FROM GWAS HITS TO TREATMENT TARGETS

EDITED BY: Jeanette Erdmann and Tanja Zeller
PUBLISHED IN: Frontiers in Cardiovascular Medicine

frontiers Research Topics

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

FROM GWAS HITS TO TREATMENT TARGETS

Topic Editors:
Jeanette Erdmann, Universität zu Lübeck, Deutsches Zentrum für
Herz-Kreislaufforschung, Germany
Tanja Zeller, Universität Hamburg, Deutsches Zentrum für
Herz-Kreislaufforschung, Germany

Genome-wide association (GWA) studies, as a prototype of large-scale OMICs studies, have advanced our understanding of the genetic basis of many common diseases. With respect to coronary artery disease (CAD) and cardiovascular risk factors, like lipids, blood pressure or BMI, they have identified hundreds of chromosomal loci that modulate disease risk. Despite their scientific success, GWA studies have been criticized for having failed so far in delivering diagnostically or therapeutically relevant products. However, the ability to achieve such goals has been strengthened recently by further layers of OMICs-based data, including large-scale transcriptomics data, and better annotation of regulatory sequences and epigenetic changes in the genome (e.g. through the ENCODE project), as well as novel tools for bioinformatics analysis, allowing a systems medicine based approach to be applied. All in all, the last decade with its "gold rush of genomic discovery" led to the identification of known and novel pathways involved in the pathogenesis of cardiovascular diseases and point to novel treatment targets. This Research Topic has gathered contributions from scientists working in the field of cardiovascular genetics who have common interests in understanding the pathomechanisms linking genetic association findings and disease to finally translate the findings from large-scale genetic studies into novel treatment options.

Table of Contents

Evaluation of 71 Coronary Artery Disease Risk Variants in a Multiethnic Cohort

Wangjing Ke [1], Kristin A. Rand [2], David V. Conti [1], Veronica W. Setiawan [1], Daniel O. Stram [1], Lynne Wilkens [3], Loic Le Marchand [3], Themistocles L. Assimes [4,5] and Christopher A. Haiman [1]*

[1] Department of Preventive Medicine, Keck School of Medicine of USC, Los Angeles, CA, United States, [2] Ancestry, San Francisco, CA , United States, [3] Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI, United States, [4] Department of Medicine, Stanford University School of Medicine, Stanford, CA, United States, [5] Cardiovascular Institute, Stanford University School of Medicine, Stanford, CA, United States

Background: Coronary heart disease (CHD) is the most common cause of death worldwide. Previous studies have identified numerous common CHD susceptibility loci, with the vast majority identified in populations of European ancestry. How well these findings transfer to other racial/ethnic populations remains unclear.

Methods and Results: We examined the generalizability of the associations with 71 known CHD loci in African American, Latino and Japanese men and women in the Multiethnic Cohort (6,035 cases and 11,251 controls). In the combined multiethnic sample, 78% of the loci demonstrated odds ratios that were directionally consistent with those previously reported ($p = 2 \times 10^{-6}$), with this fraction ranging from 59% in Japanese to 70% in Latinos. The number of nominally significant associations across all susceptibility regions ranged from only 1 in Japanese to 11 in African Americans with the most statistically significant association observed through locus fine-mapping noted for rs3832016 (OR = 1.16, $p = 2.5 \times 10^{-5}$) in the *SORT1* region on chromosome *1p13*. Lastly, we examined the cumulative predictive effect of CHD SNPs across populations with improved power by creating genetic risk scores (GRSs) that summarize an individual's aggregated exposure to risk variants. We found the GRSs to be significantly associated with risk in African Americans (OR = 1.03 per allele; $p = 4.1 \times 10^{-5}$) and Latinos (OR = 1.03; $p = 2.2 \times 10^{-8}$), but not in Japanese (OR = 1.01; $p = 0.11$).

Conclusions: While a sizable fraction of the known CHD loci appear to generalize in these populations, larger fine-mapping studies will be needed to localize the functional alleles and better define their contribution to CHD risk in these populations.

Keywords: coronary heart disease, genome wide association study (GWAS), multi-ethnic, African Americans, Latino American, Japanese Americans, *SORT1*

INTRODUCTION

Coronary heart disease (CHD) is the most common, chronic, life-threatening illness in the United States, affecting more than 11 million people (1). A study with twins has estimated the genetic contribution to the variation in CHD mortality to be 0.57 and 0.38 in males and females, respectively (2). Genome-wide association studies (GWAS) have been conducted primarily in populations of European ancestry and

have identified ~65 regions associated with CHD risk (3–11). Many of the CHD loci were identified in a large study of 22,233 case and 64,762 control of European ancestry in the CARDIoGRAMplusC4D consortium, which reported 46 genome-wide significant variants with odds ratios ranging from 1.01 to 2.08 and effect allele frequencies of 0.06–0.91 (9). More recently, 10 additional loci were reported from the same consortium in a genome-wide association study involving 61,289 cases and 126,310 controls subjects following imputation to the 1,000 Genomes Project reference panel (12). Genome-wide scans have also revealed 7 CHD risk loci in Asian populations (13–17). The known genetic risk variants for CHD are estimated to explain only 10–11% of the heritability of CHD (9, 12), suggesting that many additional genetic susceptibility loci remain to be discovered.

Several studies in Asian populations have reported successful replication of known CHD regions (17–20), with a reproducible disease association consistently noted with the *9p21* region. A limited number studies have been performed to investigate risk associated with CHD variants in minority groups such as African Americans or Latinos (21–28). In 2011, a GWAS in African Americans found a SNP, rs1859023, located at *7q21* near the *PFTK1* gene to be significantly associated with CHD (22), however this finding has never been replicated in African Americans or any other racial/ethnic group. In a study of 8,090 African Americans (~700 CHD cases) that examined known CHD risk regions, only *9p21* was found to be associated with CHD (25). In a study of 8,201 African Americans (~550 CHD cases) (26), investigators found consistent direction of effects compared to studies of European ancestry for 23 of 44 (binomial $p = 0.52$) known loci with two nominally statistically significant (rs599839 at *1p13/SORT1* and rs579459 at *4p23/ABO*). Genetic studies of CHD in Latino populations have been extremely limited. In a Costa Rican study that examined only 14 CHD SNPs in 1,898 cases with MI and 2,096 controls, 7 variants at 3 regions (*SORT1*, *CXCL12*, and *9p21*) were found to be significantly associated with risk (29). Thus, additional studies are needed to understand the generalizability and relevance of the known CHD risk loci in populations of non-European ancestry.

In this context, the objective of this study was threefold. First, we wished to determine whether associations involving 71 known susceptibility variants of CHD from 65 independent regions generalize across African-American, Latino and Japanese men and women in the Multiethnic Cohort, a study that includes over 6,000 cases and 11,000 controls. Second, we evaluated common genetic variation across each susceptibility region in an attempt to identify variation that might better define the risk associations compared to the index variants in the multiethnic sample. Lastly, we constructed genetic risk scores (GRS) summarizing one's degree of exposure to high risk alleles of CHD and evaluated to what degree this GRS contributes to population differences in CHD risk.

METHODS

Study Population

The Multiethnic Cohort study (MEC) is a large prospective cohort study that was established between 1993 and 1996. The MEC includes primarily African Americans, Japanese American,

Native Hawaiians, Latinos and European Americans living in Hawaii and California. Cohort members were recruited through Department of Motor Vehicle license files and supplemented by voter registration and Health Care Financing Administration (Medicare) files. Participating individuals were between 45 and 75 years of age, and completed a 26-page self-administered, detailed questionnaire at cohort entry (baseline data, 1993–1996). The questionnaire included basic demographic factors (including race/ethnicity and education), lifestyle factors (e.g., diet, medication use and smoking history), and chronic medical conditions. Follow-up questionnaires were also administered in years 1999 and 2003 which contained updates on participant's CHD status and lifestyle factors.

Several nested case-control studies have been assembled in the MEC for GWAS of a number of cancer and non-cancer traits (30–32) including breast cancer, prostate cancer, and type-2 diabetes, mainly in populations of non-European ancestry. In the current study, we identified CHD cases and non-cases within these nested studies for the genetic analysis of CHD risk SNPs.

The MEC study obtained written informed consent from study participants for genetic analysis, approval from the Health Science Review Board (HSIRB) at the University of Southern California, and IRB certification permitting data sharing in accordance with the NIH Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies (GWAS). Genetic data for the MEC is available on dbGAP (phs000517.v3.p1, phs000851.v1.p1, phs000356.v2.p1, phs000306.v4.p1, phs000683.v1.p1)

CHD CASE/CONTROL DEFINITIONS

CHD cases were identified through linkage of the MEC to the California Hospital Discharge Data (1990–2012) (CHDD) and the Centers for Medicare and Medicaid Services (CMS) claim files (MedPAR, outpatient) (1999–2011). Hospital discharge information was not available for the subjects from Hawaii which included 76.6% of the Japanese men and women. A CHD case was defined as having ischemic heart disease under ICD-9 codes (DX 410–414), by the principal or first diagnosis code and the principal or first procedure code. We also included cases with a primary cause of death due to myocardial infarction (ICD-9 DX410, ICD-10 I21), or other CHD conditions (ICD-9 DX411–414, ICD-10 I20, I22–25). Both prevalent and incident CHD cases were included in this study. Of the 6,035 CHD cases identified, 1,146 were identified from their baseline questionnaires at the time of enrollment in the MEC study, and a majority of these prevalent cases (1,122, 97.9%) were also identified from CHDD or Medicare.

Controls in this study were subjects with no history of heart attack or angina based on the baseline questionnaire or all subsequent follow-up questionnaires. Those taking nitrates at blood draw in subsequent examinations were also excluded. Individuals with non-primary CHD diagnosis codes (i.e., 2–24) from the CHDD and Medicare data were excluded from being either a case or control. A total of 11,251 controls were selected, of which 8,307 had at least one previous Medicare or CHDD claim (and thus would have

been identified as a case). A sensitivity analysis using controls with definite claim information was performed.

GENOTYPING AND QUALITY CONTROL

We utilized genetic data generated from case-control studies in the MEC of breast cancer, prostate cancer, and type 2 diabetes in African Americans (2,976 males and 3,539 females), Japanese Americans (2,530 males and 2,132 females), and Latinos (3,340 males and 2,769 females). Genotyping was conducted using the Illumina platform with different arrays, including the Human 1M-Duo v3.0 BeadChip (31, 32), HumanOmni2.5-Quad BeadChip (33), Human 660W-Quad BeadChip (34), and the Cardio-MetaboChip (35) (**Table S1**). We used the following exclusion criteria to remove samples whose genetic or phenotypic data were questionable: (1) unknown replicates across studies, (2) call rates < 95%, (3) samples with mismatched gender, such as male samples with >10% mean heterozygosity of SNPs on the X chromosome and/or <10% mean intensity of the Y chromosome; or female samples with <15% mean heterozygosity of SNPs on the X chromosome and/or similar mean allele intensities of SNPs on the X and Y chromosomes, (4) ancestry outliers (>4 standard deviations from the mean of the first or second principal component), and (5) first degree relatives.

A subset of 2,717 African Americans (879 CHD cases and 1,838 controls) and 1,184 Japanese Americans (302 CHD cases and 882 controls) genotyped with the Cardio-MetaboChip were missing data for 20 of the 71 SNPs; these subjects were excluded from the risk score analysis.

SNP IMPUTATION AND PRINCIPAL COMPONENTS ANALYSIS

All samples except for the African-American and Japanese samples genotyped with the Metabochip were imputed using the software IMPUTE2, based on build 37 (hg19) coordinates, to the 1000 Genome Project data phase 1 v3. Principal components were calculated by study in smartpca from EIGENSOFT (36) using a random selection of 10,000 SNPs across the genome (MAF >5% and call rate >95%).

STATISTICAL ANALYSIS

The log-additive effect of each SNP on CHD risk was estimated in PLINK using unconditional logistic regression adjusted for age, sex, BMI and the first 10 principal components to account for potential population stratification (37). All analyses were stratified by ethnicity, disease status (i.e., breast cancer, prostate cancer, or type 2 diabetes disease status). METAL was used to combine the results within and across populations, which included 18 case-control strata in the overall meta-analysis of all populations. For SNPs that were imputed, all were imputed with an IMPUTE2 INFO score >0.8 in each study and population. SNPs rs11752643 and rs3782886 in African Americans, rs180803

in Latinos, and rs6544713, rs4252120, rs2023938, rs3918226, rs3184504, and rs9982601 in Japanese had a minor allele frequency less than 1% and were not included in the ethnic-specific analysis. The cross-ethnic meta-analysis was performed on SNPs observed in at least two ethnic groups.

In addition to testing of the index SNP, we examined regional replication of the signal through testing SNPs in linkage disequilibrium (LD) with the index SNP in European ancestry groups ($r^2 \geq 0.4$ in EUR 1000 Genomes Project). Haploview (38) was used to assess pairwise tag SNPs among bins of markers in the AFR population [tagging $r^2 \geq 0.8$ for SNPs with a MAF >1% based on 1000 Genomes Project data (39)]. For each region, an alpha threshold of significance was set at 0.05 divided by the number of tag SNPs in AFR. We considered evidence of replication to be present in a region when one or more SNPs in LD with the index SNP had a *p*-value that was lower than the region-defined alpha threshold. For imputed SNPs, only those imputed with high quality (IMPUTE2 INFO score >0.8) were included in the regional replication testing. The regional association plots were generated with the LocusZoom program (40).

We also examined the aggregate effect of the CHD risk loci. Three genetic risk scores (GRS) were calculated for each individual: (1) An unweighted GRS comprised of risk-associated alleles from the 71 CHD SNPs, (2) a modified unweighted GRS (I) that substitutes the index SNP with the lead SNP reaching region-wide significance within a specific race/ethnic group of each known CHD locus, and (3) a modified unweighted GRS (II) similar to I but substituting index SNPs with the leading SNPs in each region from our cross-ethnic meta-analysis. The risk alleles for the substitution SNPs were determined based on their observed effects in our study. As outlined above, subjects genotyped with the non-GWAS Metabochip were excluded from the risk score analysis because of missing data on 20 SNPs. The risk score distributions across ethnic groups were compared using a two-sided *t*-test. The association of genetic risk scores with CHD were evaluated within each ethnicity in a logistic regression model adjusted for age, sex, BMI, and the first 10 principal components. Of the 71 SNPs selected, only one pair (rs16986953 and rs2123536) from *TTC32-WDR35* was correlated. Since the association between rs2123536 and CHD was only observed in a Chinese population (16), both SNPs were kept in the GRS analysis.

Within each population, statistical power for each SNP was calculated in the R package "gap," (41) using the allele frequency in each racial/ethnic group, and the documented OR from the literature. The allele frequency for the multiethnic sample was weighted by the sample size of each ethnic group. The power for detecting rare and common alleles within each ethnic group was calculated using QUANTO (42).

RESULTS

Descriptive characteristics of the 6,035 CHD cases and 11,251 controls stratified by sex and race/ethnicity are presented in **Table 1**. We analyzed a total of 2,376 African-American cases and

TABLE 1 | Descriptive Characteristics of CHD Cases and Controls

		Study Population (Total N = 17,286, CHD Case N = 6,035, CHD Control N = 11251)			
		African Americans (N = 6,515)			
		Male (N = 2,976)		Female (N = 3,539)	
		CHD cases (N = 1,234)	CHD controls (N = 1,742)	CHD cases (N = 1,142)	CHD controls (N = 2,397)
Ever smoked*	Yes (%)	936 (75.85)	1205 (69.17)	654 (57.27)	1120 (46.73)
	No (%)	289 (23.42)	524 (30.08)	470 (41.15)	1257 (52.44)
BMI	Mean (SD)	27.60 (4.29)	27.42 (4.24)	30.06 (6.10)	28.71 (5.76)
Age at cohort entry	Mean (SD)	63.90 (7.22)	60.16 (8.64)	62.30 (8.05)	57.29 (8.71)
		Latinos (N = 6,109)			
		Male (N = 3,340)		Female (N = 2,769)	
		CHD cases (N = 1,364)	CHD controls (N = 1,976)	CHD cases (N = 927)	CHD controls (N = 1,842)
Ever Smoked*	Yes (%)	976 (71.56)	1306 (66.09)	336 (36.25)	599 (32.52)
	No (%)	357 (26.17)	638 (32.29)	554 (59.76)	1173 (63.68)
BMI	Mean (SD)	27.72 (4.01)	27.43 (3.78)	29.37 (5.96)	27.81 (5.19)
Age at cohort entry	Mean (SD)	62.67 (6.31)	59.62 (6.98)	61.25 (6.49)	57.71 (7.16)
		Japanese Americans (N = 4,662)			
		Male (N = 2,530)		Female (N = 2,132)	
		CHD cases (N = 930)	CHD controls (N = 1,600)	CHD cases (N = 438)	CHD controls (N = 1,694)
Ever Smoked*	Yes (%)	664 (71.40)	1050 (65.63)	138 (31.51)	515 (30.40)
	No (%)	262 (28.17)	541 (33.81)	296 (67.58)	1170 (69.07)
BMI	Mean (SD)	25.50 (3.18)	25.56 (3.53)	24.40 (4.31)	24.25 (4.11)
Age at cohort entry	Mean (SD)	64.30 (6.94)	59.71 (8.76)	62.26 (7.09)	57.20 (8.24)

*Numbers don't total to 100% due to missing data.

4,139 controls, 2,291 Latino cases and 3,818 controls, and 1,368 Japanese cases and 2,294 controls. In general, compared to controls, CHD cases were slightly older at cohort entrance, were heavier in all three ethnic groups and were more likely to have ever smoked than controls in all three ethnic groups (**Table 1**). The associations of BMI and smoking with CHD were similar when further stratified by prevalent conditions, including prostate cancer, breast cancer, and diabetes (**Table S2**).

We had *a priori* greater than 80% power to detect reported per allele effect sizes for 6 out of 71 SNPs in African Americans, 9 out of 71 SNPs in Latinos, and 9 out of 71 SNPs in Japanese Americans and 16 out of 71 SNPs when combining samples from all three ethnic groups (**Figure S1**). Given the sample size in each ethnic group, we had 28.5% power to detect an OR of 1.12 (mean OR from the selected index SNPs) for a rare (MAF = 0.05) allele and 71.6% power to detect the same OR for a common (MAF = 0.20) allele in African Americans; we had 27.3% power to detect OR of 1.12 for a rare allele and 69.3% power for a common allele in Latinos; and 20% power for a rare allele and 52.7% power for a common allele in Japanese Americans.

