusive and not fully unravel yet. For further detail, we refer the reader to other excellent recent reviews on the topic (23, 53, 72, 73).

Myocardial infarction associated transcript (MIAT) was identified as a susceptible locus for MI in a Japanese population by large-scale case-control associated study (63). MIAT expression upregulation in a MI mouse model concomitant with increased cardiac interstitial fibrosis suggested a profibrotic role with a prominent impact in the MI pathogenesis (74). Furthermore, *ex-vivo* experiments with a diabetic rat model identified a regulatory feedback loop between MIAT, vascular endothelial growth factor (VEGF) and miR-150-5p. MIAT acts as a sponge for miR-150-5p and represses degradation of VEGF mediated by miR-150-5p (Figure 1B) (75). Expression of both MIAT and CDKN2B-AS1 increased in human atherosclerotic arteries suggesting a potential role of MIAT on atherosclerotic plaque development (76).

The embryonic lincRNA H19 was identified to be re-expressed in human atherosclerotic plaques and in a rat model of carotid artery injury (77, 78). Recently, a genotyping study of 4 SNPs in H19 locus demonstrated significant association with CAD in a Chinese population (26). Additional GWAS and meta-analysis studies proved association of H19 variants with blood pressure, a well-known risk factor for cardiovascular disease (24, 25). Mechanistically, H19 was proposed to modulate availability of several let-7 miRNAs by acting as a molecular sponge (79). Highly expressed in adult muscle tissue, H19 modulation of let-7 likely controls timing of muscle differentiation since H19 depletion accelerates *in vitro* muscle differentiation with a concomitant overexpression of let-7 (79). Additionally, H19 was highly up-regulated in two different mouse models of abdominal aortic aneurism whereas specific H19 knock-down limited aneurism growth by a mechanism involving decreased apoptosis of smooth muscle cells (80). Other lncRNAs that contained genetic variants associated to CAD have been identified by GWAS studies but not studied further on their putative causal mechanisms such as LOC400684 an uncharacterized antisense RNA in the Zinc Finger Protein 507 (ZNF507) locus (12) or lncRNA LINC00310 which variant rs28451064 is also associated to myocardial infarction (13).

Genome-wide analysis also revealed multiple variants associated to cardiometabolic traits such as cholesterol levels or type 2 diabetes (T2D), both of them established risk factors of cardiovascular disease. For example, genetic variant lying in the lincRNA LOC157273 associated to lipid (HDL cholesterol) (27) and glycemic (fasting insulin levels) (29) traits but also to coronary artery calcification (28). Genetic variants at LOC157273 associated to expression changes of the nearby gene PPP1R3B, a phosphatase involved in hepatic regulation of glucose (81). Another SNP (rs886424) located in the second exon of LINC00243 associated with total cholesterol and triglyceride levels (32). Expression quantitative trait loci (eQTL) analysis also associated variant rs886424 with LINC00243 expression levels of as well as numerous nearby immune-related genes including



**FIGURE 1 | (A)** Change of ANRIL expression through a variant in an enhancer region. The CAD associated variant rs10757278 lies within the binding site for the STAT1 transcription factor of enhancer region ECAD9. In lymphoid cells the binding of STAT1 to this region has been associated to decreased ANRIL expression, whereas silencing of STAT1 lead to an enhanced expression of ANRIL. The risk variant of rs10757278 disrupts the binding of STAT1 and the repression of ANRIL expression is abrogated. Increased expression of ANRIL promotes a downregulation of CDKN2B/p15 gene expression and underlines a proliferative effect which presumably increases CVD susceptibility. **(B)** Potential regulatory mechanisms of MIAT expression through different variants. Ishii et al. (63) unraveled that various variants are present in the lincRNA MIAT and associated them to myocardial infarction such as rs3132291. Some variants in Exon 5 have been associated to increased MIAT expression. Yan et al. showed in their study that MIAT can bind miR150-5p in endothelial cells and does inhibiting the degradation of its direct target VEGF. These data suggest that certain variants in the MIAT lincRNA can modify the structure of MIAT and thus leading to increased binding of miR-150-5p and consequently inhibiting the degradation of its target genes such as VEGF.