We examined evidence of replication for 71 CHD variants from 65 regions (**Table S3**). Among these variants, 69 in African Americans, 70 in Latinos, and 65 in Japanese Americans had a MAF >1% and were included in the analysis. Compared to the null expectation that one-half of the examined SNPs show consistent direction of effects as previously reported, 65.2% (45 of 69, binomial p = 0.008) SNPs in African Americans, 70.0% (49 of 70, binomial p = 5.5 × 10^{-4}) in Latinos, 58.5% (38 of 65, binomial p = 0.11) in Japanese, and 77.5% (55 of 71, binomial p = 2.0 × 10^{-6}) in the combined multiethnic sample had the same direction of association as previously reported. In African Americans, nominally statistically significant associations ($p < 0.05$) and consistent directional effects were observed for 11 index SNPs in *PPAP2B, SORT1, IL6R, REST-NOA1, BTN2A1, SLC22A3-LPAL2-LPA, 9p21, CXCL12, SH2B3,*

and *KCNE2*. In Latinos, nominal evidence of association ($p < 0.05$) and consistent directional effects were observed with 8 index SNPs at *SORT1, APOB, NOS3, LPL, ZHF259-APOA5-APOA1, MFGE8-ABHD2, FURIN-FES,* and *BCAS3*. In Japanese, only 1 index SNP at *9p21* was nominally significant and directionally consistent. In the combined multiethnic sample, 10 index SNPs at *PPAP2B, SORT1, IL6R, REST-NOA1, EDNRA, PHACTR1, BTN2A1, NOS3, 9p21,* and *CXCL12* were directionally consistent and nominally statistically significant.

We observed evidence of regional replication for 6 regions in African Americans, 3 in Latinos, 1 in Japanese Americans, and 10 in the combined sample when examining SNPs correlated with the index SNPs (**Table S4**; see Methods). The previously reported index SNP in four of the 10 regions was not significant at the 0.05 level, but correlated SNPs with p-values smaller than the region specific significance levels were detected in these four regions: *SLC22A4-SLC22A5* and *RAI1-PEMT-RASD1* in African Americans, *TTC32-WDR35* in the multiethnic analysis, and *APOE-APOC1* in Latinos and the multiethnic sample.

The most statistically significant association was observed at the *SORT1* locus (**Figure 1**). Two index SNPs in complete LD (rs602633 and rs599839) were initially reported from GWAS in European ancestry populations. The index SNP rs602633 was associated with risk in African Americans (OR = 1.13; p = 0.004), Latinos (OR = 1.11; p = 0.04), and in the cross-ethnic meta-analysis (OR = 1.11; p = 7.8×10^{-4}), but not in Japanese Americans (OR = 1.01, p = 0.88). The most significant association in the region was with variant rs3832016 (OR = 1.16; p = 2.5×10^{-5} in the multiethnic sample), an INDEL (−/T) in high LD with rs602633 in EUR (r^2 = 0.96) and with a MAF of 0.35 in African Americans, 0.20 in Latinos, and 0.07 in Japanese Americans. A previous fine-mapping study of the *SORT1* region at *1p13* implicated a nearby non-coding polymorphism (rs12740374) to be the likely functional variant and to affect lipoprotein metabolism (43). SNP rs12740374 is in high

FIGURE 1 | Results for the *SORT1* region on chromosome 1p13 from the multiethnic meta-analysis. The r^2 shown is for the EUR group in the 1000 Genomes Project relative to the index SNP rs602633. The stronger signal, rs3832016, is also highlighted. This regional association plot was generated with the LocusZoom plot (40).
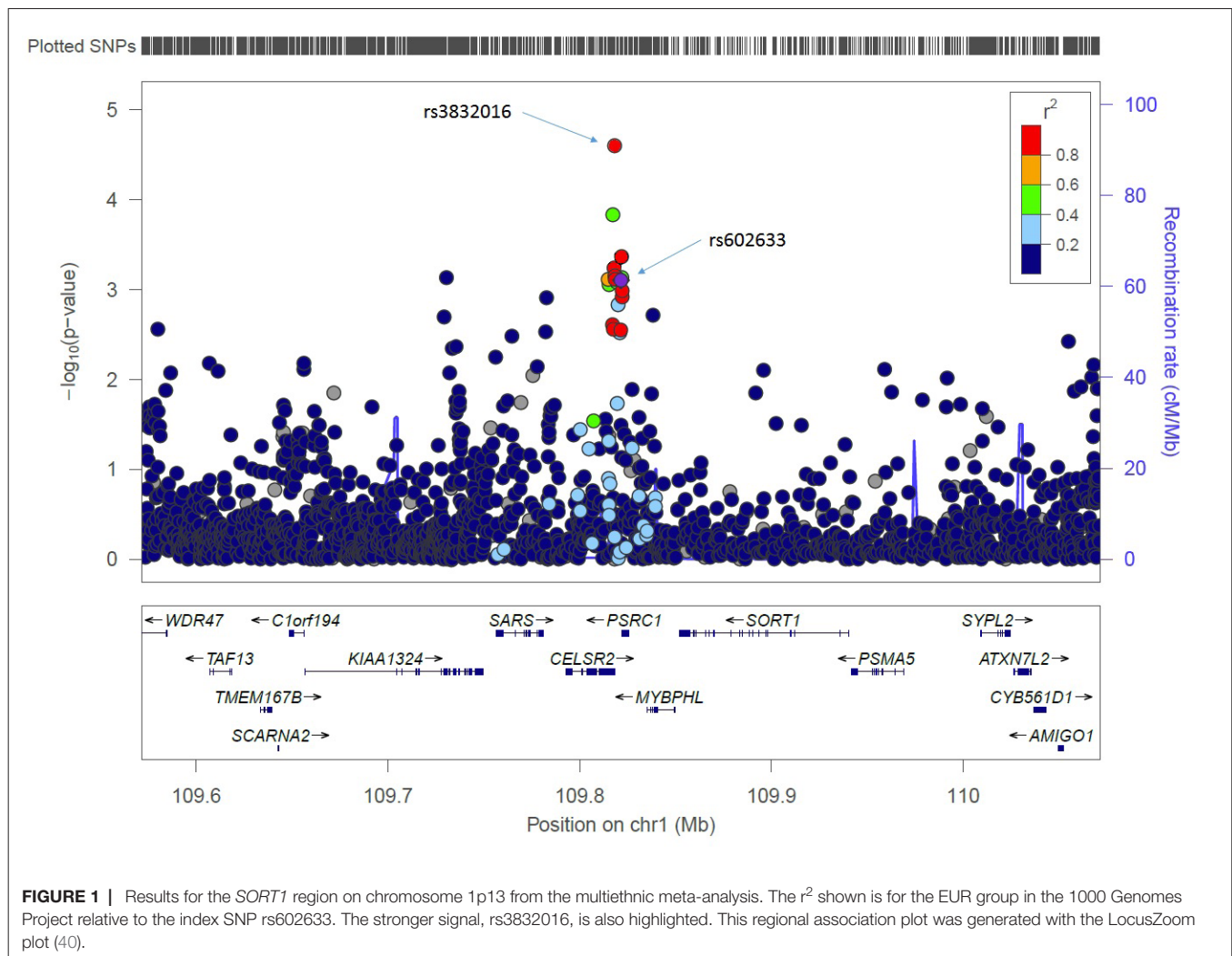
LD not only with the index SNP rs602633 ($r^2 = 0.90$ in EUR) but also with rs3832016 ($r^2 = 0.94$ in EUR). Variant rs12740374 was less strongly associated with risk in the current study ($p = 0.008$ in African Americans with MAF = 0.25, $p = 0.08$ in Latinos with MAF = 0.20, $p = 0.93$ in Japanese Americans with MAF = 0.07, and $p = 0.003$ in the combined multiethnic analysis).

Other regions where evidence of regional replication was observed in African Americans include *PPAP2B* (rs72664341, $p = 0.00018$), *SLC22A4-SLC22A5* (rs17689550, $p = 0.006$), *SLC22A3-LPAL2-LPA* (rs4709431, $p = 0.0077$), *SH2B3* (rs10774625, $p = 0.0047$) and *RAI1-PEMT-RASD1* (rs9899364, $p = 4.5 \times 10^{-4}$). In Latinos, evidence for regional replication was observed at *MFGE8-ABHD2* (rs8037001, $p = 0.0017$), *FURIN-FES* (rs8182016, $p = 1.1 \times 10^{-4}$), and *APOE-APOC1* (rs7412, $p = 0.0043$). In the Japanese, regional replication was only observed at *9p21* (rs10811656, $p = 0.0015$). Five of the 10 regions that replicated in the multiethnic analysis were also significant in ethnic-specific analyses, whereas the remaining 5 regions were detected with significant regional associations in one or more of the ethnic-specific populations alone (*TTC32-WDR35, APOB, EDNRA, PHACTR1* and *BCAS3*; **Table S4**).

Genetic risk scores (GRSs) were used to compare the distribution of genetic risk between populations. Japanese Americans carried, on average, more risk alleles (70.26 ± 4.61, mean ± SD) in comparison to African Americans and Latinos (67.03 ± 4.73 and 68.37 ± 5.11, respectively) (**Table 2**; **Table S5**). The greater number of risk alleles resulted in the distribution of the GRS to be shifted to the right in Japanese Americans compared to African Americans and Latinos (**Figure 2**). The distribution of the GRS was slightly higher in cases than in controls for every group (two-sided t-test, AA $p = 4.4 \times 10^{-5}$, LA $p = 4.5 \times 10^{-7}$, and JA $p = 0.28$). Only minor changes in the distribution of the GRS were noted when we included regionally significant leading SNPs from each ancestry (modified risk score I), or from the cross-ethnic meta-analysis (modified risk score II) (**Table S5**). The average risk scores remained highest in Japanese Americans whereas differences between African Americans and Latinos were reduced, especially when comparing CHD cases from these two ethnic groups (modified risk score II, $p = 0.14$).

The unweighted risk scores were statistically significantly associated with CHD risk in African Americans (per allele OR = 1.03, $p = 4.1 \times 10^{-5}$) and Latinos OR = 1.03, ($p = 2.2 \times$

TABLE 2 | Associations of the genetic risk score with CHD by ethnicity

CHD Risk Score (Mean ± SD, p)		African Americans Case = 67.67 ± 4.63 Control = 67.03 ± 4.73 $p* = 4.4 \times 10^{-5}$		Latinos Case = 69.05 ± 5.13 Control = 68.37 ± 5.11 $p* = 4.5 \times 10^{-7}$		Japanese Americans Case = 70.44 ± 4.60 Control = 70.26 ± 4.61 $p* = 0.28$	
		OR[†]	p[†]	OR[†]	p[†]	OR[†]	p[†]
	Continuous	1.031	4.09E-5	1.031	2.23E-8	1.014	0.11
	Quartile 1	Reference		Reference		Reference	
	Quartile 2	1.084	0.42	1.048	0.55	1.001	0.99
Unweighted risk score	Quartile 3	1.132	0.21	1.190	0.03	1.288	0.02
	Quartile 4	1.397	6.39E-4	1.393	1.98E-5	1.093	0.42
	Continuous	1.028	2.12E-4	1.022	5.27E-5	1.009	0.29
	Quartile 1	Reference		Reference		Reference	
	Quartile 2	1.033	0.75	1.085	0.30	1.107	0.36
	Quartile 3	1.252	0.02	1.205	0.02	1.214	0.08
Modified risk score I[‡]	Quartile 4	1.285	0.01	1.263	0.003	1.081	0.48
	Continuous	1.018	1.71E-3	1.028	9.06E-8	1.016	0.045
	Quartile 1	Reference		Reference		Reference	
	Quartile 2	0.976	0.81	1.066	0.42	1.062	0.59
	Quartile 3	1.278	0.01	1.182	0.03	1.342	0.007
Modified risk score II[§]	Quartile 4	1.280	0.01	1.393	2.0E-5	1.175	0.15

*Two-sided t-test

†Logistic regression model adjusted for age, gender, BMI, and the first 10 principal components

‡Risk score that includes ethnic-specific regional leading SNPs

§Risk score that includes cross-ethnic regional leading SNPs

10^{-8}), but only weakly associated with CHD risk in Japanese Americans (OR = 1.01, p = 0.11) (**Table 2**). When comparing individuals within GRSs in the top quartile to individuals in the bottom quartile, we found both African Americans (OR = 1.40) and Latinos (OR = 1.39) to have a statistically significant ~40% increase in risk (**Table 2**). The analogous risk was lower (~10%) and not significant in Japanese-Americans (OR = 1.09). Results were similar for the modified risk scores (**Table 2**).

To evaluate the effect of existing conditions on the results, we repeated the analysis excluding cancer or diabetes cases; the ORs were comparable to those observed in each ethnic group and in the entire sample (**Table S6**).

A sensitivity analysis was also performed on the selected index SNPs using controls refined to those with medical claims from Medicare or CHDD. Despite the loss of statistical power due to smaller sample size, the effect sizes were comparable to those observed when using the entire control sample (**Table S7**).

DISCUSSION

We evaluated 71 SNPs associated with CHD risk within 65 risk regions in a large multi-ethnic sample of African Americans, Latinos, and Japanese Americans and found that a statistically significant proportion of SNPs exhibited consistent directions of effect beyond the 50% expected by chance. However, only a subset of 11, 8, and 1 of these SNPs were found to be nominally statistically significant in African Americans, Latinos, and Japanese Americans, respectively. Exploration of common genetic variation in these CHD-associated regions provided additional support for

association at 10 regions, with different ethnic-specific or cross-ethnic leading SNPs. These replication results provide additional evidence for shared common genetic effects across ethnicities, with previous studies only replicating signals at *9p21* (21, 24–26), *SORT1* (26), and *ABO* (26) in African Americans, and *SORT1*, *CXCL12*, and *9p21* in Latinos (29). This is the first report of *BTN2A1*, a region initially reported in Japanese, replicating in African Americans.

Japanese Americans had a higher GRS on average when compared to African Americans and Latinos. However, the GRS was more strongly associated with CHD in African Americans and Latinos compared to the Japanese Americans. We note that the genetic markers reported from previous discovery efforts are unlikely to be the functional alleles. The correlation between the index and functional SNPs may vary depending on the LD structure of each ancestral group, which may contribute to the difference in the ethnic-specific odds ratios. In addition to having limited statistical power to replicate associations with index SNPs within and across these populations, differences in LD may serve as an alternative explanation for the lack of replication. In an attempt to address such issues, we conducted regional association testing and constructed modified risk scores incorporating regional association results. When substituting the index SNPs with leading SNPs from the regional analyses, the differences in the modified risk score distributions and per-allele aggregate effects were only modified slightly, but differences were still noted, particularly between the Japanese and the other populations. The reasons for these differences are unclear. Our findings may reflect the severity of subclinical coronary atherosclerosis among Japanese participants in the MEC that is on average greater than the severity observed in Africans and Hispanics (44). Although our analyses are
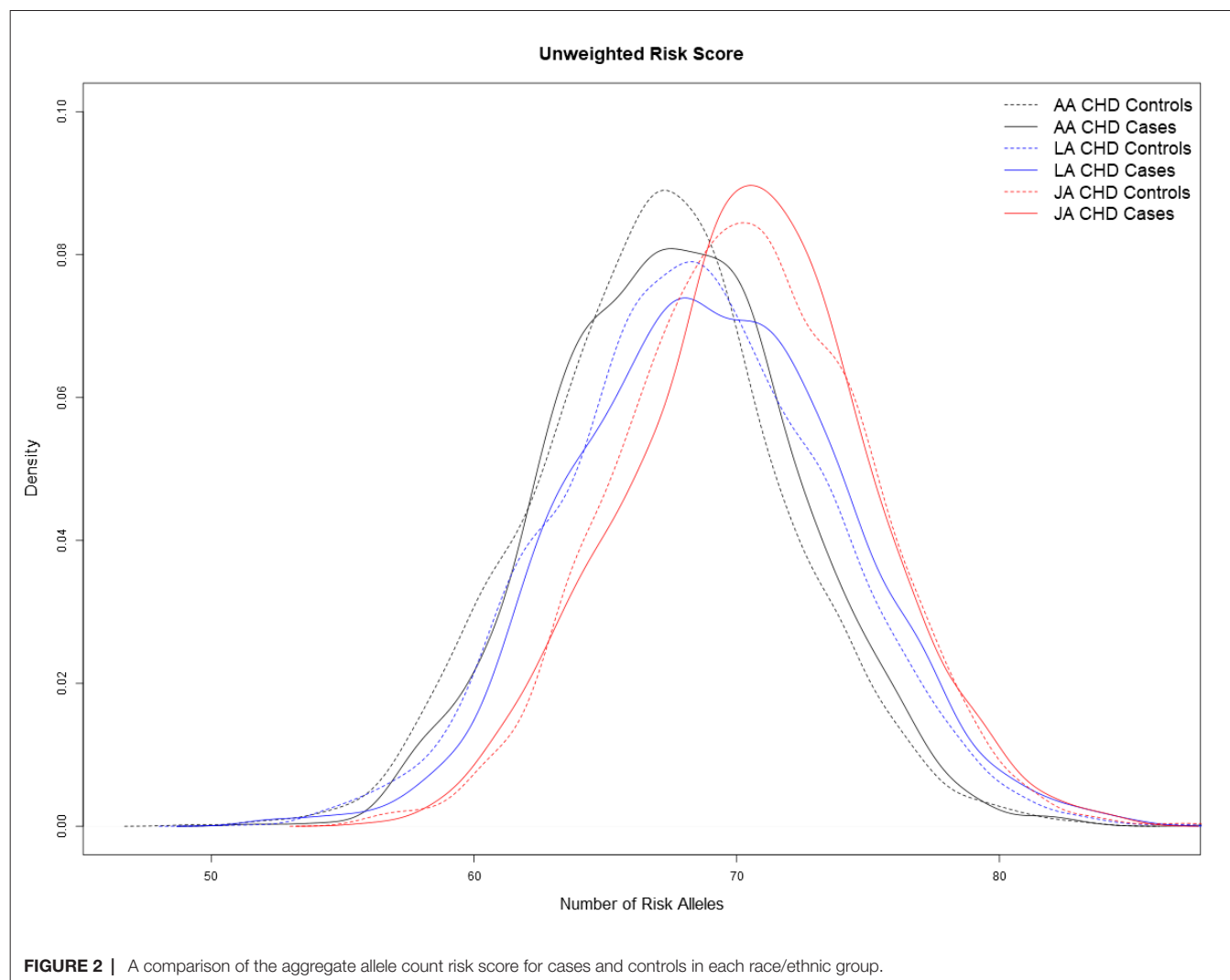
FIGURE 2 | A comparison of the aggregate allele count risk score for cases and controls in each race/ethnic group.

preliminary, we deem it unlikely that these known risk alleles are major contributors to race/ethnic differences in the incidence of CHD, as the incidence of CHD in Japanese is lower than that in the other two groups (45). Given that Japanese had a higher average GRS compared to other ethnic groups, but their population risk is lower, it is possible that functional variants within CHD susceptibility genes not included in our GRS disproportionally affect non-Japanese race/ethnic groups. Alternatively, environmental risk factors such as suboptimal diet and smoking may be less prevalent in Japanese and primarily responsible for the lower rates of CHD despite the higher genetic risk. It is difficult to directly compare the GRS distribution reported in this study to those in studies in European ancestry populations as the methods and number of selected SNPs vary (46–53). The vast majority of studies in European ancestry populations have observed statistically significant per allele relative risks of 1.02–1.12 and relative risks of 1.5–1.9 in comparing the highest versus lowest quintile or quartile of the GRS. Our findings in African Americans and Latinos are generally consistent with these reports albeit smaller effect sizes were noted, perhaps due to differences in LD between the index and functional SNPs.

Our study has a number of limitations. First, the information used to define CHD cases and controls was based on a combination of health care claims data as well as self-report on questionnaires. Some of the Japanese cases from Hawaii may have been missed due to the lack of CHDD records. Of the 1,089 Japanese participants whose CHDD records were available (in California), 426 CHD cases were identified, with 103 classified as cases based solely on CHDD records. Given the same ratio, about 338 Japanese CHD cases from Hawaii where CHDD was not available, may have been misclassified as controls. Assuming an equal distribution of genotypes in these missed cases compared to recognized cases, this misclassification would result in effects being biased towards the null and a reduced power to detect associations. Similar misclassification may apply as Medicare or CHDD data were not available for all controls. In the sensitivity analysis, limiting controls to those with claims data, fewer SNPs reached nominal statistical significance (0.05) however effect sizes were relatively comparable to those observed in the entire control sample. In an attempt to increase specificity when using Medicare and CHDD claims, we only included CHD cases identified from the primary and the first

diagnosis codes, with individuals identified with CHD beyond the primary diagnosis excluded from being a case or a control. The validity of our case and control definitions is indirectly supported by the observed associations of case-control status with known risk factors and by the detection of more directionally consistent genetic associations than expected. Another limitation is the potential inclusion of inappropriately labeled CHD deaths, as CHD is often reported on death certificates when the cause of death is unclear. However, of the 1,005 CHD deaths, 718 also had prior claims data from CHDD or Medicare (**Figure S2**), suggesting a high consistency between mortality records and health care claims.

Another limitation of the study was the selection of CHD cases and controls among MEC participants conditional on three existing medical conditions. However, in sensitivity analyses limited to those without existing conditions, we observed robust consistency in terms of effect size and effect direction between this subset and the entire sample.

We present the largest replication study of established GWAS loci for CHD in Latinos and Japanese Americans conducted to date. However, our power was still limited to detect the originally reported effect sizes even in the combined multiethnic sample. We observed a higher GRS in Japanese Americans compared with African Americans or Latinos but the GRS was paradoxically not significantly associated with CHD risk in Japanese Americans despite observing strong associations in the other two groups. Substantially larger samples that include multiple racial/ethnic groups will help to identify the functional alleles in these regions and characterize their associations with CHD risk and contributions to CHD disparities among ethnically diverse populations.

ETHICS STATEMENT

The MEC study obtained informed consent from study participants and approval from the Health Science Review Board (HSIRB) at the University of Southern California and obtained IRB certification permitting data sharing in accordance with the NIH Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association studies (GWAS).

AUTHOR CONTRIBUTIONS

WK conducted the analysis and wrote the paper. KR contributed to the analysis. DC contributed to the analysis. VS contributed data for the manuscript. DS contributed to the analysis of the manuscript. LW contributed data for the manuscript. LL contributed data for the manuscript. TA contributed data and assisted with the writing of the manuscript. CH oversaw the project, contributed to the data, analysis and writing of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fcvm.2018.00019/full#supplementary-material

TABLE S1 | Genetic arrays by studies.

TABLE S2 | Detailed demographic information on participants by ethnicity, gender, and existing conditions.

TABLE S3 | Associations with CHD SNPs in the MEC populations.

TABLE S4 | Results of regional replication by index or leading SNPs.

TABLE S5 | Comparison of risk score distribution.

TABLE S6 | Sensitivity analysis: comparing subpopulation with/without existing conditions.

TABLE S7 | Sensitivity analysis: associations with CHD SNPs using controls with claims.

FIGURE S1 | Statistical power for known CHD SNPs.

FIGURE S2 | CHD case sources venn diagram.

REFERENCES

1. Selwyn AP, Braunwald E. Ischemic heart disease. In: Braunwald E, Fauci AS, Kasper DL, Hauser SL, Longo DL, Jameson JL et al., editors. *Harrison's Principles of Internal Medicine*. 15th ed. New York: McGraw-Hill (2001). p. 1399–410.
2. Zdravkovic S, Wienke A, Pedersen NL, Marenberg ME, Yashin AI, de Faire U. Heritability of death from coronary heart disease: a 36-year follow-up of 20 966 Swedish twins. *J Intern Med* (2002) 252(3):247–54. doi: 10.1046/j.1365-2796.2002.01029.x
3. 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(5830):1488–91. doi: 10.1126/science.1142447
4. Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, Mayer B, et al. Genomewide association analysis of coronary artery disease. *N Engl J Med* (2007) 357(5):443–53. doi: 10.1056/NEJMoa072366

5. Wild PS, Zeller T, Schillert A, Szymczak S, Sinning CR, Deiseroth A, et al. A genome-wide association study identifies LIPA as a susceptibility gene for coronary artery disease. *Circ Cardiovasc Genet* (2011) 4(4):403–12. doi: 10.1161/CIRCGENETICS.110.958728
6. Schunkert H, König IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet* (2011) 43(4):333–8. doi: 10.1038/ng.784
7. Peden JF, Hopewell JC, Saleheen D, Chambers JC, Hager J, Coronary Artery Disease (C4D) Genetics Consortium. A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. *Nat Genet* (2011) 43(4):339–44. doi: 10.1038/ng.782
8. Erdmann J, Willenborg C, Nahrstaedt J, Preuss M, König IR, Baumert J, et al. Genome-wide association study identifies a new locus for coronary artery disease on chromosome 10p11.23. *Eur Heart J* (2011) 32(2):158–68. doi: 10.1093/eurheartj/ehq405
9. Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, Thompson JR, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet* (2013) 45(1):25–33. doi: 10.1038/ng.2480

10. Davies RW, Wells GA, Stewart AF, Erdmann J, Shah SH, Ferguson JF, et al. A genome-wide association study for coronary artery disease identifies a novel susceptibility locus in the major histocompatibility complex. *Circ Cardiovasc Genet* (2012) 5(2):217–25. doi: 10.1161/CIRCGENETICS.111.961243

11. Vargas JD, Manichaikul A, Wang XQ, Rich SS, Rotter JI, Post WS, et al. Common genetic variants and subclinical atherosclerosis: The Multi-Ethnic Study of Atherosclerosis (MESA). *Atherosclerosis* (2016) 245:230–6. doi: 10.1016/j.atherosclerosis.2015.11.034

12. 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(10):1121–30. doi: 10.1038/ng.3396

13. Ozaki K, Sato H, Inoue K, Tsunoda T, Sakata Y, Mizuno H, et al. SNPs in BRAP associated with risk of myocardial infarction in Asian populations. *Nat Genet* (2009) 41(3):329–33. doi: 10.1038/ng.326

14. Aoki A, Ozaki K, Sato H, Takahashi A, Kubo M, Sakata Y, et al. SNPs on chromosome 5p15.3 associated with myocardial infarction in Japanese population. *J Hum Genet* (2011) 56(1):47–51. doi: 10.1038/jhg.2010.141

15. Wang F, Xu CQ, He Q, Cai JP, Li XC, Wang D, et al. Genome-wide association identifies a susceptibility locus for coronary artery disease in the Chinese Han population. *Nat Genet* (2011) 43(4):345–9. doi: 10.1038/ng.783

16. Lu X, Wang L, Chen S, He L, Yang X, Shi Y, et al. Genome-wide association study in Han Chinese identifies four new susceptibility loci for coronary artery disease. *Nat Genet* (2012) 44(8):890–4. doi: 10.1038/ng.2337

17. Takeuchi F, Yokota M, Yamamoto K, Nakashima E, Katsuya T, Asano H, et al. Genome-wide association study of coronary artery disease in the Japanese. *Eur J Hum Genet* (2012) 20(3):333–40. doi: 10.1038/ejhg.2011.184

18. Assimes TL, Lee IT, Juang JM, Guo X, Wang TD, Kim ET, et al. Genetics of coronary artery disease in Taiwan: a cardiometabochip study by the taichi consortium. *PLoS ONE* (2016) 11(3):e0138014. doi: 10.1371/journal.pone.0138014

19. Lee JY, Lee BS, Shin DJ, Woo Park K, Shin YA, Joong Kim K, et al. A genome-wide association study of a coronary artery disease risk variant. *J Hum Genet* (2013) 58(3):120–6. doi: 10.1038/jhg.2012.124

20. Shen GQ, Li L, Rao S, Abdullah KG, Ban JM, Lee BS, et al. Four SNPs on chromosome 9p21 in a South Korean population implicate a genetic locus that confers high cross-race risk for development of coronary artery disease. *Arterioscler Thromb Vasc Biol* (2008) 28(2):360–5. doi: 10.1161/ATVBAHA.107.157248

21. Assimes TL, Knowles JW, Basu A, Iribarren C, Southwick A, Tang H, et al. Susceptibility locus for clinical and subclinical coronary artery disease at chromosome 9p21 in the multi-ethnic ADVANCE study. *Hum Mol Genet* (2008) 17(15):2320–8. doi: 10.1093/hmg/ddn132

22. Barbalic M, Reiner AP, Wu C, Hixson JE, Franceschini N, Eaton CB, et al. Genome-wide association analysis of incident coronary heart disease (CHD) in African Americans: a short report. *PLoS Genet* (2011) 7(8):e1002199. doi: 10.1371/journal.pgen.1002199

23. Franceschini N, Carty C, Bůzková P, Reiner AP, Garrett T, Lin Y, et al. Association of genetic variants and incident coronary heart disease in multiethnic cohorts: the PAGE study. *Circ Cardiovasc Genet* (2011) 4(6):661–72. doi: 10.1161/CIRCGENETICS.111.960096

24. Kral BG, Mathias RA, Suktitipat B, Ruczinski I, Vaidya D, Yanek LR, et al. A common variant in the CDKN2B gene on chromosome 9p21 protects against coronary artery disease in Americans of African ancestry. *J Hum Genet* (2011) 56(3):224–9. doi: 10.1038/jhg.2010.171

25. Lettre G, Palmer CD, Young T, Ejebe KG, Allayee H, Benjamin EJ, et al. Genome-wide association study of coronary heart disease and its risk factors in 8,090 African Americans: the NHLBI CARe Project. *PLoS Genet* (2011) 7(2):e1001300. doi: 10.1371/journal.pgen.1001300

26. Franceschini N, Hu Y, Reiner AP, Buyske S, Nalls M, Yanek LR, et al. Prospective associations of coronary heart disease loci in African Americans using the MetaboChip: the PAGE study. *PLoS ONE* (2014) 9(12):e113203. doi: 10.1371/journal.pone.0113203

27. Manichaikul A, Wang XQ, Musani SK, Herrington DM, Post WS, Wilson JG, et al. Association of the lipoprotein receptor SCARB1 common missense variant rs4238001 with incident coronary heart disease. *PLoS ONE* (2015) 10(5):e0125497. doi: 10.1371/journal.pone.0125497

28. Zhang L, Buzkova P, Wassel CL, Roman MJ, North KE, Crawford DC, et al. Lack of associations of ten candidate coronary heart disease risk genetic variants and subclinical atherosclerosis in four US populations: the Population Architecture using Genomics and Epidemiology (PAGE) study. *Atherosclerosis* (2013) 228(2):390–9. doi: 10.1016/j.atherosclerosis.2013.02.038

29. Qi L, Campos H. Genetic predictors for cardiovascular disease in hispanics. *Trends Cardiovasc Med* (2011) 21(1):15–20. doi: 10.1016/j.tcm.2012.01.002

30. Han Y, Hazelett DJ, Wiklund F, Schumacher FR, Stram DO, Berndt SI, et al. Integration of multiethnic fine-mapping and genomic annotation to prioritize candidate functional SNPs at prostate cancer susceptibility regions. *Hum Mol Genet* (2015) 24(19):5603–18. doi: 10.1093/hmg/ddv269

31. Feng Y, Stram DO, Rhie SK, Millikan RC, Ambrosone CB, John EM, et al. A comprehensive examination of breast cancer risk loci in African American women. *Hum Mol Genet* (2014) 23(20):5518–26. doi: 10.1093/hmg/ddu252

32. Han Y, Signorello LB, Strom SS, Kittles RA, Rybicki BA, Stanford JL, et al. Generalizability of established prostate cancer risk variants in men of African ancestry. *Int J Cancer* (2015) 136(5):1210–7. doi: 10.1002/ijc.29066

33. Williams AL, Jacobs SB, Moreno-Macías H, Huerta-Chagoya A, Churchhouse C, Márquez-Luna Cet al, et al. Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. *Nature* (2014) 506(7486):97–101. doi: 10.1038/nature12828

34. Siddiq A, Couch FJ, Chen GK, Lindström S, Eccles D, Millikan RC, et al. A meta-analysis of genome-wide association studies of breast cancer identifies two novel susceptibility loci at 6q14 and 20q11. *Hum Mol Genet* (2012) 21(24):5373–84. doi: 10.1093/hmg/dds381

35. Voight BF, Kang HM, Ding J, Palmer CD, Sidore C, Chines PS, et al. The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. *PLoS Genet* (2012) 8(8):e1002793. doi: 10.1371/journal.pgen.1002793

36. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet* (2006) 2(12):e190. doi: 10.1371/journal.pgen.0020190

37. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* (2006) 38(8):904–9. doi: 10.1038/ng1847

38. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* (2005) 21(2):263–5. doi: 10.1093/bioinformatics/bth457

39. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* (2012) 491(7422):56–65. doi: 10.1038/nature11632

40. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* (2010) 26(18):2336–7. doi: 10.1093/bioinformatics/btq419

41. Zhao JH. Gap : genetic analysis package. *J Stat Softw* (2007) 23(8):1–18. doi: 10.18637/jss.v023.i08

42. Gauderman WJ, Morrison JM. (2006) *QUANTO 1.1: A computer program for power and sample size calculations for genetic-epidemiology studies.* http://hydra.usc.edu/gxe

43. Musunuru K, Strong A, Frank-Kamenetsky M, Lee NE, Ahfeldt T, Sachs KV, et al. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature* (2010) 466(7307):714–9. doi: 10.1038/nature09266

44. Salfati E, Nandkeolyar S, Fortmann SP, Sidney S, Hlatky MA, Quertermous T, et al. Susceptibility Loci for clinical coronary artery disease and subclinical coronary atherosclerosis throughout the life-course. *Circ Cardiovasc Genet* (2015) 8(6):803–11. doi: 10.1161/CIRCGENETICS.114.001071

45. Holland AT, Wong EC, Lauderdale DS, Palaniappan LP. Spectrum of cardiovascular diseases in Asian-American racial/ethnic subgroups. *Ann Epidemiol* (2011) 21(8):608–14. doi: 10.1016/j.annepidem.2011.04.004

46. Vaarhorst AA, Lu Y, Heijmans BT, Dollé ME, Böhringer S, Putter H, et al. Literature-based genetic risk scores for coronary heart disease: the Cardiovascular Registry Maastricht (CAREMA) prospective cohort study. *Circ Cardiovasc Genet* (2012) 5(2):202–9. doi: 10.1161/CIRCGENETICS.111.960708

47. Krarup NT, Borglykke A, Allin KH, Sandholt CH, Justesen JM, Andersson EA, et al. A genetic risk score of 45 coronary artery disease risk variants associates with increased risk of myocardial infarction in 6041 Danish individuals. *Atherosclerosis* (2015) 240(2):305–10. doi: 10.1016/j.atherosclerosis.2015.03.022

48. Paynter NP, Chasman DI, Paré G, Buring JE, Cook NR, Miletich JP, et al. Association between a literature-based genetic risk score and cardiovascular events in women. *JAMA* (2010) 303(7):631–7. doi: 10.1001/jama.2010.119

49. Ripatti S, Tikkanen E, Orho-Melander M, Havulinna AS, Silander K, Sharma A, et al. A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses. *Lancet* (2010) 376(9750):1393–400. doi: 10.1016/S0140-6736(10)61267-6

50. Khera AV, Emdin CA, Drake I, Natarajan P, Bick AG, Cook NR, et al. Genetic risk, adherence to a healthy lifestyle, and coronary disease. *N Engl J Med* (2016) 375(24):2349–58. doi: 10.1056/NEJMoa1605086

51. Thanassoulis G, Peloso GM, Pencina MJ, Hoffmann U, Fox CS, Cupples LA, et al. A genetic risk score is associated with incident cardiovascular disease and coronary artery calcium: the Framingham heart study. *Circ Cardiovasc Genet* (2012) 5(1):113–21. doi: 10.1161/CIRCGENETICS.111.961342

52. Brautbar A, Pompeii LA, Dehghan A, Ngwa JS, Nambi V, Virani SS, et al. A genetic risk score based on direct associations with coronary heart disease improves coronary heart disease risk prediction in the Atherosclerosis Risk in Communities (ARIC), but not in the Rotterdam and Framingham offspring, studies. *Atherosclerosis* (2012) 223(2):421–6. doi: 10.1016/j.atherosclerosis.2012.05.035

53. Ganna A, Magnusson PK, Pedersen NL, de Faire U, Reilly M, Arnlöv J, et al. Multilocus genetic risk scores for coronary heart disease prediction. *Arterioscler Thromb Vasc Biol* (2013) 33(9):2267–72. doi: 10.1161/ATVBAHA.113.301218

Integrative Bioinformatics Approaches for Identification of Drug Targets in Hypertension

Daiane Hemerich [1,2], Jessica van Setten [1], Vinicius Tragante [1] and Folkert W. Asselbergs [1,3,4,5]*

[1] Department of Cardiology, University Medical Center Utrecht, University of Utrecht, Utrecht, Netherlands, [2] CAPES Foundation, Ministry of Education of Brazil, Brasília, Brazil, [3] Durrer Center for Cardiovascular Research, Netherlands Heart Institute, Utrecht, Netherlands, [4] Institute of Cardiovascular Science, Faculty of Population Health Sciences, University College London, London, United Kingdom, [5] Farr Institute of Health Informatics Research and Institute of Health Informatics, University College London, London, United Kingdom

High blood pressure or hypertension is an established risk factor for a myriad of cardiovascular diseases. Genome-wide association studies have successfully found over nine hundred loci that contribute to blood pressure. However, the mechanisms through which these loci contribute to disease are still relatively undetermined as less than 10% of hypertension-associated variants are located in coding regions. Phenotypic cell-type specificity analyses and expression quantitative trait loci show predominant vascular and cardiac tissue involvement for blood pressure-associated variants. Maps of chromosomal conformation and expression quantitative trait loci (eQTL) in critical tissues identified 2,424 genes interacting with blood pressure-associated loci, of which 517 are druggable. Integrating genome, regulome and transcriptome information in relevant cell-types could help to functionally annotate blood pressure associated loci and identify drug targets.