immediate early response 3 (IER3) and several HLA forms (32). IER3 was reported to inhibit pro-inflammatory cytokines but the exact role of LINC00243 in immune-function and its putative

link to cardiometabolic diseases requires further evaluation. One of the SNPs associated to T2D (rs231362) in the KCNQ1 locus overlaps both KCNQ1OT1 lincRNA antisense and the intron 11

of *KCNQ1* (32). Several other polymorphisms in *KCNQ1* locus associated also with cardiovascular events (82) and some showed protective effect against arrhythmic risk in long-QT syndrome (83). Both *KCNQ1OT1* and *CDKN2B-AS1* were shown to be valid predictors of left ventricle dysfunction after an MI (71). *KCNQ1OT1* is an imprinted gene that is expressed only from the paternal allele and responsible to silence a proximal cluster of genes (84). Mechanistically, *KCNQ1OT1* acts as a scaffold for the chromatin modifiers HMT G9a and PRC2 as well as DNA methyltransferase *Dnmt1* which exerts gene repression by histone modifications and DNA methylation, respectively (84).

Finally, the ARIC (Atherosclerosis Risk in Communities) study intended to establish genetic loci associated to ECG global electrical heterogeneity (GEH) and consequently changes in QT measurements and one of the identified loci contained the lncRNA *LINC02137* (33). *LINC02137* was highly expressed in human heart atrial-appendage region and eQTL analysis showed that variant rs4784934 significantly associated with the expression of *LINC02137* and gene *NDRG4* in atrial tissue. *NDRG4* was reported to be necessary for sodium channel trafficking in the nervous system but also associated with cardiomyopathy (85).

## FUTURE PERSPECTIVES OF LNCRNA GENETIC VARIANTS

Determination of potential causality among genetic variants associated with cardiovascular and cardiometabolic diseases remains a challenging future task. In the case of the functional analysis of lncRNAs it is important to consider their low expression levels and high degree of tissue and cell type specificity. For example, tissue-specific expression quantitative trait loci (eQTL) analysis of lncRNAs is a strong tool to associate certain variants to downstream effectors. Genotype-Tissue Expression (GTEx) project provides the possibility to study tissue-specific gene expression and regulation on large scale with 44 various tissues in 449 individuals, which allowed to build up a resourceful platform in order to identify genetic associations both for local (cis eQTLs) and distal (trans eQTLs) effects (86). Nonetheless, it is relevant to indicate some limitations inherent to this analysis tool such as the inability to detect small size effect eQTLs due to multiple test burden, or the fact that eQTL effects are strongly tissue specific which hinders the inference of functionality and therefore caution must be taken to extrapolate conclusions to other tissues.

Novel lncRNA were localized near leukocyte enhancers and close to GWAS identified risk variants for autoimmune diseases suggesting alterations in enhancers or super enhancers might be associated to changes in phenotype and disease risk (87). SNP in close proximity or even in far distance (e.g., in trans location to the variant), may help unravel the complex regulatory events of cardiovascular disease including underlying importance of enhancers or super-enhancers (88). Yet, the term “super-enhancer” is under debate since a clear definition has not been established and their functional properties do not

necessarily set them apart from regular enhancers (89). Another task for future studies is to determine the role of lncRNAs and their genetic variants in the maintenance and remodeling of the chromatin structure that drives interactions between enhancers and transcription initiation sites. Chromosome Conformation Capture (C3) technologies such as HiC (90, 91) or chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) (92) will be useful as genome-wide approaches to study chromatin structural changes and to define the impact of genetic variants in long-range chromatin interactomes.

The advent of new sequencing technologies that improve current throughput, length of reads and cost will increase the number of annotated lncRNAs and help to define their complex transcript models. One of such technologies is capture long-read sequencing (CLS), a technique that uses lncRNA capture enrichment with nanopore technology, which allows sequencing of longer fragments (~1.5 kb) for characterizing the lncRNA structure (93). This highly promising approach would greatly improve the task of defining exon connectivity and therefore splicing transcript models.

Another feature to improve is our ability to predict and characterize lncRNA structural motifs and their underlying functional domains. Computational analysis approaches are able to predict the formation of loops and simple helices but are not so successful to define more complex motifs (94). New high-throughput techniques based on new generation sequencing (NGS) technologies emerged to define new motifs and validate computational predictions in a genome-wide scale (94). These methods use diverse RNA nucleases (ssRNA or dsRNA) or chemical probes in combination with NGS to analyze full transcriptomes in techniques such as Parallel Analysis of RNA Structure (PARS) (95), Fragmentation Sequencing (96) or Selective 2' hydroxyl acylation analyzed by primer extension (SHAPE) (97, 98). For a detailed functional characterization of lncRNAs, novel identify structural domains should be linked to interactome information that can be obtained with novel technologies such as ChIRP (99) and CHART (100). These techniques allow the identification of specific lncRNA interacting partners such as RBPs and can also delimit the interaction sites to specific domains within the RNA molecule.

Lastly, it will be relevant to understand the potential regulatory effects that genetic variants within lncRNA have on regulation of CpG islands in cardiometabolic disorders (32). In fact, an integrative analysis of 11 human data sets generated a reference human epigenome as a framework to characterize GWAS variants that alter the epigenomic profile during complex human diseases (101), which can be also used to profile the non-coding genome.

In summary, in the post-GWAS era many relevant factors must be considered in order to study the effect of genetic variation in lncRNA, some of which comprise differential tissue expression, splicing isoforms models, RNA structural prediction and functional domain identification, and identification of lncRNA interacting partners such as RBPs. The high proportion of disease-associated SNPs lying in non-coding regions highlighted their functional relevance and prompted a

better understanding of lncRNA biology as well as regulatory regions such as enhancer to unravel their potential role in cardiometabolic diseases. The expansion of the GWAS field to explore the functionality of lncRNA but also other non-coding RNAs will provide potential novel regulatory causal mechanisms of cardiovascular disease. This research area warrants interesting new insights into underlying mechanisms that determine the genetic component of human disease and will clear the path toward a personalized medicine approach.

## REFERENCES

- Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. *Nature* (2015) 526:68–74. doi: 10.1038/nature15393
- Genomes Project C, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, et al. A map of human genome variation from population-scale sequencing. *Nature* (2010) 467:1061–73. doi: 10.1038/nature09534
- Pazin MJ, Using the ENCODE resource for functional annotation of genetic variants. *Cold Spring Harb Protoc.* (2015) 2015:522–36. doi: 10.1101/pdb.top084988
- Qu H, Fang X. A brief review on the Human Encyclopedia of DNA Elements (ENCODE) project. *Genom Prot Bioinform.* (2013) 11:135–41. doi: 10.1016/j.gpb.2013.05.001
- de Hoon M, Shin JW, Carninci P. Paradigm shifts in genomics through the FANTOM projects. *Mammal Genome* (2015) 26:391–402. doi: 10.1007/s00335-015-9593-8
- Helgadottir A, Thorleifsson G, Manolescu A, Gretarsdottir S, Blondal T, Jonasdottir A, et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science* (2007) 316:1491–3. doi: 10.1126/science.1142842
- McPherson R, Pertsemlidis A, Kavaslar N, Stewart A, Roberts R, Cox DR, et al. A common allele on chromosome 9 associated with coronary heart disease. *Science* (2007) 316:1488–91. doi: 10.1126/science.1142447
- Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, Mayer B, et al. Genomewide association analysis of coronary artery disease. *N Eng J Med.* (2007) 357:443–53. doi: 10.1056/NEJMoa072366
- Wellcome Trust Case Control C. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* (2007) 447:661–78. doi: 10.1038/nature05911
- Erdmann J, Grosshennig A, Braund PS, König IR, Hengstenberg C, Hall AS, et al. New susceptibility locus for coronary artery disease on chromosome 3q22.3. *Nature Genet.* (2009) 41:280–2. doi: 10.1038/ng.307
- Consortium CAD, Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet.* (2013) 45:25–33. doi: 10.1038/ng.2480
- van der Harst P, Verweij N. Identification of 64 novel genetic loci provides an expanded view on the genetic architecture of coronary artery disease. *Circ Res.* (2018) 122:433–43. doi: 10.1161/CIRCRESAHA.117.312086
- Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet.* (2015) 47:1121–30. doi: 10.1038/ng.3396
- Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. *Nature* (2009) 461:747–53. doi: 10.1038/nature08494
- Chasman DI. Peering into the future of CAD genomics. *Circ Res.* (2018) 122:391–3. doi: 10.1161/CIRCRESAHA.117.312502
- Boyle EA, Li YI, Pritchard JK. An expanded view of complex traits: from polygenic to omnigenic. *Cell* (2017) 169:1177–86. doi: 10.1016/j.cell.2017.05.038

## AUTHOR CONTRIBUTIONS

AK and HG screened the literature on the topic, drafted, wrote and revised the article. UL revised the article.

## ACKNOWLEDGMENTS

We acknowledge support from the German Research Foundation (DFG) and the Open Access Publication Fund of Charité—Universitätsmedizin Berlin.