Keywords: hypertension, blood pressure, epigenetic regulation, GWAS, data integration, functional annotation, drug target identification.

INTRODUCTION

Elevated blood pressure (BP) or hypertension is a heritable chronic disorder (1–3), considered the single largest contributing risk factor in disease burden and premature mortality (4). High systolic and/or diastolic BP reflects a higher risk of cardiovascular diseases (4). Genome-wide association studies (GWAS) have found association of 905 loci to BP traits (systolic - SBP, diastolic - DBP and pulse pressure -PP) to date (**Table S1**) (5–33). The use of larger sample sizes has helped to identify additional variants, as demonstrated by the most recent study including over 1 million people that has identified 535 novel BP loci (33). Still, this collective effort thus far has not entirely elucidated the complete genetic contribution to BP, estimated to be approximately 50–60% (34).

To add to this complexity, 90.7% of the 905 BP-associated index variants are located in intronic or intergenic regions (**Table S1**). Causal variants are also difficult to pinpoint because of linkage disequilibrium (LD) (35). There is now vast evidence that non-coding variants associated with disease interrupt the action of regulatory elements crucial in relevant tissues for that particular disease (36). BP loci are not only linked to cardiovascular disease but also to other diseases (**Figure 1**), suggesting that BP-associated variants can result in a wide range of phenotypes. Tissue specificity of genetic loci may be relevant for organ specific disease progression. For example, variants altering expression in heart may more likely affect disease progression through heart-
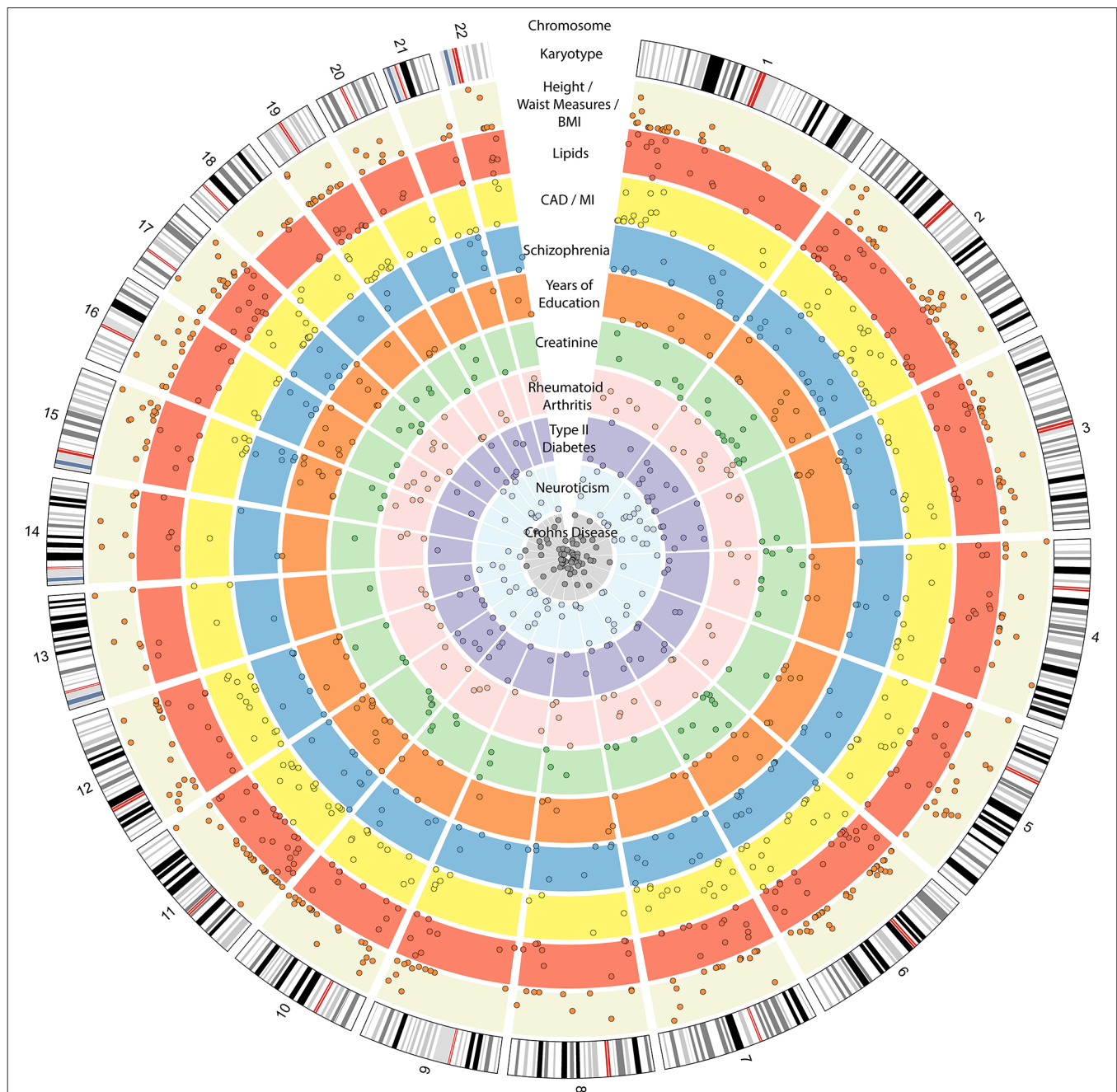
FIGURE 1 | Circos plot showing the 10 traits from the GWAS catalog (37) with the largest number of loci also associated to BP, as identified by PhenoScanner (38) at *p* < 0.05 (**Supplemental Methods**). The outer ring represents the genomic/chromosomal location (hg19). The following inner rings show the associations to different traits. Beige: body measurements (height, body mass index (BMI), weight, waist/hip ratio, hip circumference, waist circumference. *N* = 358). Red: lipids (high-density lipoprotein (HDL), low-density lipoprotein (LDL), triglycerides, total cholesterol. *N* = 226). Yellow: coronary artery disease (CAD)/myocardial infarction (MI) (*N* = 206). Blue: schizophrenia (*N* = 135). Orange: years of education attendance (*N* = 101). Light green: creatinine (*N* = 88). Light pink: rheumatoid arthritis (*N* = 78). Purple: type II diabetes (*N* = 73). Light turquoise: neuroticism (*N* = 69). Light grey: Crohn's disease (*N* = 67).

mediated processes rather than kidney-mediated processes, and some patients may suffer of left ventricular hypertrophy while others may develop nephropathy. Thus, investigating the influence of BP variants in critical cell-types is essential in understanding disease risk and biology, and assessing the possible translation of an associated locus into a drug target. The public availability

of regulatory annotations in several tissues by projects such as ENCODE (39), Roadmap (40) and GTEx (41, 42) has enabled integration of epigenetic modifications, expression quantitative trait loci (eQTLs) and –omics information with GWAS data. Integrative approaches are useful for prioritizing genes from known GWAS loci for functional follow-up, detecting novel

FIGURE 2 | Diagram of analytical steps that can be followed for variant prioritization and translation of association to a potential drug target. Each step is accompanied by examples of publicly available data (green boxes on the left) and tools (yellow boxes on the right) that can be used.
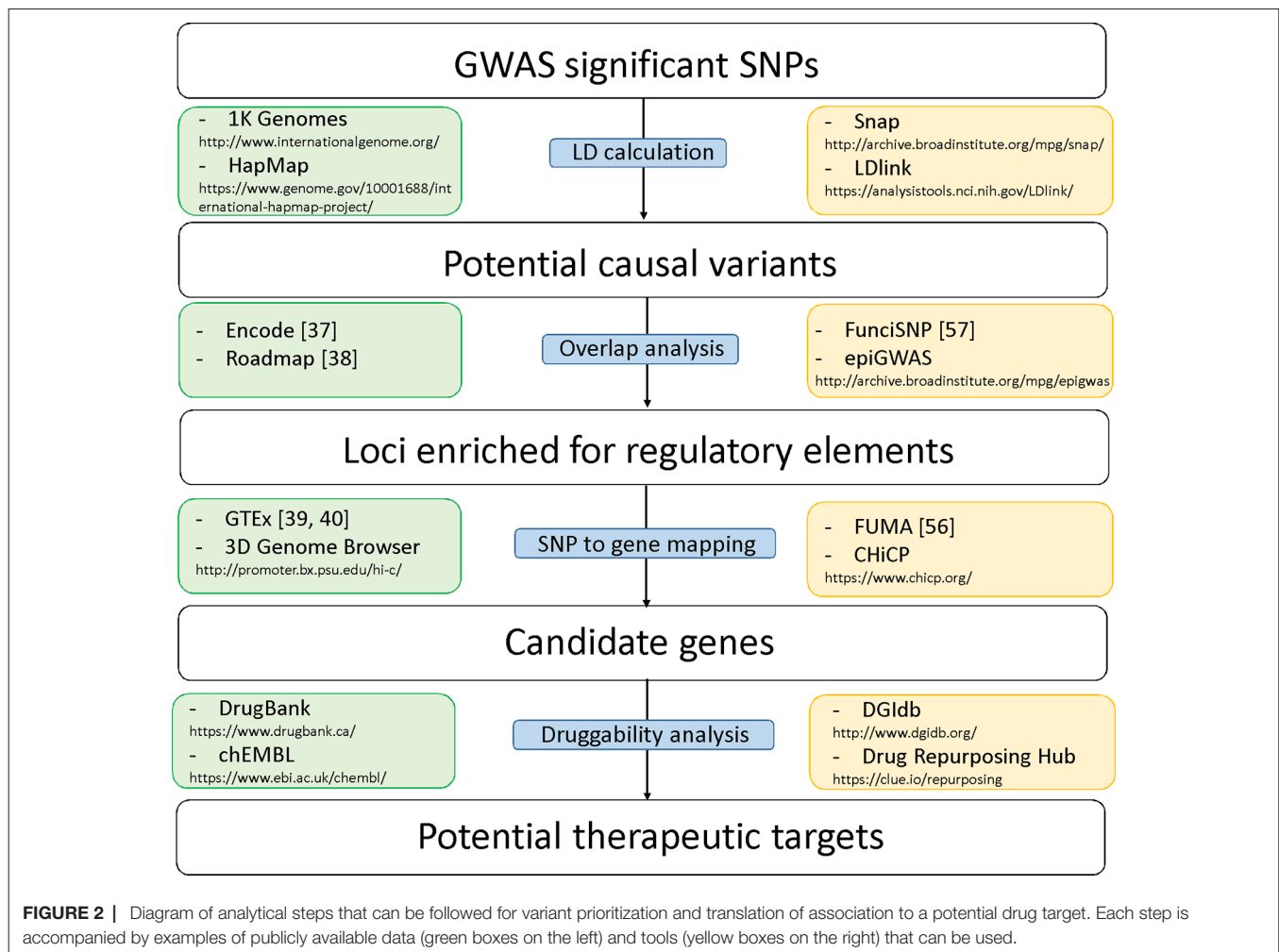
gene-trait associations, inferring the directions of associations, and potential druggability (43–46).

Here we summarize the advances made in recent years towards unraveling the mechanisms of non-coding BP variants in disease progression with the resources mentioned above. We focus on integrative approaches that aim to prioritize BP-associated SNPs located in regulatory regions of the genome for follow-up studies (**Figure 2**). Genetic and molecular aspects of hypertension have been reviewed previously by others (47, 48).

INTEGRATIVE APPROACHES USING – OMICS DATA

Remarkable advances have been made recently towards a better comprehension of BP genetics, the biology of disease and translation towards new therapeutics, boosted by the widespread application of high-throughput genotyping technologies. At the same time, most BP-associated variants are non-coding, making the conversion of statistical associations into target genes a great challenge. SIFT (49, 50), PROVEAN (51), PolyPhen (52), CONDEL (53) and more recently CADD (54) are scoring algorithms developed for

predicting the effect of amino acid changes. Only 98 out of the 905 lead BP-associated SNPs reflect a CADD score above 12.37 (**Table S2**), a threshold suggested by Kicher et al. as deleterious (54). However, the causal variant inside the locus might reflect a different CADD score than the lead SNP, and pinpointing the mechanisms disturbed by the variation remains a challenge.

New strategies that make use of regulatory annotations in disease-relevant tissues have greatly expanded our ability to investigate the processes involved in BP. In particular, annotation of histone modifications and regions of open chromatin allow the identification of active transcription in specific-cell types. Similarly, maps of DNA variants affecting expression in a cell-type specific manner will be integral in BP loci interpretation. A list of cardiovascular-related cell-types researched by the ENCODE Project is presented by Munroe et al. (55). Such data can be integrated with GWAS results using bioinformatics tools (56–58). For instance, FUMA provides extensive functional annotation for all SNPs in associated loci and annotates the identified genes in biological context (57). FunciSNP investigates functional SNPs in regulatory regions of interest (58). Ensemble's Variant Effect Predictor (VEP) determines the effect of variants on genes, transcripts, and protein sequence, as well as regulatory regions,

also outputting SIFT, Polyphen and CADD scores for each variant, among other information (59). Although such integrative tools are useful for variant prioritization and interpretation, not all take into consideration tissue specificity aspects. RegulomeDB, for example, is a database that annotates SNPs with known and predicted regulatory elements in the intergenic regions of the human genome, calculating a score that reflects its evidence for regulatory potential (60). However, the scoring procedure can only be performed across all available tissue types. In addition, several databases containing a broad range of tissues were made publicly available since the last update of RegulomeDB, that could be included in the tool. Together, these resources have been useful in prioritizing genes and variants in associated loci for functional follow-up experiments in many post-GWAS analyses, and can be implemented in interpretation of BP-associated loci.

Transcription Regulation: Histone Modifications and Open Chromatin

As genomic coordinates of active regulatory elements may be mapped using unique functions of chromatin, the characterization of chromatin changes in the genome in specific cell-types can be used to identify DNA variants disturbing active regulatory elements. The four core chromatin histones, H2A, H2B, H3 and H4, can suffer posttranslational modifications, such as acetylation or methylation (61). These histone modifications indicate active (euchromatin)

or repressed (heterochromatin) chromatin structure, defining regulation and gene transcription (62, 63). Acetylation of histones H3 and H4, and H3 methylation at Lys4 (H3K4me3), for instance, correlate with gene transcription, whereas methylation at Lys9 correlates with gene silencing (62, 64). These modifications provide a robust readout of active regulatory positions in the genome, and have been employed for annotation in several studies (23). Histone modifications influencing arterial pressure have been observed in many tissues, including vascular smooth muscle (65). An updated phenotypic cell-type specificity analysis of the 905 BP loci using H3K4me3 mark in 125 tissues is shown in **Figure 3**. The most significant cell-types are cardiovascular-related (**Supplemental Methods**, **Table S3**). Other tissues with high rank in specificity are smooth muscle, fetal adrenal gland, embryonic kidney cells, CD34 and stem-cell derived CD56 +mesoderm cultured cells. These results are consistent with analyses using DNase I hypersensitivity sites (DHSs), which indicate likely binding sites of transcription factors. These results add more evidence that BP loci are enriched on regions of open chromatin (19, 20, 23, 33) (**Figure S1**), regulating transcription in a broad range of tissues.

Methylation

In addition to histone modifications that promote transcription, BP loci have also been studied for their enrichment on DNA methylation, known to have the opposite regulatory effect. The

FIGURE 3 | Ranked tissues after phenotypic cell-type specificity analysis of 905 BP SNPs using 125 H3K4me3 datasets on human tissue (**Supplemental Methods**, **Table S3**).