- Wray NR, Wijmenga C, Sullivan PF, Yang J, Visscher PM. Common disease is more complex than implied by the core gene omnigenic model. *Cell* (2018) 173:1573–580. doi: 10.1016/j.cell.2018.05.051
- Edwards SL, Beesley J, French JD, Dunning AM. Beyond GWASs: illuminating the dark road from association to function. *Am J Hum Genet.* (2013) 93:779–97. doi: 10.1016/j.ajhg.2013.10.012
- Hrdlickova B, de Almeida RC, Borek Z, Withoff S. Genetic variation in the non-coding genome: Involvement of micro-RNAs and long non-coding RNAs in disease. *Biochim Biophys Acta* (2014) 1842:1910–22. doi: 10.1016/j.bbdis.2014.03.011
- Mirza AH, Kaur S, Brorsson CA, Pociot F. Effects of GWAS-associated genetic variants on lncRNAs within IBD and T1D candidate loci. *PLoS ONE* (2014) 9:e105723. doi: 10.1371/journal.pone.0105723
- Volders PJ, Verheggen K, Menschaert G, Vandepoele K, Martens L, Vandesompele J, et al. An update on LNCipedia: a database for annotated human lncRNA sequences. *Nucl Acids Res.* (2015) 43:D174–80. doi: 10.1093/nar/gkv295
- Zhao W, Rasheed A, Tikkanen E, Lee JJ, Butterworth AS, Howson JMM, et al. Identification of new susceptibility loci for type 2 diabetes and shared etiological pathways with coronary heart disease. *Nat Genet.* (2017) 49:1450–7. doi: 10.1038/ng.3943
- Aguilo F, Di Cecilia S, Walsh MJ. Long Non-coding RNA ANRIL and polycomb in human cancers and cardiovascular disease. *Curr Top Microbiol Immunol.* (2016) 394:29–39. doi: 10.1007/82\_2015\_455
- Sung YJ, Winkler TW, de Las Fuentes L, Bentley AR, Brown MR, Kraja AT, et al. A large-scale multi-ancestry genome-wide study accounting for smoking behavior identifies multiple significant loci for blood pressure. *Am J Hum Genet.* (2018) 102:375–400. doi: 10.1016/j.ajhg.2018.01.015
- Tragante V, Barnes MR, Ganesh SK, Lanktree MB, Guo W, Franceschini N, et al. Gene-centric meta-analysis in 87,736 individuals of European ancestry identifies multiple blood-pressure-related loci. *Am J Hum Genet.* (2014) 94:349–60. doi: 10.1016/j.ajhg.2013.12.016
- Gao W, Zhu M, Wang H, Zhao S, Zhao D, Yang Y, et al. Association of polymorphisms in long non-coding RNA H19 with coronary artery disease risk in a Chinese population. *Mutat Res.* (2015) 772:15–22. doi: 10.1016/j.mrfmmm.2014.12.009
- Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, et al. Discovery and refinement of loci associated with lipid levels. *Nat Genet.* (2013) 45:1274–83. doi: 10.1038/ng.2797
- Inouye M, Ripatti S, Kettunen J, Lyttikainen LP, Oksala N, Laurila PP, et al. Novel Loci for metabolic networks and multi-tissue expression studies reveal genes for atherosclerosis. *PLoS Genet.* (2012) 8:e1002907. doi: 10.1371/journal.pgen.1002907
- Manning AK, Hivert MF, Scott RA, Grimsby JL, Bouatia-Naji N, Chen H, et al. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat Genet.* (2012) 44:659–69. doi: 10.1038/ng.2274
- Tomer Y, Dolan LM, Kahaly G, Divers J, D'Agostino RB, Jr, et al. Genome wide identification of new genes and pathways in patients with both autoimmune thyroiditis and type 1 diabetes. *J Autoimmunity* (2015) 60:32–9. doi: 10.1016/j.jaut.2015.03.006

31. Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, Welch RP, et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nature Genet.* (2010) 42:579–89. doi: 10.1038/ng.609
32. Ghanbari M, Peters MJ, de Vries PS, Boer CG, van Rooij JGJ, Lee YC, et al. A systematic analysis highlights multiple long non-coding RNAs associated with cardiometabolic disorders. *J Hum Genet.* (2018) 63:431–46. doi: 10.1038/s10038-017-0403-x
33. Arking DE, Pulit SL, Crotti L, van der Harst P, Munroe PB, Koopmann TT, et al. Genetic association study of QT interval highlights role for calcium signaling pathways in myocardial repolarization. *Nat Genet.* (2014) 46:826–36. doi: 10.1038/ng.3014
34. Marian AJ. Recent developments in cardiovascular genetics and genomics. *Circ Res.* (2014) 115:e11–7. doi: 10.1161/CIRCRESAHA.114.305054
35. Ning S, Yue M, Wang P, Liu Y, Zhi H, Zhang Y, et al. LincSNP 2.0: an updated database for linking disease-associated SNPs to human long non-coding RNAs and their TFBSs. *Nucl Acids Res.* (2017) 45:D74–8. doi: 10.1093/nar/gkw945
36. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucl Acids Res.* (2017) 45:D896–901. doi: 10.1093/nar/gkw1133
37. Li MJ, Liu Z, Wang P, Wong MP, Nelson MR, Kocher JP, et al. GWASdb v2: an update database for human genetic variants identified by genome-wide association studies. *Nucl Acids Res.* (2016) 44:D869–76. doi: 10.1093/nar/gkv1317
38. Ballantyne RL, Zhang X, Nunez S, Xue C, Zhao W, Reed E, et al. Genome-wide interrogation reveals hundreds of long intergenic noncoding RNAs that associate with cardiometabolic traits. *Hum Mol Genet.* (2016) 25:3125–41. doi: 10.1093/hmg/ddw154
39. Dechamethakun S, Muramatsu M. Long noncoding RNA variations in cardiometabolic diseases. *J Hum Genet.* (2017) 62:97–104. doi: 10.1038/jhg.2016.70
40. Ghanbari M, de Vries PS, de Looper H, Peters MJ, Schurmann C, Yaghothkar H, et al. A genetic variant in the seed region of miR-4513 shows pleiotropic effects on lipid and glucose homeostasis, blood pressure, and coronary artery disease. *Hum Mutat.* (2014) 35:1524–31. doi: 10.1002/humu.22706
41. Gong J, Liu W, Zhang J, Miao X, Guo AY. lncRNASNP: a database of SNPs in lncRNAs and their potential functions in human and mouse. *Nucl Acids Res.* (2015) 43:D181–6. doi: 10.1093/nar/gku1000
42. Qian J, Nunez S, Reed E, Reilly MP, Foulkes AS. A simple test of class-level genetic association can reveal novel cardiometabolic trait loci. *PLoS ONE* (2016) 11:e0148218. doi: 10.1371/journal.pone.0148218
43. Noh JH, Kim KM, McClusky WG, Abdelmohsen K, Gorospe M. Cytoplasmic functions of long noncoding RNAs. *Wiley Interdiscip Rev RNA* (2018). 9:e1471. doi: 10.1002/wrna.1471
44. Engreitz JM, Ollikainen N, Guttman M. Long non-coding RNAs: spatial amplifiers that control nuclear structure and gene expression. *Nat Rev Mol Cell Biol.* (2016) 17:756–70. doi: 10.1038/nrm.2016.126
45. Mercer TR, Mattick JS. Structure and function of long noncoding RNAs in epigenetic regulation. *Nat Struct Mol Biol.* (2013) 20:300–7. doi: 10.1038/nsmb.2480
46. Quinn JJ, Chang HY. Unique features of long non-coding RNA biogenesis and function. *Nat Rev Genet.* (2016) 17:47–62. doi: 10.1038/nrg.2015.10
47. Bhartiya D, Scaria V. Genomic variations in non-coding RNAs: structure, function and regulation. *Genomics* (2016) 107:59–68. doi: 10.1016/j.ygeno.2016.01.005
48. Bhartiya D, Jalali S, Ghosh S, Scaria V. Distinct patterns of genetic variations in potential functional elements in long noncoding RNAs. *Hum Mutat.* (2014) 35:192–201. doi: 10.1002/humu.22472
49. Nurnberg ST, Zhang H, Hand NJ, Bauer RC, Saleheen D, Reilly MP, et al. From loci to biology: functional genomics of genome-wide association for coronary disease. *Circ Res.* (2016) 118:586–606. doi: 10.1161/CIRCRESAHA.115.306464
50. Cai L, Chang H, Fang Y, Li G. A comprehensive characterization of the function of lincRNAs in transcriptional regulation through long-range chromatin interactions. *Sci Rep.* (2016) 6:36572. doi: 10.1038/srep36572