FIGURE 4 | Diagram illustrating the results of our integrative approach.
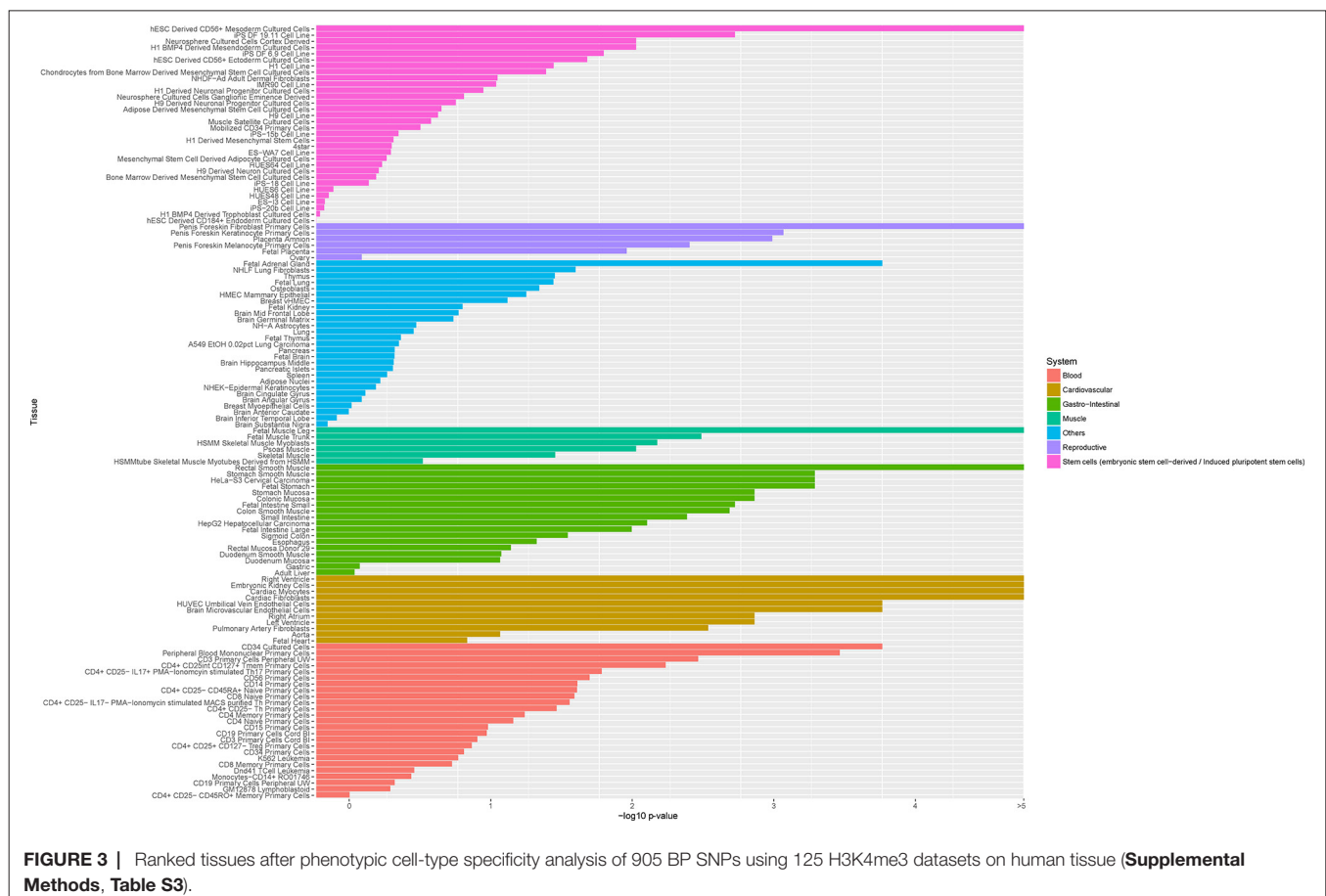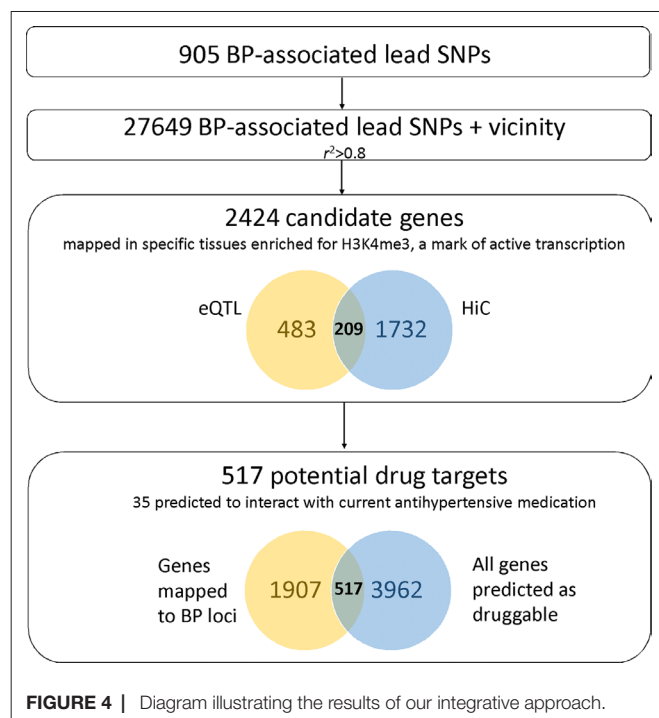
methylation of CpG sites, presented by CpG islands in promoters, affects binding of transcription factors, resulting in gene silencing (66, 67). Abnormal CpG methylation is found in hypertension (68–70), and in many other complex diseases (71, 72). Recently, Kato et al. identified a ~2 fold enrichment associating BP variants and local DNA methylation (19). The study also demonstrates that DNA methylation in blood correlates with methylation in several other tissues. These observations add to previous indications on the function of DNA methylation in regulating BP.

Measuring the Impact of BP Risk Alleles on Gene Expression: eQTLs

Expression quantitative trait loci (eQTL) are regions harbouring nucleotides correlating with alterations in gene expression (73). Linking transcription levels to complex traits has been a follow-up step adopted by many studies (43, 74–76), driven by the increase in available data of expression patterns across tissues and populations (33, 46, 77–81). Warren et al. found that 55.1% of their identified BP-associated loci have SNPs with eQTLs in at least one tissue from GTex repository (41), with arterial tissue most frequently observed (29.9% of loci had eQTL in aorta and/or tibial artery) (21). A great enrichment of eQTLs in artery was also observed by Evangelou et al., who identified 92 novel loci with eQTL enrichment in arterial tissue and 48 in adrenal tissue (33). In summary, these studies also suggest that BP loci exert a regulatory effect mostly in vascular and cardiac tissues.

Finding the Targets: Chromosome Confirmation Capture Techniques

Mapping variation to target genes is one of the greatest challenges in the post-GWAS era, and different strategies have been developed to

this end (82). One approach is the use of chromosome confirmation capture [3C (83), 4C (84, 85), Hi-C (86, 87)]. These techniques capture chromosome interactions (88), resulting in networks of interacting genetic loci (84, 85).

Warren et al. made use of this resource to investigate the target genes of non-coding SNPs, using Hi-C data from endothelial cells (HUVECs). Distal potential genes were found on 21 loci, and these genes were enriched for regulators of cardiac hypertrophy in pathway analysis (20). Kraja et al. also explored long-range chromatin interactions using endothelial precursor cell Hi-C data (89, 90), finding the link between an associated loci and a gene known to affect cell growth and death (91). More recently, Evangelou et al. used chromatin interaction Hi-C data from HUVECs (92), neural progenitor cells (NPC), mesenchymal stem cells (MSC) and tissue from the aorta and adrenal gland (93) to identify distal affected genes. They found 498 novel loci that contained a potential regulatory SNP, and in 484 loci long-range interactions were found in at least one cell-type (33).

A list of human HiC data available on BP relevant tissues is presented in **Table S4**. An updated version of variant to gene mapping making use of this chromatin conformation data is shown in **Table S5**. Promoter regions of 1,941 genes were found to interact with the 27,649 candidate SNPs (905 BP associated SNPs and vicinity) (Supplemental Methods, **Figure 4**). Integration with eQTL data on relevant tissues confirmed 209 of the genes mapped, and added additional 483 genes. One main goal of understanding biological mechanisms of GWAS associations and affected genes is to be able to therapeutically target them. Assessment of the druggability of a BP-associated locus depends on several factors, but overlap of these results with a recent effort on druggability suggests that 517 of these 2,424 genes are druggable (94), and 35 mapped genes are also predicted to interact with common drugs for treatment of hypertension (**Table S2**, **Figure 4**, **Supplemental Methods**). Interestingly, 1,774 of the genes mapped are physically located outside BP-associated loci. These results support the hypothesis that BP GWAS loci act on tissue specific regulatory gene networks. Importantly, they also show that the use of long range chromatin interaction maps can reliably identify target genes even outside the risk locus.

DISCUSSION AND CONCLUSIONS

GWAS have pinpointed over 900 loci associated with BP, and increasing sample size has shown to be crucial to identify more signals (33). However, efforts are needed to translate these results into biological inferences on causal mechanisms and understanding of disease biology. The integration of data beyond the DNA sequence is crucial to identify genes involved in BP regulated by epigenetic mechanisms.

BP variants show eQTL, histone modification and open chromatin enrichment in a broad range of tissues, mostly vascular and cardiac-related. As the interplay of regulatory elements is highly cell-type specific, the study of changes that influence chromatin structure and accessibility needs to be extended to a broad range of tissues and conditions, including disease and its stages. Rosa-Garrido et al. observed chromatin structural abnormalities when comparing healthy and diseased cardiac myocytes, concluding

that heart failure involves altered enhancer-gene interactions (95). Thus, alterations in chromatin structure underlying heart disease perturbs significant interactions that contribute to gene expression. This finding suggests that high resolution chromatin conformation and epigenetic data in disease state can help in understanding how regulatory variants confer risk to disease. The availability of data in different populations will also allow fine-mapping and functional annotation across ethnic groups.

By mapping of BP-associated variants to genes using maps of chromosomal conformation in specific cell-types, we identified 1,941 genes, of which 209 show supported by eQTL mapping. Of all genes mapped (n = 2,424), 517 are predicted as druggable and 35 are predicted to interact with common antihypertensive drugs. These include successful cases such as *APOB* gene, predicted to be targeted by Ibersartan, an angiotensin II receptor antagonist used mainly for the treatment of hypertension (96). Interestingly, in this analysis we were also able to identify *ABCC9* gene on both eQTL and HiC mapping, a gene that interacts with Minoxidil. Although originally developed as an antihypertensive vasodilator, side effects provided limitations and currently its main application occurs topically for treatment of hair loss (97, 98). This highlights the several factors involved in druggability of a target and need for extensive validation and trials. With *in-silico* experimental evidence supporting a plausible mechanism for association, definitive assignment of functions to putative cis-regulatory elements requires perturbation of these elements. Although the majority of associated variants add only modest effects on risk, more studies suggest combinations of SNPs are frequently necessary in order to explain these effects (99–101). CRISPR–Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats) editing technology (102) permits targeted manipulation of epigenetic mechanisms linked to risk alleles (103). Finally, genes that show consequent differential expression can be further validated *in vivo* with the use of animal models.

In summary, the integrative approaches presented in this review help understanding the underlying biology of GWAS loci by mapping SNPs to genes and determine cell and tissue-specificity. The increase in availability of regulatory data in a broad range of tissues and disease states will expand the possibilities for integration and interpretation of association results. Studies validating the genes prioritized may identify new drug targets, enabling more effective prevention and treatment of hypertension and its consequences.

AUTHOR CONTRIBUTIONS

DH, VT and FA contributed in study conception and design. DH was responsible for analysis and interpretation of data and drafting of manuscript. DH, VT, JS and FA provided critical revision and final approval of the manuscript.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fcvm.2018.00025/full#supplementary-material

REFERENCES

1. Muñoz M, Pong-Wong R, Canela-Xandri O, Rawlik K, Haley CS, Tenesa A. Evaluating the contribution of genetics and familial shared environment to common disease using the UK Biobank. *Nat Genet* (2016) 48(9):980–3. doi: 10.1038/ng.3618
2. Feinleib M, Garrison RJ, Fabsitz R, Christian JC, Hrubec Z, Borhani NO, et al. The NHLBI twin study of cardiovascular disease risk factors: methodology and summary of results. *Am J Epidemiol* (1977) 106(4):284–95. doi: 10.1093/oxfordjournals.aje.a112464
3. Mongeau JG, Biron P, Sing CF. The influence of genetics and household environment upon the variability of normal blood pressure: the montreal adoption survey. *Clin Exp Hypertens A* (1986) 8(4-5):653–60. doi: 10.3109/10641968609046581
4. Bromfield S, Muntner P. High blood pressure: the leading global burden of disease risk factor and the need for worldwide prevention programs. *Curr Hypertens Rep* (2013) 15(3):134–6. doi: 10.1007/s11906-013-0340-9
5. Levy D, Ehret GB, Rice K, Verwoert GC, Launer LJ, Dehghan A, et al. Genome-wide association study of blood pressure and hypertension. *Nat Genet* (2009) 41(6):677–87. doi: 10.1038/ng.384
6. Newton-Cheh C, Larson MG, Vasan RS, Levy D, Bloch KD, Surti A, et al. Association of common variants in NPPA and NPPB with circulating natriuretic peptides and blood pressure. *Nat Genet* (2009) 41(3):348–53. doi: 10.1038/ng.328

7. Talmud PJ, Drenos F, Shah S, Shah T, Palmen J, Verzilli C, et al. *Gene-centric association signals for lipids and apolipoproteins identified via the HumanCVD BeadChip*. *Am J Hum Genet* (2009) 85(5):628–42. doi: 10.1016/j.ajhg.2009.10.014
8. Padmanabhan S, Melander O, Johnson T, Di Blasio AM, Lee WK, Gentilini D, et al. Genome-wide association study of blood pressure extremes identifies variant near UMOD associated with hypertension. *PLoS Genet* (2010) 6(10):e1001177. doi: 10.1371/journal.pgen.1001177
9. Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, Chasman DI, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* (2011) 478(7367):103–9. doi: 10.1038/nature10405
10. Franceschini N, Reiner AP, Heiss G. Recent findings in the genetics of blood pressure and hypertension traits. *Am J Hypertens* (2011) 24(4):392–400. doi: 10.1038/ajh.2010.218
11. IBC 50K CAD Consortium. Large-scale gene-centric analysis identifies novel variants for coronary artery disease. *PLoS Genet* (2011) 7(9):e1002260.
12. Johnson AD, Newton-Cheh C, Chasman DI, Ehret GB, Johnson T, Rose L, et al. Association of hypertension drug target genes with blood pressure and hypertension in 86,588 individuals. *Hypertension* (2011) 57(5):903–10. doi: 10.1161/HYPERTENSIONAHA.110.158667
13. Johnson T, Gaunt TR, Newhouse SJ, Padmanabhan S, Tomaszewski M, Kumari M, et al. Blood pressure loci identified with a gene-centric array. *Am J Hum Genet* (2011) 89(6):688–700. doi: 10.1016/j.ajhg.2011.10.013

14. Padmanabhan S, Newton-Cheh C, Dominiczak AF. Genetic basis of blood pressure and hypertension. *Trends Genet* (2012) 28(8):397–408. doi: 10.1016/j.tig.2012.04.001

15. Franceschini N, Fox E, Zhang Z, Edwards TL, Nalls MA, Sung YJ, et al. Genome-wide association analysis of blood-pressure traits in African-ancestry individuals reveals common associated genes in African and non-African populations. *Am J Hum Genet* (2013) 93(3):545–54. doi: 10.1016/j.ajhg.2013.07.010

16. Ganesh SK, Tragante V, Guo W, Guo Y, Lanktree MB, Smith EN, et al. Loci influencing blood pressure identified using a cardiovascular gene-centric array. *Hum Mol Genet* (2013) 22(8):1663–78. doi: 10.1093/hmg/dds555

17. 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(3):349–60. doi: 10.1016/j.ajhg.2013.12.016

18. Surendran P, Drenos F, Young R, Warren H, Cook JP, Manning AK, et al. Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. *Nat Genet* (2016) 48(10):1151–61. doi: 10.1038/ng.3654

19. Kato N, Loh M, Takeuchi F, Verweij N, Wang X, Zhang W, et al. Trans-ancestry genome-wide association study identifies 12 genetic loci influencing blood pressure and implicates a role for DNA methylation. *Nat Genet* (2015) 47(11):1282–93. doi: 10.1038/ng.3405

20. Warren HR, Evangelou E, Cabrera CP, Gao H, Ren M, Mifsud B, et al. Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. *Nat Genet* (2017) 49(3):403–15. doi: 10.1038/ng.3768

21. Hoffmann TJ, Ehret GB, Nandakumar P, Ranatunga D, Schaefer C, Kwok PY, et al. Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. *Nat Genet* (2017) 49(1):54–64. doi: 10.1038/ng.3715

22. Liu C, Kraja AT, Smith JA, Brody JA, Franceschini N, Bis JC, et al. Meta-analysis identifies common and rare variants influencing blood pressure and overlapping with metabolic trait loci. *Nat Genet* (2016) 48(10):1162–70. doi: 10.1038/ng.3660

23. Ehret GB, Ferreira T, Chasman DI, Jackson AU, Schmidt EM, Johnson T, et al. The genetics of blood pressure regulation and its target organs from association studies in 342,415 individuals. *Nat Genet* (2016) 48(10):1171–84. doi: 10.1038/ng.3667

24. Ganesh SK, Chasman DI, Larson MG, Guo X, Verwoert G, Bis JC, et al. *Effects of long-term averaging of quantitative blood pressure traits on the detection of genetic associations. Am. J. Hum. Genet.* (2014) 95(1):49–65. doi: 10.1016/j.ajhg.2014.06.002

25. Kato N, Takeuchi F, Tabara Y, Kelly TN, Go MJ, Sim X, et al. Meta-analysis of genome-wide association studies identifies common variants associated with blood pressure variation in east Asians. *Nat Genet* (2011) 43(6):531–8. doi: 10.1038/ng.834

26. Wang Y, O'Connell JR, Mcardle PF, Wade JB, Dorff SE, Shah SJ, et al. From the Cover: Whole-genome association study identifies STK39 as a hypertension susceptibility gene. *Proc Natl Acad Sci USA* (2009) 106(1):226–31. doi: 10.1073/pnas.0808358106

27. Wain LV, Verwoert GC, O'Reilly PF, Shi G, Johnson T, Johnson AD. et al. Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. *Nat Genet* (2011) 43(10):1005–11. doi: 10.1038/ng.922

28. Simino J, Shi G, Bis JC, Chasman DI, Ehret GB, Gu Xet al. Gene-age interactions in blood pressure regulation: a large-scale investigation with the CHARGE, Global BPgen, and ICBP Consortia. *Am J Hum Genet* (2014) 95(1):24–38. doi: 10.1016/j.ajhg.2014.05.010

29. Fox ER, Young JH, Li Y, Dreisbach AW, Keating BJ, Musani SK. et al. Association of genetic variation with systolic and diastolic blood pressure among African Americans: the candidate gene association resource study. *Hum Mol Genet* (2011) 20(11):2273–84. doi: 10.1093/hmg/ddr092

30. Zhu X, Feng T, Tayo BO, Liang J, Young JH, Franceschini Net al. Meta-analysis of correlated traits via summary statistics from GWASs with an application in hypertension. *Am J Hum Genet* (2015) 96(1):21–36. doi: 10.1016/j.ajhg.2014.11.011

31. Ho JE, Levy D, Rose L, Johnson AD, Ridker PM, Chasman DI. Discovery and replication of novel blood pressure genetic loci in the Women's Genome Health Study. *J Hypertens* (2011) 29(1):62–9. doi: 10.1097/HJH.0b013e3283406927

32. Parmar PG, Taal HR, Timpson NJ, Thiering E, Lehtimäki T, Marinelli M, et al. International genome-wide association study consortium identifies novel loci associated with blood pressure in children and adolescents. *Circ Cardiovasc Genet* (2016) 9(3):266–78. doi: 10.1161/CIRCGENETICS.115.001190

33. Evangelou E. Genetic analysis of over one million people identifies 535 novel loci for blood pressure. *bioRxiv* (2017).