51. Broadbent HM, Peden JF, Lorkowski S, Goel A, Ongen H, Green F, et al. Susceptibility to coronary artery disease and diabetes is encoded by distinct, tightly linked SNPs in the ANRIL locus on chromosome 9p. *Hum Mol Genet.* (2008) 17:806–14. doi: 10.1093/hmg/ddm352
52. Pasmant E, Laurendeau I, Heron D, Vidaud M, Vidaud D, Bieche I. Characterization of a germ-line deletion, including the entire INK4/ARF locus, in a melanoma-neural system tumor family: identification of ANRIL, an antisense noncoding RNA whose expression coclusters with ARF. *Cancer Res.* (2007) 67:3963–9. doi: 10.1158/0008-5472.CAN-06-2004
53. Pasmant E, Sabbagh A, Vidaud M, Bieche I. ANRIL, a long, noncoding RNA, is an unexpected major hotspot in GWAS. *FASEB J.* (2011) 25:444–8. doi: 10.1096/fj.10-172452
54. Jarinova O, Stewart AF, Roberts R, Wells G, Lau P, Naing T, et al. Functional analysis of the chromosome 9p21.3 coronary artery disease risk locus. *Arterioscler Thromb Vasc Biol.* (2009) 29:1671–7. doi: 10.1161/ATVBAHA.109.189522
55. Yu W, Gius D, Onyango P, Muldoon-Jacobs K, Karp J, Feinberg AP, et al. Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. *Nature* (2008) 451:202–6. doi: 10.1038/nature06468
56. Diabetes Genetics Initiative of Broad Institute of H. Mit LU, Novartis Institutes of BioMedical R, Saxena R, Voight BF, Lyssenko V, Burt NP, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* (2007) 316:1331–6. doi: 10.1126/science.1142358
57. Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* (2007) 316:1341–5. doi: 10.1126/science.1142382
58. Cunnington MS, Santibanez Koref M, Mayosi BM, Burn J, Keavney B. Chromosome 9p21 SNPs associated with multiple disease phenotypes correlate with ANRIL expression. *PLoS Genet.* (2010) 6:e1000899. doi: 10.1371/journal.pgen.1000899
59. Harismendy O, Notani D, Song X, Rahim NG, Tanasa B, Heintzman N, et al. 9p21 DNA variants associated with coronary artery disease impair interferon-gamma signalling response. *Nature* (2011) 470:264–8. doi: 10.1038/nature09753
60. Holdt LM, Hoffmann S, Sass K, Langenberger D, Scholz M, Krohn K, et al. Alu elements in ANRIL non-coding RNA at chromosome 9p21 modulate atherogenic cell functions through trans-regulation of gene networks. *PLoS Genet.* (2013) 9:e1003588. doi: 10.1371/journal.pgen.1003588
61. Kotake Y, Nakagawa T, Kitagawa K, Suzuki S, Liu N, Kitagawa M, Xiong Y. Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene. *Oncogene* (2011) 30:1956–62. doi: 10.1038/onc.2010.568
62. Yap KL, Li S, Munoz-Cabello AM, Raguz S, Zeng L, Mujtaba S, et al. Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Mol Cell* (2010) 38:662–74. doi: 10.1016/j.molcel.2010.03.021
63. Ishii N, Ozaki K, Sato H, Mizuno H, Saito S, Takahashi A, et al. Identification of a novel non-coding RNA, MIAT, that confers risk of myocardial infarction. *J Hum Genet.* (2006) 51:1087–99. doi: 10.1007/s10038-006-0070-9
64. Folkersen L, Kyriakou T, Goel A, Peden J, Malarstig A, Paulsson-Berne G, et al. Relationship between CAD risk genotype in the chromosome 9p21 locus and gene expression. Identification of eight new ANRIL splice variants. *PLoS ONE* (2009) 4:e7677. doi: 10.1371/journal.pone.0007677
65. Holdt LM, Beutner F, Scholz M, Gielen S, Gabel G, Bergert H, et al. ANRIL expression is associated with atherosclerosis risk at chromosome 9p21. *Arterioscler Thromb Vasc Biol.* (2010) 30:620–7. doi: 10.1161/ATVBAHA.109.196832
66. Bochenek G, Hasler R, El Mokhtari NE, König IR, Loos BG, Jepsen S, et al. The large non-coding RNA ANRIL, which is associated with atherosclerosis, periodontitis and several forms of cancer, regulates ADIPOR1, VAMP3 and C11ORF10. *Hum Mol Genet.* (2013) 22:4516–27. doi: 10.1093/hmg/ddt299
67. Hubberten M, Bochenek G, Chen H, Hasler R, Wiehe R, Rosenstiel P, et al. Linear isoforms of the long noncoding RNA CDKN2B-AS1 regulate the c-myc-enhancer binding factor RBMS1. *Eur J Hum Genet.* (2018). doi: 10.1038/s41431-018-0210-7. [Epub ahead of print].
68. Holdt LM, Stahringer A, Sass K, Pichler G, Kulak NA, Wilfert W, et al. Circular non-coding RNA ANRIL modulates ribosomal RNA maturation and atherosclerosis in humans. *Nat Commun.* (2016) 7:12429. doi: 10.1038/ncomms12429