34. Kupper N, Willemsen G, Riese H, Posthuma D, Boomsma DI, de Geus EJ. Heritability of daytime ambulatory blood pressure in an extended twin design. *Hypertension* (2005) 45(1):80–5. doi: 10.1161/01.HYP.0000149952.84391.54

35. Cowper-Sal Lari R, Zhang X, Wright JB, Bailey SD, Cole MD, Eeckhoute J, et al. Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression. *Nat Genet* (2012) 44(11):1191–8. doi: 10.1038/ng.2416

36. Trynka G, Sandor C, Han B, Xu H, Stranger BE, Liu XS, et al. Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat Genet* (2013) 45(2):124–30. doi: 10.1038/ng.2504

37. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci USA* (2009) 106(23):9362–7. doi: 10.1073/pnas.0903103106

38. Staley JR, Blackshaw J, Kamat MA, Ellis S, Surendran P, Sun BB, et al. PhenoScanner: a database of human genotype-phenotype associations. *Bioinformatics* (2016) 32(20):3207–9. doi: 10.1093/bioinformatics/btw373

39. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* (2012) 489(7414):57–74. doi: 10.1038/nature11247

40. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A. et al. Integrative analysis of 111 reference human epigenomes. *Nature* (2015) 518(7539):317–30. doi: 10.1038/nature14248

41. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* (2013) 45(6):580–5. doi: 10.1038/ng.2653

42. Carithers LJ, Moore HM. The Genotype-Tissue Expression (GTEx) project. *Biopreserv Biobank* (2015) 13(5):307–8. doi: 10.1089/bio.2015.29031.hmm

43. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet* (2016) 48(5):481–7. doi: 10.1038/ng.3538

44. Pickrell JK, Berisa T, Liu JZ, Ségurel L, Tung JY, Hinds DA. Detection and interpretation of shared genetic influences on 42 human traits. *Nat Genet* (2016) 48(7):709–17. doi: 10.1038/ng.3570

45. Visscher PM, Yang J. A plethora of pleiotropy across complex traits. *Nat Genet* (2016) 48(7):707–8. doi: 10.1038/ng.3604

46. Battle A, Brown CD, Engelhardt BE, Montgomery SB, GTEx Consortium Laboratory. Genetic effects on gene expression across human tissues. *Nature* (2017) 550(7675):204–13. doi: 10.1038/nature24277

47. Padmanabhan S, Caulfield M, Dominiczak AF. Genetic and molecular aspects of hypertension. *Circ Res* (2015) 116(6):937–59. doi: 10.1161/CIRCRESAHA.116.303647

48. Wang X, Prins BP, Sõber S, Laan M, Snieder H. Beyond genome-wide association studies: new strategies for identifying genetic determinants of hypertension. *Curr Hypertens Rep* (2011) 13(6):442–51. doi: 10.1007/s11906-011-0230-y

49. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* (2009) 4(7):1073–81. doi: 10.1038/nprot.2009.86

50. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* (2003) 31(13):3812–4. doi: 10.1093/nar/gkg509

51. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. *PLoS ONE* (2012) 7(10):e46688. doi: 10.1371/journal.pone.0046688

52. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat Methods* (2010) 7(4):248–9. doi: 10.1038/nmeth0410-248

53. González-Pérez A, López-Bigas N. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *Am J Hum Genet* (2011) 88(4):440–9. doi: 10.1016/j.ajhg.2011.03.004

54. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* (2014) 46(3):310–5. doi: 10.1038/ng.2892

55. Munroe PB, Barnes MR, Caulfield MJ. Advances in blood pressure genomics. *Circ Res* (2013) 112(10):1365–79. doi: 10.1161/CIRCRESAHA.112.300387

56. Tak YG, Farnham PJ. Making sense of GWAS: using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome. *Epigenetics Chromatin* (2015) 8:57. doi: 10.1186/s13072-015-0050-4

57. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* (2017) 8(1):p. 1826. doi: 10.1038/s41467-017-01261-5

58. Coetzee SG, Rhie SK, Berman BP, Coetzee GA, Noushmehr H. FunciSNP: an R/bioconductor tool integrating functional non-coding data sets with genetic association studies to identify candidate regulatory SNPs. *Nucleic Acids Res* (2012) 40(18):e139. doi: 10.1093/nar/gks542

59. Mclaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl Variant Effect Predictor. *Genome Biol* (2016) 17(1):122. doi: 10.1186/s13059-016-0974-4

60. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* (2012) 22(9):1790–7. doi: 10.1101/gr.137323.112

61. Marmorstein R, Trievel RC. Histone modifying enzymes: structures, mechanisms, and specificities. *Biochim Biophys Acta* (2009) 1789(1):58–68. doi: 10.1016/j.bbagrm.2008.07.009

62. Jenuwein T, Allis CD. Translating the histone code. *Science* (2001) 293(5532):1074–80. doi: 10.1126/science.1063127

63. Sullivan BA, Karpen GH. Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin. *Nat Struct Mol Biol* (2004) 11(11):1076–83. doi: 10.1038/nsmb845

64. Fischle W, Wang Y, Allis CD. Binary switches and modification cassettes in histone biology and beyond. *Nature* (2003) 425(6957):475–9. doi: 10.1038/nature02017

65. Wise IA, Charchar FJ. Epigenetic Modifications in Essential Hypertension. *Int J Mol Sci* (2016) 17(4):451. doi: 10.3390/ijms17040451

66. Holliday R, Pugh JE. DNA modification mechanisms and gene activity during development. *Science* (1975) 187(4173):226–32. doi: 10.1126/science.1111098

67. Riggs AD. X inactivation, differentiation, and DNA methylation. *Cytogenet Cell Genet* (1975) 14(1):9–25. doi: 10.1159/000130315

68. Frey FJ. Methylation of CpG islands: potential relevance for hypertension and kidney diseases. *Nephrol Dial Transplant* (2005) 20(5):868–9. doi: 10.1093/ndt/gfh746

69. Smolarek I, Wyszko E, Barciszewska AM, Nowak S, Gawronska I, Jablecka A, et al. Global DNA methylation changes in blood of patients with essential hypertension. *Med Sci Monit* (2010) 16(3):CR149–55.

70. Raftopoulos L, Katsi V, Makris T, Tousoulis D, Stefanadis C, Kallikazaros I. Epigenetics, the missing link in hypertension. *Life Sci* (2015) 129:22–6. doi: 10.1016/j.lfs.2014.08.003

71. Muhonen P, Holthofer H. Epigenetic and microRNA-mediated regulation in diabetes. *Nephrol Dial Transplant* (2009) 24(4):1088–96. doi: 10.1093/ndt/gfn728

72. Mill J, Tang T, Kaminsky Z, Khare T, Yazdanpanah S, Bouchard L, et al. Epigenomic profiling reveals DNA-methylation changes associated with major psychosis. *Am J Hum Genet* (2008) 82(3):696–711. doi: 10.1016/j.ajhg.2008.01.008

73. Albert FW, Kruglyak L. The role of regulatory variation in complex traits and disease. *Nat Rev Genet* (2015) 16(4):197–212. doi: 10.1038/nrg3891

74. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BW, et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet* (2016) 48(3):245–52. doi: 10.1038/ng.3506

75. Musunuru K, Strong A, Frank-Kamenetsky M, Lee NE, Ahfeldt T, Sachs KV, et al. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature* (2010) 466(7307):714–9. doi: 10.1038/nature09266

76. Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJet al. A gene-based association method for mapping traits using reference transcriptome data. *Nat. Genet.* (2015) 47(9):1091–8. doi: 10.1038/ng.3367

77. Xia K, Shabalin AA, Huang S, Madar V, Zhou YH, Wang W, et al. seeQTL: a searchable database for human eQTLs. *Bioinformatics* (2012) 28(3):451–2. doi: 10.1093/bioinformatics/btr678

78. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* (2015) 348(6235):648–60. doi: 10.1126/science.1262110

79. Yang TP, Beazley C, Montgomery SB, Dimas AS, Gutierrez-Arcelus M, Stranger BE, et al. Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. *Bioinformatics* (2010) 26(19):2474–6. doi: 10.1093/bioinformatics/btq452

80. Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* (2013) 45(10):1238–43. doi: 10.1038/ng.2756

81. Lappalainen T, Sammeth M, Friedländer MR, 't Hoen PA, Monlong J, Rivas MAet al. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* (2013) 501(7468):506–11. doi: 10.1038/nature12531

82. Spain SL, Barrett JC. Strategies for fine-mapping complex traits. *Hum Mol Genet* (2015) 24(R1):R111–R119. doi: 10.1093/hmg/ddv260

83. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. *Science* (2002) 295(5558):1306–11. doi: 10.1126/science.1067799

84. Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, de Wit E, et al. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet* (2006) 38(11):1348–54. doi: 10.1038/ng1896

85. Zhao Z, Tavoosidana G, Sjölinder M, Göndör A, Mariano P, Wang S, et al. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat Genet* (2006) 38(11):1341–7. doi: 10.1038/ng1891

86. Belton JM, Mccord RP, Gibcus JH, Naumova N, Zhan Y, Dekker J. Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods* (2012) 58(3):268–76. doi: 10.1016/j.ymeth.2012.05.001

87. van Berkum NL, Lieberman-Aiden E, Williams L, Imakaev M, Gnirke A, Mirny LA, et al. Hi-C: a method to study the three-dimensional architecture of genomes. *J Vis Exp* (2010) (39):1869. doi: 10.3791/1869

88. Rivera CM, Ren B. Mapping human epigenomes. *Cell* (2013) 155(1):39–55. doi: 10.1016/j.cell.2013.09.011

89. Javierre BM, Burren OS, Wilder SP, Kreuzhuber R, Hill SM, Sewitz Set al. Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell* (2016) 167(5):1369–84. doi: 10.1016/j.cell.2016.09.037

90. Schofield EC, Carver T, Achuthan P, Freire-Pritchett P, Spivakov M, Todd JA, et al. CHiCP: a web-based tool for the integrative and interactive visualization of promoter capture Hi-C datasets. *Bioinformatics* (2016) 32(16):2511–3. doi: 10.1093/bioinformatics/btw173

91. Kraja AT, Cook JP, Warren HR, Surendran P, Liu C, Evangelou E, et al. New Blood Pressure-Associated Loci Identified in Meta-Analyses of 475 000 Individuals. *Circ Cardiovasc Genet* (2017) 10(5):e001778. doi: 10.1161/CIRCGENETICS.117.001778

92. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* (2014) 159(7):1665–80. doi: 10.1016/j.cell.2014.11.021

93. Schmitt AD, Hu M, Jung I, Xu Z, Qiu Y, Tan CL, et al. A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cell Rep* (2016) 17(8):2042–59. doi: 10.1016/j.celrep.2016.10.061

94. Finan C, Gaulton A, Kruger FA, Lumbers RT, Shah T, Engmann J, et al. The druggable genome and support for target identification and validation in drug development. *Sci Transl Med* (2017) 9(383):eaag1166. doi: 10.1126/scitranslmed.aag1166

95. Rosa-Garrido M, Chapski DJ, Schmitt AD, Kimball TH, Karbassi E, Monte E, et al. High-resolution mapping of chromatin conformation in cardiac myocytes reveals structural remodeling of the epigenome in heart failure. *Circulation* (2017) 136(17):1613–25. doi: 10.1161/CIRCULATIONAHA.117.029430

96. Liljedahl U, Lind L, Kurland L, Berglund L, Kahan T, Syvänen AC. Single nucleotide polymorphisms in the apolipoprotein B and low density lipoprotein receptor genes affect response to antihypertensive treatment. *BMC Cardiovasc Disord* (2004) 4(1):16. doi: 10.1186/1471-2261-4-16

97. Chandrashekar BS, Nandhini T, Vasanth V, Sriram R, Navale S. Topical minoxidil fortified with finasteride: an account of maintenance of hair density after replacing oral finasteride. *Indian Dermatol Online J* (2015) 6(1):17–20. doi: 10.4103/2229-5178.148925

98. Harakalova M, van Harssel JJ, Terhal PA, van Lieshout S, Duran K, Renkens I, et al. Dominant missense mutations in ABCC9 cause Cantú syndrome. *Nat Genet* (2012) 44(7):793–6. doi: 10.1038/ng.2324

99. Trynka G, Hunt KA, Bockett NA, Romanos J, Mistry V, Szperl A, et al. Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nat Genet* (2011) 43(12):1193–201. doi: 10.1038/ng.998

100. Stahl EA, Wegmann D, Trynka G, Gutierrez-Achury J, Do R, Voight BF, et al. Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nat Genet* (2012) 44(5):483–9. doi: 10.1038/ng.2232

101. Edwards SL, Beesley J, French JD, Dunning AM. Beyond GWASs: illuminating the dark road from association to function. *Am J Hum Genet* (2013) 93(5):779–97. doi: 10.1016/j.ajhg.2013.10.012

102. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* (2013) 339(6121):819–23. doi: 10.1126/science.1231143

103. Atlasi Y, Stunnenberg HG. The interplay of epigenetic marks during stem cell differentiation and development. *Nat Rev Genet* (2017) 18(11):643–58. doi: 10.1038/nrg.2017.57

frontiers
in Cardiovascular Medicine

Roles of the Chr.9p21.3 *ANRIL* Locus in Regulating Inflammation and Implications for Anti-Inflammatory Drug Target Identification

Ghazal Aarabi [1], Tanja Zeller [2,3], Guido Heydecke [1], Matthias Munz [4,5,6], Arne Schäfer S [4] and Udo Seedorf [1]*

[1] Department of Prosthetic Dentistry, Center for Dental and Oral Medicine, University Medical Center Hamburg-Eppendorf, Hamburg, Germany, [2] Department of General and Interventional Cardiology, University Heart Center Hamburg (UHZ), University Medical Center Hamburg-Eppendorf, Hamburg, Germany, [3] Deutsches Zentrum für Herz-Kreislauf-Forschung (DZHK), Partner Site Hamburg/Lübeck/Kiel, Hamburg, Germany, [4] Center of Dento-Maxillo-Facial Medicine, Department of Periodontology and Synoptic Dentistry, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany, [5] Institute for Cardiogenetics, University of Lübeck, Lübeck, Germany, [6] University Heart Center Lübeck, Lübeck, Germany

Periodontitis (PD) is a common gingival infectious disease caused by an over-aggressive inflammatory reaction to dysbiosis of the oral microbiome. The disease induces a profound systemic inflammatory host response, that triggers endothelial dysfunction and pro-thrombosis and thus may aggravate atherosclerotic vascular disease and its clinical complications. Recently, a risk haplotype at the *ANRIL/CDKN2B-AS1* locus on chromosome 9p21.3, that is not only associated with coronary artery disease / myocardial infarction (CAD/MI) but also with PD, could be identified by genome-wide association studies. The locus encodes ANRIL - a long non-coding RNA (lncRNA) which, like other lncRNAs, regulates genome methylation via interacting with specific DNA sequences and proteins, such as DNA methyltranferases and polycomb proteins, thereby affecting expression of multiple genes by *cis* and *trans* mechanisms. Here, we describe ANRIL regulated genes and metabolic pathways and discuss implications of the findings for target identification of drugs with potentially anti-inflammatory activity in general.

Keywords: periodontitis, inflammation, ANRIL, 9p21.3, drug target, anti-inflammatory agents, coronary artery disease, CDKN2B-AS1

INTRODUCTION

Periodontitis (PD) is an inflammatory disease that involves the osseous, connective, and epithelial, tissues surrounding the teeth (1). Bacteria attached to the teeth along the gingival margin form a biofilm, which may trigger an immune response in the adjacent gingival tissue. If the biofilm is not removed and persists, it can induce gingivitis characterized by swelling, redness and bleeding (2). If the bacterial biofilm and the accompanying inflammatory reaction migrate apically along the root surface and penetrate into the tooth supporting structures the gingival inflammation becomes PD (3), which exists in two forms, chronic periodontitis (CP) and a more severe, early onset form called aggressive periodontitis (AgP) (4). In the US almost 50% of adults aged 30 years or above

have CP, including 30% with moderate and 8.5% with severe PD (5). Compared with CP, AgP is less frequent (prevalence: <0.1%). PD is a complex inflammatory disease, which is influenced considerably by interactions between environmental, lifestyle and genetic factors. Some individuals develop PD at young age, although they have similar lifestyle habits and environmental context compared to individuals who do not develop the disease. Therefore, it is considered that early-age of disease onset often indicates a genetic predisposition (6). The genetic susceptibility to PD has been examined extensively by GWAS (7–10) and seven common variants were identified, three of which met the genome-wide significance thresholds. Of the latter three, one (GLT6D1, glycosyltransferase 6 domain containing 1) is specific for AgP, whereas the other two (SIGLEC5, sialic acid binding Ig like lectin 5; DEFA1A3, defensin alpha 1/alpha 3) are associated with both AgP and CP (8, 10, 11). However, to date no associations that met the genome-wide significance threshold for common and rare alleles could be identified for CP alone. It is considered that these not signnificant findings are caused by the small sample sizes that were employed. Yet, some loci give suggestive evidence for association with PD. This evidence is based on independent replication in samples of the same disease phenotype with sufficient statistical power, independent validation of the associations in samples of different disease manifestations, like AgP and CP, and independent identification through different unbiased systematic approaches. According to these criteria, the following loci in addition to GLT6D1, SIGLEC5 and DEFA1A3 may currently be considered to be associated with CP and/or AgP: ANRIL (antisense noncoding RNA in the INK4 locus), NPY (neuropeptide Y), PF4 (platelet factor 4), PLG (plasminogen), VAMP3 (vesicle associated membrane protein 3) (10, 12–20).

Results obtained from longitudinal epidemiological studies support that CAD and CP are associated with each other (21), although the causative relationship between CAD and CP has remained ambiguous (22). Interestingly, variants at ANRIL, PLG and VAMP3 were reported to be associated with periodontal phenotypes and also with CAD [recently reviewed in ref. (23)]. Of these, ANRIL is the most significant risk locus of CAD and the association of ANRIL with PD was replicated repeatedly. In this narrative review, we summarize recent publications on the impact of this locus on chronic inflammation and to discuss potential approaches and strategies to identify new drug targets related to anti-inflammatory therapies in general.

THE CHR.9P21.3 RISK REGION IS SHARED BETWEEN PERIODONTITIS AND CAD/MI AND AFFECTS GENE EXPRESSION OF MULTIPLE GENES IN DIFFERENT CELL TYPES

The 9p21.3 risk haplotype at ANRIL/CDKN2B-AS1 had initially been identified by GWAS of CAD (24), and was shortly later identified by Schaefer et al. as one of the first genetic risk factors of AgP (17, 25–27) [see (**Table 1**) for a comparison of the association statistics of the relevant 9p21.3 lead SNPs related to AgP and coronary heart disease].

The core risk haplotype of ~50 kb, that is shared between CAD/MI and PD encodes the 3'end of a long ncRNA called "antisense non-coding RNA in the INK4 locus (ANRIL)" (also designated CDKN2BAS) (17, 25). Its sequence is oriented antisense relative to cyclin-dependent kinase inhibitor 2B (CDKN2B), which is located adjacent to the core CAD/PD region. Together with CDKN2A, which is located further upstream of ANRIL, this region harbors a hotspot for multiple complex human diseases and traits (28). Adjacent is a tightly linked locus for diabetes (29) which is neither associated with CAD (29) nor PD (17).

Given the extended region of high linkage disequilibrium at the 9p21.3 locus and the large number of transcriptional regulatory elements that are present in the CAD risk region, it is currently not entirely clear whether the risk of CAD and PD is mediated solely by ANRIL or whether its neighbors, CDKN2B and CDKN2A - two well-known tumor suppressor genes involved in cell cycle arrest and malignant transformation in certain cancers (30) - contribute to the mechanism. Knockout mice lacking CDKN2B do not only develop a cancer-related phenotype but also advanced aneurysms, accelerated smooth muscle cell apoptosis and medial arterial thinning (31), suggesting a potential involvement of CDKN2B not only in cancer but also in vascular disease. CAD risk SNP rs1537373 affects CDKN2B expression in human coronary artery smooth muscle cells, aorta and the mammary artery (32), and CDKN2B has been shown to regulate inflammatory cytokine production and the clearance of smooth muscle cell-derived apoptotic bodies during atherosclerosis (33). Miller et al. (32) recently investigated the role of SNP rs1537373 in the expression of ANRIL. This variant resides in a large haplotype block of linked variants including the highly replicated CAD SNP, rs4977574 and the CAD and PD lead SNP rs1333049 (17, 34). Although rs1537373 does not affect a known transcription factor binding motif, it is located at

TABLE 1 | Summary of the Chromosome 9p21.3 Locus Associated with Coronary Artery Disease and Periodontitis.

SNP	OR (AgP)	OR (CHD)	P (AgP)	P (CHD)	CI 95% (AgP)	CI 95% (CHD)	N (AgP)	N (CHD)
rs2891168	1.44	1.42	4.4 E-3	1.1 E-6	1.12–1.86	1.23–1.64		
rs1333042	1.44	1.42	4.8 E-3	1.2 E-6	1.12–1.85	1.23–1.64	159/736	1,104/736
rs1333048	1.48	1.39	2.5 E-3	7.6 E-6	1.15–1.92	1.20–1.60		

Association statistics of tree haplotype tagging SNPs at the relevant chromosome 9p21.3 risk region, multiplicative model adjusted for smoking, diabetes, and gender in a logistic regression model. AgP: aggressive periodontitis (generalized), CHD, coronary heart disease (disease onset <55 years), OR: odds ratio, CI: confidence interval, P: P-value obtained from a Wald test, N: number of cases/controls. Data extracted from ref. (17).

a site of accessible chromatin. Allele-specific transcription factor binding and histone H3 lysine 27 acetylation around rs1537373 indicated that the native chromatin structure may be affected by the genotype, which was consistent with the observed *cis* eQTL affecting CDKN2B rather than ANRIL in aortic tissues (32). It appears noteworthy in this context that SNP rs1537373 was earlier demonstrated to be also strongly associated with coronary artery calcification (35). If bone marrow lacking murine Cdkn2a was transplanted to the atherosclerosis prone Ldlr(-/-) mouse model, the Cdkn2a-deficient recipients exhibited accelerated atherosclerosis, a higher number of pro-inflammatory monocytes, and increased monocyte/macrophage proliferation compared to controls (36). Thus besides CDKN2B, also CDKN2A has some plausibility for being involved in the pathogenesis of vascular inflammation [see the review by Hannou et al. (37) for further information].

The location of the core risk haplotype of CAD/MI and PD at the 3'end of *ANRIL* implies that the encoded long ncRNA is a prime functional candidate involved in the risk mediating mechanism(s). *ANRIL* is a lowly expressed gene consisting of 20 exons whose transcripts could be detected in a wide variety of cell-types and tissues, including smooth muscle cells, endothelial cells, and cells of the immune system that are known to be involved in atherogenesis (29, 38, 39). Originally, two splice variants were demonstrated in normal human testis and signals using PCR with primers derived from exons 14–16 were also obtained in a range of other tissues (40). Subsequently, many additional splice variants could be identified in various cell-types (38, 41, 42). ANRIL is subject to a complex pathway of alternative splicing which may differ from tissue to tissue and which may be influenced by the presence of SNPs interfering with the function of splice signals.

ANRIL expression was reported to be tightly linked to the *ANRIL* genotype due to disruption of an inhibitory STAT1 binding site in risk allele carriers (43), which would be expected to impair the IFNγ signaling response. However, results published by Almontashiri et al. argued against an involvement of IFNγ in the mechanism underlying the association of the 9p21.3 genotype with CAD risk (44). The CAD risk allele of SNP rs564398, which is one of the SNPs most strongly correlated with ANRIL expression, was predicted to disrupt a Ras Responsive Element Binding protein (RREB) 1 binding site in the 9p21.3 locus (45, 46). RREB may be involved in up-regulating CDKN2B in a Ras-dependent manner by down-regulating ANRIL. Besides stimulating VSMC senescence, Ras has also been implicated to contribute to atherogenesis by affecting vascular inflammation (47). The local functional influence of variants in the 9p21.3 region on gene expression has been examined by many other studies in a variety of tissues and cells (41, 45, 48–52). The results confirmed that the CAD risk variants in the 9p21.3 region are strongly associated with *ANRIL* expression and also with expression of the adjacent loci (*CDKN2A*, *CDKN2B*), albeit much more moderately. However, there is some inconsistency concerning the direction of the effect. Earlier studies suggested associations between CAD risk variants and lower *ANRIL* expression in vascular smooth muscle cells, whole blood cells and purified peripheral blood T-cells (49, 53, 54). In contrast, the study by Holdt et al. (51), in which specifically the long ANRIL transcript (ENST00000428597) was measured,

demonstrated that the CAD risk haplotype was associated with higher *ANRIL* expression in whole blood cells and peripheral blood mononuclear cells. Also Zhao et al. found higher expression of this transcript in transformed beta-lymphocytes collected from genotyped donors who carried the CAD risk variant rs7865618 (55). In the latter study, all CAD risk variants assayed in the study were associated with the same directions of the effects.

In addition to the linear form of ANRIL, there also exists a circular ANRIL RNA form (38). Recently, Holdt et al. (56) showed that circular ANRIL may be athero-protective by regulating rRNA maturation. In their model, pescadillo homologue 1 (PES1, a 60S-preribosomal assembly factor) binds to circular ANRIL, which impairs ribosome biogenesis and exonuclease-mediated pre-rRNA processing. The resulting nucleolar stress induces activation of p53, which triggers apoptosis and inhibits proliferation, thereby preventing the accumulation of vascular smooth muscle cells and foam cells at the sites of the atherosclerotic lesion. The balance between atherogenic linear and athero-protective circular ANRIL may be critical for the impact of ANRIL on disease progression. Conversely, a recently published study came to the opposite conclusion, namely that circular ANRIL may be pro-atherogenic (57). In this study, circular antisense ANRIL was used to investigate the inflammatory response of vascular endothelial cells *in vivo* in a rat model of coronary atherosclerosis which was established by injecting rats on a high fat diet with vitamin D3 (57). Circular antisense ANRIL lowered circular ANRIL in vascular endothelial cells along with the levels of several pro-atherogenic markers (serum cholesterol, triglycerides, LDL, IL-1, IL-6, MMP-9, CRP, cANRIL, Bax, caspase-3) and the rates of endothelial cell apoptosis, while HDL levels and bcl-2 expression were increased. In contrast, induction of circular ANRIL expression promoted atherosclerosis by increasing pro-inflammatory properties in vascular endothelial cells and by raising serum lipid and pro-inflammatory cytokine levels. These results were consistent with the hypothesis, that inhibiting circular ANRIL expression would be anti-inflammatory and would reduce vascular endothelial cell apoptosis, which in turn would protect against atherosclerosis in this animal model.

In earlier studies, it could be demonstrated that the epigenetic silencer polycomb repressive complexes 1 and 2 (PRC1 and PRC2) and PRC-associated activating proteins RYBP and YY1 can bind to ANRIL (58, 59), suggesting that ANRIL may be able to modulate epigenetic regulation of target gene expression in *cis* and *trans*. It could be demonstrated *in vitro* by inducible knock-down approaches in T-Rex 293 HEK cells that silencing of two proximal *ANRIL* transcripts altered expression of *ADIPOR1*, *VAMP3* and *TMEM258* (60) (see **Table 2** for a list of genes regulated by *ANRIL*). ADIPOR1 is a high-affinity receptor for globular adiponectin, which is involved, amongst others, in PPARα (peroxisome proliferator activated receptor alpha) and AMPK (AMP-activated protein kinase) signaling (62). PPARα activation could prevent experimentally induced bone-loss in animal studies (63). AMPK and PPARα act as key regulators of glucose and fatty acid metabolism in the liver. Adiponectin levels are inversely correlated with BMI, body fat and severity of CAD (64). Globular adiponectin also increases insulin sensitivity by stimulating cellular glucose uptake via increasing recruitment of glucose transporter 4 (GLUT4) to the plasma membrane and inducing *GLUT4* expression (65). Besides these

TABLE 2 | *ANRIL*-Regulated Genes

Gene	Mode	Gene description	Tissue / cell type	Disease	Ref.
ANRIL	cis	ANRIL, long ncRNA	PBMC, atherosclerotic plaque	CAD, PD	(39)
CDKN2A	cis	Cyclin-dependent kinase Inhibitor 2A, tumor suppressor	ANRIL knock-down in VSMC	Cancer	(45)
CDKN2B	cis	Cyclin-dependent kinase Inhibitor 2B, tumor suppressor	ANRIL knock-down in VSMC	Cancer	(45)
ADIPOR1	trans	Adiponectin receptor 1, glucose, lipid metabolism	inducible ANRIL knock-down in T-Rex 293 HEK cells	Diabetes, CVD	(13)
VAMP3	trans	Vesicle-associated membrane protein 3, IL-6, TNFα secretion	inducible ANRIL knock-down in T-Rex 293 HEK cells	Inflammation / cancer	(13)
C11ORF10	trans	TMEM258, N-glycosylation	inducible ANRIL knock-down in T-Rex 293 HEK cells	Unknown	(13)
DUT	trans	Deoxyuridine Triphosphatase, nucleotide metabolism	Transformed B cells	Unknown	(55)
EIF1AY	trans	Eukaryotic translation initiation factor 1A (Y-chr.), mRNA Translation	Transformed B cells	Unknown	(55)
CASP14	trans	Caspase-14, Inflammation, apoptosis	Transformed B cells	Psoriasis	(55)
ABCA1	trans	ATP binding cassette transporter A1, sterol transport	Transformed B cells	CAD/MI	(55)
DHRS9	trans	Dehydrogenase reductase 9, retinol metabolism	Transformed B cells	Unknown	(55)
CARD8	trans	Caspase recruitment domain 8, inflammasome	ANRIL knock-down / over-expression in HUVEC, HepG2 cells	Inflammation	(61)

CAD, coronary artery disease; CVD, cardiovascular disease; HUVEC, human umbilical vein endothelial cells; IL, interleukin; PBMC, peripheral blood mononuclear cells; VSMC, vascular smooth muscle cells

metabolic roles, adiponectin also has anti-inflammatory activity by activating tissue inhibitors of metalloproteinases, IL-10, and by suppressing lipopolysaccharide-activated *TNF* (tumor necrosis factor) expression and phagocytic activity (66, 67). The effect of ANRIL on *VAMP3* expression (**Table 2**) may be important, because VAMP3 belongs to the VAMP/synaptobrevin family involved in phagocytosis and trafficking of TNF-α-containing secretory vesicles to the cell surface required for TNF-α secretion (68).

Genome-wide *cis* and *trans* effects of the variants in the 9p21.3 region on gene expression were recently studied by Zhao et al. (55), who employed the SNP-set (Sequence) Kernel Association Test [SKAT, (69)] on genotyped transformed beta-lymphocytes collected from 801 participants from the Genetic Epidemiology Network of Arteriopathy (GENOA) study. The results demonstrated a significant association between the CAD and PD risk variants in the region with the expression of the long linear *ANRIL* transcript containing the coding information of all 20 exons except exon 13. In addition to this *cis*-regulatory effect, several *trans* eQTLs could also be identified (**Table 2**). The affected genes were *DUT* (Deoxyuridine Triphosphatase also known as UTPase), *EIF1AY* (Eukaryotic Translation Initiation Factor 1A, Y-Linked), *CASP14*

(Caspase 14), *ABCA1* (ATP-binding cassette transporter A1), and *DHRS9* (Dehydrogenase/Reductase 9) (**Table 2**) (55).

The *DUT* gene product is an essential enzyme of nucleotide metabolism, which is required for the hydrolysis of dUTP into dUMP and inorganic pyrophosphate. The enzyme plays an important role in controlling the relative cellular levels of dUTP/ dTTP (70). Lack or inhibition of dUTPase result in elevated levels of uracil in the DNA, which triggers DNA repair and may induce the formation of DNA double strand breaks, somatic mutations, and apoptosis (71).

CASP14 is involved in cell apoptosis and is over-expressed in skin, the oral epithelium, bone, heart, and epithelial tumors (72). *EIF1AY* encodes a translation initiation factor which seems to be required for maximal rate of protein biosynthesis (73) and *DHRS9* is involved in retinol and steroid metabolism (74). *ABCA1* plays a well-known role in atherosclerosis (75); but its contribution to PD is unclear. It was proposed that LPS from *P. gingivalis*, which is the most important pathogen involved in PD, may suppress *ABCA1* expression during periodontitis via miRNA-mediated mechanisms (76). To further investigate the potential biological implications of the *trans*-effected genes, Zhao et al. (55) performed

gene enrichment analysis on basis of the KEGG Pathway databank. The enriched pathways included "retinol metabolism", "TGF-β signaling", and "N-glycan biosynthesis". Retinol metabolism was at the top of the list of enriched pathways, in which *LRAT* (lecithin retinol acyltransferase), *ADH1* (alcohol dehydrogenase 1), *DHRS9*, *DHRS4L2* (dehydrogenase/reductase 9 and 4 like 2), and CYP26B1 (cytochrome P450 retinoid metabolizing protein) were significantly associated. The importance of TGF-β signaling in the pathogenesis of PD is well-known, since anti-TGF-β antibodies can inhibit the recruitment of leukocytes and the destruction of cartilage and bone at the periodontal lesion sites during periodontitis (77). Another reported downstream target regulated by *ANRIL* is *CARD8* (caspase recruitment domain-containing protein 8) (**Table 2**) (61). The *CARD8* SNP rs2043211 is significantly associated with ischemic stroke; but its involvement in PD is unclear. The *CARD8*

gene product is a component of the inflammasome together with other proteins. ANRIL is induced by pro-inflammatory factors, such as TNFα and IFN-γ, via activation of NF-κB (**Figure 1**) (78). The transcription factor Yin yang 1 (YY1) can bind to ANRIL and the ANRIL-YY1 complex interacts with the promoter of *IL6/8* to activate *IL6* and *IL8* expression, two cytokines with well established roles in CAD/MI and PD.

Taken together, these findings seem to suggest that *ANRIL* exerts its effects through epigenetic regulation of a great variety of target genes. The common theme seems to be its involvement in expression regulation of genes that play important roles in inflammation, immunity, cell apoptosis and survival, cell proliferation, and metabolism. Many of the reported *trans* regulated genes clearly have plausible roles in CAD and PD as well. Nevertheless, at this stage, we find it premature to formulate a unifying theory that
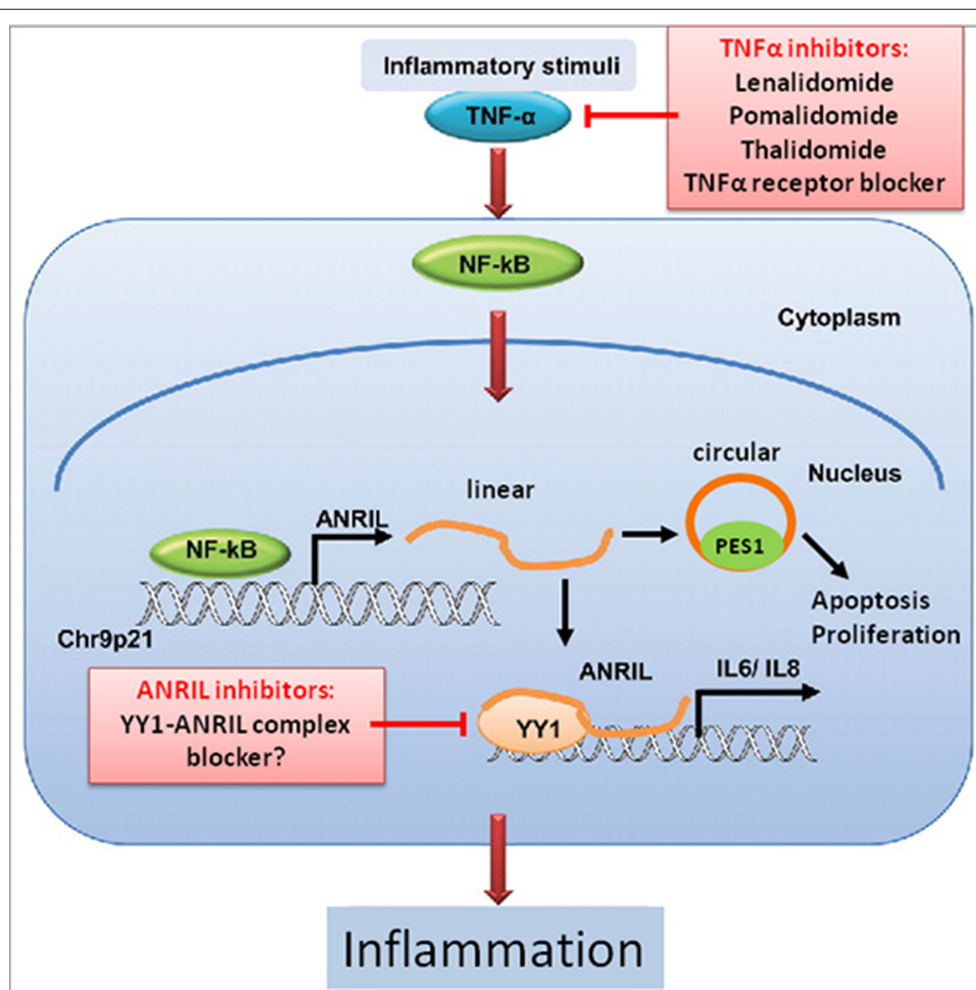
FIGURE 1 | Hypothetical roles of linear and circular ANRIL lncRNA in regulating inflammation and cell survival in human vascular endothelial cells and potential drug targets. TNF-α triggers NF-κB activation, which induces *ANRIL* transcription (66). Linear ANRIL can be converted to circular ANRIL (38). Linear ANRIL interacts with the transcription factor yin yang-1 (YY1) to form a functional complex that binds to and regulates expression of target genes such as IL-6/8. Circular ANRIL interacts with pescadillo homologue 1 (PES1) to form a complex with the pre-ribosomal assembly complex, that impairs ribosome biogenesis, leading to activation of p53 and a subsequent increase in apoptosis and decrease in the proliferative rate (41). This pathway may promote atheroprotection by eliminating over-proliferating cells in atherosclerotic plaques. Neither TNFα nor NF-κB antagonists do seem suitable for wide-spread use in anti-inflammatory therapies of PD or CAD, because of their serious side effects. Since ANRIL is located downstream of TNFα and NF-κB, ANRIL or its downstream targets may be better suited as drug targets to inhibit the pro-inflammatory activities linked to this signaling pathway [modified according to ref. (78)].

would be consistent with at least the majority of the findings. Most concerning is the apparent complete lack of replication of *trans* regulated genes between the published studies. The reasons for this striking inconsistency may have something to do with the diversity of the experimental approaches and cell-types that have been employed to date. The genome-wide approaches may lack sufficient power to detect some of the differentially expressed genes identified by targeted strategies (55). Antisense approaches are difficult to control due to the complex cell-type specific alternative splicing pathways (38, 41, 42) and findings coming from rodent animal models may not be relevant for humans, since rodent and human ANRIL are evolutionary not well conserved and they differ structurally substantially from each other (79).