69. Motterle A, Pu X, Wood H, Xiao Q, Gor S, Ng FL, et al. Functional analyses of coronary artery disease associated variation on chromosome 9p21 in vascular smooth muscle cells. *Hum Mol Genet.* (2012) 21:4021–9. doi: 10.1093/hmg/dds224
70. Cheng J, Cai MY, Chen YN, Li ZC, Tang SS, Yang XL, et al. Variants in ANRIL gene correlated with its expression contribute to myocardial infarction risk. *Oncotarget* (2017) 8:12607–19. doi: 10.18632/oncotarget.14721
71. Vausort M, Wagner DR, Devaux Y. Long noncoding RNAs in patients with acute myocardial infarction. *Circ Res.* (2014) 115:668–77. doi: 10.1161/CIRCRESAHA.115.303836
72. Aarabi G, Zeller T, Heydecke G, Munz M, Schafer A, Seedorf U. Roles of the Chr.9p21.3 ANRIL locus in regulating inflammation and implications for anti-inflammatory drug target identification. *Front Cardiovasc Med.* (2018) 5:47. doi: 10.3389/fcvm.2018.00047
73. Hannou SA, Wouters K, Paumelle R, Staels B. Functional genomics of the CDKN2A/B locus in cardiovascular and metabolic disease: what have we learned from GWASs? *Trends Endocrinol Metabol.* (2015) 26:176–84. doi: 10.1016/j.tem.2015.01.008
74. Qu X, Du Y, Shu Y, Gao M, Sun F, Luo S, et al. MIAT Is a Pro-fibrotic Long non-coding RNA governing cardiac fibrosis in post-infarct myocardium. *Sci Rep.* (2017) 7:42657. doi: 10.1038/srep42657
75. Yan B, Yao J, Liu JY, Li XM, Wang XQ, Li YJ, et al. lncRNA-MIAT regulates microvascular dysfunction by functioning as a competing endogenous RNA. *Circ Res.* (2015) 116:1143–56. doi: 10.1161/CIRCRESAHA.116.305510
76. Arslan S, Berkan O, Lalem T, Ozbilum N, Goksel S, Korkmaz O, et al. Long non-coding RNAs in the atherosclerotic plaque. *Atherosclerosis* (2017) 266:176–81. doi: 10.1016/j.atherosclerosis.2017.10.012
77. Han DK, Khaing ZZ, Pollock RA, Haudenschild CC, Liu G. H19, a marker of developmental transition, is reexpressed in human atherosclerotic plaques and is regulated by the insulin family of growth factors in cultured rabbit smooth muscle cells. *J Clin Invest.* (1996) 97:1276–85. doi: 10.1172/JCI118543
78. Kim DK, Zhang L, Dzau VJ, Pratt RE. H19, a developmentally regulated gene, is reexpressed in rat vascular smooth muscle cells after injury. *J Clin Invest.* (1994) 93:355–60. doi: 10.1172/JCI116967
79. Kallen AN, Zhou XB, Xu J, Qiao C, Ma J, Yan L, et al. The imprinted H19 lncRNA antagonizes let-7 microRNAs. *Mol Cell.* (2013) 52:101–12. doi: 10.1016/j.molcel.2013.08.027
80. Li DY, Busch A, Jin H, Chernogubova E, Pelisek J, Karlsson J, et al. H19 induces abdominal aortic aneurysm development and progression. *Circulation* (2018) 138:1551–68. doi: 10.1161/CIRCULATIONAHA.117.032184
81. Mehta MB, Shewale SV, Sequeira RN, Millar JS, Hand NJ, Rader DJ. Hepatic protein phosphatase 1 regulatory subunit 3B (Ppp1r3b) promotes hepatic glycogen synthesis and thereby regulates fasting energy homeostasis. *J Biol Chem.* (2017) 292:10444–54. doi: 10.1074/jbc.M116.766329
82. Olszak-Waskiewicz M, Dziuk M, Kubik L, Kaczanowski R, Kucharczyk K. Novel KCNQ1 mutations in patients after myocardial infarction. *Cardiol J.* (2008) 15:252–60.
83. Duchatelet S, Crotti L, Peat RA, Denjoy I, Itoh H, Berthet M, et al. Identification of a KCNQ1 polymorphism acting as a protective modifier against arrhythmic risk in long-QT syndrome. *Circ Cardiovasc Genet.* (2013) 6:354–61. doi: 10.1161/CIRCGENETICS.113.000023
84. Marchese FP, Huarte M. Long non-coding RNAs and chromatin modifiers: their place in the epigenetic code. *Epigenetics* (2014) 9:21–6. doi: 10.4161/epi.27472
85. Tereshchenko LG, Sotoodehnia N, Sitlani CM, Ashar FN, Kabir M, Biggs ML, et al. Genome-wide associations of global electrical heterogeneity ECG phenotype: the ARIC (Atherosclerosis Risk in Communities) Study and CHS (Cardiovascular Health Study). *J Am Heart Assoc.* (2018) 7:e008160. doi: 10.1161/JAHA.117.008160
86. Consortium GT, Laboratory DA, Coordinating Center -Analysis Working G, Statistical Methods groups-Analysis Working G, Enhancing Gg, Fund NIHC, et al. Genetic effects on gene expression across human tissues. *Nature* (2017) 550:204–13. doi: 10.1038/nature24277
87. Aune TM, Crooke PS, 3rd, Patrick AE, Tossberg JT, Olsen NJ, Spurlock CF, 3rd. Expression of long non-coding RNAs in autoimmunity and linkage to enhancer function and autoimmune disease risk genetic variants. *J Autoimmunity* (2017) 81:99–109. doi: 10.1016/j.jaut.2017.03.014
88. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-Andre V, Sigova AA, et al. Super-enhancers in the control of cell identity and disease. *Cell* (2013) 155:934–47. doi: 10.1016/j.cell.2013.09.053
89. Pott S, Lieb JD. What are super-enhancers? *Nat Genet.* (2015) 47:8–12. doi: 10.1038/ng.3167
90. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* (2009) 326:289–93. doi: 10.1126/science.1181369
91. Mishra A, Hawkins RD. Three-dimensional genome architecture and emerging technologies: looping in disease. *Genome Med.* (2017) 9:87. doi: 10.1186/s13073-017-0477-2
92. Li X, Luo OJ, Wang P, Zheng M, Wang D, Piecuch E, et al. Long-read ChIA-PET for base-pair-resolution mapping of haplotype-specific chromatin interactions. *Nat Protoc.* (2017) 12:899–915. doi: 10.1038/nprot.2017.012
93. Uszczynska-Ratajczak B, Lagarde J, Frankish A, Guigo R, Johnson R. Towards a complete map of the human long non-coding RNA transcriptome. *Nat Rev Genet.* (2018) 19:535–48. doi: 10.1038/s41576-018-0017-y
94. Mortimer SA, Kidwell MA, Doudna JA. Insights into RNA structure and function from genome-wide studies. *Nat Rev Genet.* (2014) 15:469–79. doi: 10.1038/nrg3681
95. Kertesz M, Wan Y, Mazar E, Rinn JL, Nutter RC, Chang HY, Segal E. Genome-wide measurement of RNA secondary structure in yeast. *Nature* (2010) 467:103–7. doi: 10.1038/nature09322
96. Underwood JG, Uzilov AV, Katzman S, Onodera CS, Mainzer JE, Mathews DH, et al. FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. *Nat Methods* (2010) 7:995–1001. doi: 10.1038/nmeth.1529
97. Spitale RC, Flynn RA, Zhang QC, Crisalli P, Lee B, Jung JW, et al. Structural imprints *in vivo* decode RNA regulatory mechanisms. *Nature* (2015) 519:486–90. doi: 10.1038/nature14263
98. Wilkinson KA, Merino EJ, Weeks KM. Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat Protoc.* (2006) 1:1610–6. doi: 10.1038/nprot.2006.249
99. Chu C, Quinn J, Chang HY. Chromatin isolation by RNA purification (ChIRP). *J Vis Exp.* (2012) 61:3912. doi: 10.3791/3912
100. Simon MD, Wang CI, Kharchenko PV, West JA, Chapman BA, Alekseyenko AA, et al. The genomic binding sites of a noncoding RNA. *Proc Natl Acad Sci USA.* (2011) 108:20497–502. doi: 10.1073/pnas.1113536108
101. Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenyk M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. *Nature* (2015) 518:317–30. doi: 10.1038/nature14248

**Conflict of Interest Statement:** 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.

The reviewer AS declared a shared affiliation, with no collaboration, with the authors to the handling Editor.

Copyright © 2018 Giral, Landmesser and Kratzer. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.