IMPLICATIONS OF THE CHR.9P21.3 *ANRIL* LOCUS FOR DRUG TARGET IDENTIFICATION

Zhou et al. (78) showed that *ANRIL* expression is up-regulated via the TNFα/NF-κB signaling pathway under inflammatory stress conditions (**Figure 1**). Since endothelial cell-specific inhibition of NF-κB protects mice from atherosclerosis (80), and since *ANRIL* is a downstream target of TNFα/NF-κB signaling, targeting TNFα or NF-κB may theoretically be considered to be athero-protective via inhibiting ANRIL-YY1-mediated IL-6/8 production. Several TNFα receptor antagonists (mostly antibodies) have been tested for safety and efficiency for modulating pro-inflammatory cytokine release in the treatment of rheumatoid arthritis (81). However, clinical trials have shown that these receptor antagonists are associated with increased risks of malignancies and serious infections (81). Since *ANRIL* is located downstream of TNFα and NF-κB, it may

be better suited as drug target. However, given the important role of *ANRIL* transcripts in controlling cell growth, its expression is likely precisely regulated. Possibly, putative drug targeting options may come to mind from a better understanding of the precise downstream effects of the linear and circular *ANRIL* lncRNAs on expression of genes involved in chronic inflammatory pathways, suggesting that such work has potential to identify new drug targets for anti-inflammatory intervention.

AUTHOR CONTRIBUTIONS

GA and US had the initial idea of writing a review and proposed the topic. Moreover, they conducted extensive literature search and created the first draft of the manuscript. TZ, GH, AS and MM integrated the different information and also wrote and submited the manuscript.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

1. Kinane DF, Loe H, Theilade E, Jensen SB, Kinane DF, Loe H et al Pathogenesis of Periodontitis. In: Lindhe J, Lang NP, Karring T, editors. *Clinical Periodontology and Implant Dentistry*. Oxford: Blackwell Munksgaard (2008). p. 256–69.

2. Loe H, Theilade E, Jensen SB. Experimental gingivitis in man. *J Periodontol* (1965) 36:177–87. doi: 10.1902/jop.1965.36.3.177

3. Robinson PJ. Gingivitis: a prelude to periodontitis? *J Clin Dent* (1995) 6:41–5.

4. Kinane DF. Berglundh: Pathogenesis of Periodontitis. In: Lindhe J, Lang NP, Karring T, editors. *Clinical Periodontology and Implant Dentistry*. Oxford: Blackwell Munksgaard (2008).

5. Thornton-Evans G, Eke P, Wei L, Palmer A, Moeti R, Hutchins S, et al. Periodontitis among adults aged ≥30 years - United States, 2009-2010. *MMWR supplements* (2013) 62(3):129–35.

6. Vieira AR, Albandar JM. Role of genetic factors in the pathogenesis of aggressive periodontitis. *Periodontol. 2000* (2014) 65(1):92–106. doi: 10.1111/prd.12021

7. Divaris K, Monda KL, North KE, Olshan AF, Reynolds LM, Hsueh WC, et al. Exploring the genetic basis of chronic periodontitis: a genome-wide association study. *Hum Mol Genet* (2013) 22(11):2312–24. doi: 10.1093/hmg/ddt065

8. Schaefer AS, Richter GM, Nothnagel M, Manke T, Dommisch H, Jacobs G, et al. A genome-wide association study identifies GLT6D1 as a susceptibility locus for periodontitis. *Hum Mol Genet* (2010) 19(3):553–62. doi: 10.1093/hmg/ddp508

9. Teumer A, Holtfreter B, Völker U, Petersmann A, Nauck M, Biffar R, et al. Genome-wide association study of chronic periodontitis in a general German population. *J Clin Periodontol* (2013) 40(11):977–85. doi: 10.1111/jcpe.12154

10. Munz M, Willenborg C, Richter GM, Jockel-Schneider Y, Graetz C, Staufenbiel I, et al. A genome-wide association study identifies nucleotide variants at SIGLEC5 and DEFA1A3 as risk loci for periodontitis. *Hum Mol Genet* (2017) 26(13):2577–88. doi: 10.1093/hmg/ddx151

11. Hashim NT, Linden GJ, Ibrahim ME, Gismalla BG, Lundy FT, Hughes FJ, et al. Replication of the association of GLT6D1 with aggressive periodontitis in a Sudanese population. *J Clin Periodontol* (2015) 42(4):319–24. doi: 10.1111/jcpe.12375

12. Freitag-Wolf S, Dommisch H, Graetz C, Jockel-Schneider Y, Harks I, Staufenbiel I, et al. Genome-wide exploration identifies sex-specific genetic effects of alleles upstream NPY to increase the risk of severe periodontitis in men. *J Clin Periodontol* (2014) 41(12):1115–21. doi: 10.1111/jcpe.12317

13. Bochenek G, Häsler 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(22):4516–27. doi: 10.1093/hmg/ddt299

14. Divaris K, Monda KL, North KE, Olshan AF, Lange EM, Moss K, et al. Genome-wide association study of periodontal pathogen colonization. *J Dent Res* (2012) 91(7 Suppl):S21–8. doi: 10.1177/0022034512447951

15. Munz M, Chen H, Jockel-Schneider Y, Adam K, Hoffman P, Berger K, et al. A haplotype block downstream of plasminogen is associated with chronic and aggressive periodontitis. *J Clin Periodontol* (2017) 44(10):962–70. doi: 10.1111/jcpe.12749

16. Schaefer AS, Bochenek G, Jochens A, Ellinghaus D, Dommisch H, Güzeldemir-Akçakanat E, et al. Genetic evidence for PLASMINOGEN as a shared genetic risk factor of coronary artery disease and periodontitis. *Circ Cardiovasc Genet* (2015) 8(1):159–67–. doi: 10.1161/CIRCGENETICS.114.000554

17. Schaefer AS, Richter GM, Groessner-Schreiber B, Noack B, Nothnagel M, El Mokhtari NE, et al. Identification of a shared genetic susceptibility locus for coronary heart disease and periodontitis. *PLoS Genet* (2009) 5(2):e1000378. doi: 10.1371/journal.pgen.1000378

18. Ernst FD, Uhr K, Teumer A, Fanghänel J, Schulz S, Noack B, et al. Replication of the association of chromosomal region 9p21.3 with generalized aggressive periodontitis (gAgP) using an independent case-control cohort. *BMC Med Genet* (2010) 11(1):119. doi: 10.1186/1471-2350-11-119

19. Shusterman A, Munz M, Richter G, Jepsen S, Lieb W, Krone B, et al. The PF4/PPBP/CXCL5 gene cluster is associated with periodontitis. *J Dent Res* (2017) 96(8):945–52. doi: 10.1177/0022034517706311

20. Schaefer AS, Bochenek G, Jochens A, Ellinghaus D, Dommisch H, Güzeldemir-Akçakanat E, et al. Genetic evidence for PLASMINOGEN as a shared genetic risk factor of coronary artery disease and periodontitis. *Circ Cardiovasc Genet* (2015) 8(1):159–67. doi: 10.1161/CIRCGENETICS.114.000554

21. Dietrich T, Jimenez M, Krall Kaye EA, Vokonas PS, Garcia RI. Age-dependent associations between chronic periodontitis/edentulism and risk of coronary heart disease. *Circulation* (2008) 117(13):1668–74. doi: 10.1161/CIRCULATIONAHA.107.711507

22. Lockhart PB, Bolger AF, Papapanou PN, Osinbowale O, Trevisan M, Levison ME, et al. Periodontal disease and atherosclerotic vascular disease: does the evidence support an independent association?: a scientific statement from the American Heart Association. *Circulation* (2012) 125(20):2520–44. doi: 10.1161/CIR.0b013e31825719f3

23. Aarabi G, Zeller T, Seedorf H, Reissmann DR, Heydecke G, Schaefer AS, et al. Genetic susceptibility contributing to periodontal and cardiovascular disease. *J Dent Res* (2017) 96(6):610–7. doi: 10.1177/0022034517699786

24. 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(5830):1488–91. doi: 10.1126/science.1142447

25. Ernst FD, Uhr K, Teumer A, Fanghänel J, Schulz S, Noack B, et al. Replication of the association of chromosomal region 9p21.3 with generalized aggressive periodontitis (gAgP) using an independent case-control cohort. *BMC Med Genet* (2010) 11:119. doi: 10.1186/1471-2350-11-119

26. Schaefer AS, Richter GM, Dommisch H, Reinartz M, Nothnagel M, Noack B, et al. CDKN2BAS is associated with periodontitis in different European populations and is activated by bacterial infection. *J Med Genet* (2011) 48(1):38–47. doi: 10.1136/jmg.2010.078998

27. Schaefer AS, Bochenek G, Manke T, Nothnagel M, Graetz C, Thien A, et al. Validation of reported genetic risk factors for periodontitis in a large-scale replication study. *J Clin Periodontol* (2013) 40(6):563–72. doi: 10.1111/jcpe.12092

28. Pasmant E, Sabbagh A, Vidaud M, Bièche I. ANRIL, a long, noncoding RNA, is an unexpected major hotspot in GWAS. *Faseb J* (2011) 25(2):444–8. doi: 10.1096/fj.10-172452

29. 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(6):806–14. doi: 10.1093/hmg/ddm352

30. Boström J, Meyer-Puttlitz B, Wolter M, Blaschke B, Weber RG, Lichter P, et al. Alterations of the tumor suppressor genes CDKN2A (p16(INK4a)), p14(ARF), CDKN2B (p15(INK4b)), and CDKN2C (p18(INK4c)) in atypical and anaplastic meningiomas. *Am J Pathol* (2001) 159(2):661–9. doi: 10.1016/S0002-9440(10)61737-3

31. Leeper NJ, Raiesdana A, Kojima Y, Kundu RK, Cheng H, Maegdefessel L, et al. Loss of CDKN2B promotes p53-dependent smooth muscle cell apoptosis and aneurysm formation. *Arterioscler Thromb Vasc Biol* (2013) 33(1):e1–e10. doi: 10.1161/ATVBAHA.112.300399

32. Miller CL, Pjanic M, Wang T, Nguyen T, Cohain A, Lee JD, et al. Integrative functional genomics identifies regulatory mechanisms at coronary artery disease loci. *Nat Commun* (2016) 7:12092. doi: 10.1038/ncomms12092

33. Kojima Y, Downing K, Kundu R, Miller C, Dewey F, Lancero H, et al. Cyclin-dependent kinase inhibitor 2B regulates efferocytosis and atherosclerosis. *J Clin Invest* (2014) 124(3):1083–97. doi: 10.1172/JCI70391

34. Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, Mayer B, et al. Genomewide association analysis of coronary artery disease. *N Engl J Med* (2007) 357(5):443–53. doi: 10.1056/NEJMoa072366

35. Pechlivanis S, Mühleisen TW, Möhlenkamp S, Schadendorf D, Erbel R, Jöckel KH, et al. Risk loci for coronary artery calcification replicated at 9p21 and 6q24 in the Heinz Nixdorf Recall Study. *BMC Med Genet* (2013) 14:23. doi: 10.1186/1471-2350-14-23

36. Kuo CL, Murphy AJ, Sayers S, Li R, Yvan-Charvet L, Davis JZ, et al. Cdkn2a is an atherosclerosis modifier locus that regulates monocyte/macrophage proliferation. *Arterioscler Thromb Vasc Biol* (2011) 31(11):2483–92. doi: 10.1161/ATVBAHA.111.234492

37. 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 Metab* (2015) 26(4):176–84. doi: 10.1016/j.tem.2015.01.008

38. Burd CE, Jeck WR, Liu Y, Sanoff HK, Wang Z, Sharpless NE. Expression of linear and novel circular forms of an INK4/ARF-associated non-coding RNA correlates with atherosclerosis risk. *PLoS Genet* (2010) 6(12):e1001233. doi: 10.1371/journal.pgen.1001233

39. Holdt LM, Beutner F, Scholz M, Gielen S, Gäbel G, Bergert H, et al. ANRIL expression is associated with atherosclerosis risk at chromosome 9p21. *Arterioscler Thromb Vasc Biol* (2010) 30(3):620–7. doi: 10.1161/ATVBAHA.109.196832

40. Pasmant E, Laurendeau I, Héron D, Vidaud M, Vidaud D, Bièche 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(8):3963–9. doi: 10.1158/0008-5472.CAN-06-2004

41. Folkersen L, Kyriakou T, Goel A, Peden J, Mälarstig 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(11):e7677. doi: 10.1371/journal.pone.0007677

42. 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(7):e1003588. doi: 10.1371/journal.pgen.1003588

43. Harismendy O, Notani D, Song X, Rahim NG, Tanasa B, Heintzman N, et al. 9p21 DNA variants associated with coronary artery disease impair interferon-γ signalling response. *Nature* (2011) 470(7333):264–8. doi: 10.1038/nature09753

44. Almontashiri NA, Fan M, Cheng BL, Chen HH, Roberts R, Stewart AF. Interferon-γ activates expression of p15 and p16 regardless of 9p21.3 coronary artery disease risk genotype. *J Am Coll Cardiol* (2013) 61(2):143–7. doi: 10.1016/j.jacc.2012.08.1020

45. Congrains A, Kamide K, Oguro R, Yasuda O, Miyata K, Yamamoto E, et al. Genetic variants at the 9p21 locus contribute to atherosclerosis through modulation of ANRIL and CDKN2A/B. *Atherosclerosis* (2012) 220(2):449–55. doi: 10.1016/j.atherosclerosis.2011.11.017

46. Gschwendtner A, Bevan S, Cole JW, Plourde A, Matarin M, Ross-Adams H, et al. Sequence variants on chromosome 9p21.3 confer risk for atherosclerotic stroke. *Ann Neurol* (2009) 65(5):531–9. doi: 10.1002/ana.21590

47. Minamino T, Yoshida T, Tateno K, Miyauchi H, Zou Y, Toko H, et al. Ras induces vascular smooth muscle cell senescence and inflammation in human atherosclerosis. *Circulation* (2003) 108(18):2264–9. doi: 10.1161/01.CIR.0000093274.82929.22

48. 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(10):1671–7. doi: 10.1161/ATVBAHA.109.189522

49. Liu Y, Sanoff HK, Cho H, Burd CE, Torrice C, Mohlke KL, et al. INK4/ARF transcript expression is associated with chromosome 9p21 variants linked to atherosclerosis. *PLoS ONE* (2009) 4(4):e5027. doi: 10.1371/journal.pone.0005027

50. Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, Thompson JR, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet* (2013) 45(1):25–33. doi: 10.1038/ng.2480

51. Holdt LM, Beutner F, Scholz M, Gielen S, Gäbel G, Bergert H, et al. ANRIL expression is associated with atherosclerosis risk at chromosome

9p21. *Arterioscler Thromb Vasc Biol* (2010) 30(3):620–7. doi: 10.1161/ATVBAHA.109.196832

52. Congrains A, Kamide K, Katsuya T, Yasuda O, Oguro R, Yamamoto K, et al. CVD-associated non-coding RNA, ANRIL, modulates expression of atherogenic pathways in VSMC. *Biochem Biophys Res Commun* (2012) 419(4):612–6. doi: 10.1016/j.bbrc.2012.02.050

53. 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(4):e1000899. doi: 10.1371/journal.pgen.1000899

54. 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(18):4021–9. doi: 10.1093/hmg/dds224

55. Zhao W, Smith JA, Mao G, Fornage M, Peyser PA, Sun YV, et al. The cis and trans effects of the risk variants of coronary artery disease in the Chr9p21 region. *BMC Med Genomics* (2015) 8:21. doi: 10.1186/s12920-015-0094-0

56. 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

57. Song CL, Wang JP, Xue X, Liu N, Zhang XH, Zhao Z, et al. Effect of circular ANRIL on the inflammatory response of vascular endothelial cells in a rat model of coronary atherosclerosis. *Cell Physiol Biochem* (2017) 42(3):1202–12. doi: 10.1159/000478918

58. Kotake Y, Nakagawa T, Kitagawa K, Suzuki S, Liu N, Kitagawa M, et al. Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene. *Oncogene* (2011) 30(16):1956–62. doi: 10.1038/onc.2010.568

59. González I, Busturia A. High levels of dRYBP induce apoptosis in Drosophila imaginal cells through the activation of reaper and the requirement of trithorax, dredd and dFADD. *Cell Res* (2009) 19(6):747–57. doi: 10.1038/cr.2009.29

60. Bochenek G, Häsler 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(22):4516–27. doi: 10.1093/hmg/ddt299

61. Bai Y, Nie S, Jiang G, Zhou Y, Zhou M, Zhao Y, et al. Regulation of CARD8 expression by ANRIL and association of CARD8 single nucleotide polymorphism rs2043211 (p.C10X) with ischemic stroke. *Stroke* (2014) 45(2):383–8. doi: 10.1161/STROKEAHA.113.003393

62. Yamauchi T, Nio Y, Maki T, Kobayashi M, Takazawa T, Iwabu M, et al. Targeted disruption of AdipoR1 and AdipoR2 causes abrogation of adiponectin binding and metabolic actions. *Nat Med* (2007) 13(3):332–9. doi: 10.1038/nm1557

63. Stunes AK, Westbroek I, Gustafsson BI, Fossmark R, Waarsing JH, Eriksen EF, et al. The peroxisome proliferator-activated receptor (PPAR) alpha agonist fenofibrate maintains bone mass, while the PPAR gamma agonist pioglitazone exaggerates bone loss, in ovariectomized rats. *BMC Endocr Disord* (2011) 11:11. doi: 10.1186/1472-6823-11-11

64. Hara K, Yamauchi T, Imai Y, Manabe I, Nagai R, Kadowaki T. Reduced adiponectin level is associated with severity of coronary artery disease. *Int Heart J* (2007) 48(2):149–53. doi: 10.1536/ihj.48.149

65. Ceddia RB, Somwar R, Maida A, Fang X, Bikopoulos G, Sweeney G. Globular adiponectin increases GLUT4 translocation and glucose uptake but reduces glycogen synthesis in rat skeletal muscle cells. *Diabetologia* (2005) 48(1):132–9. doi: 10.1007/s00125-004-1609-y

66. Kumada M, Kihara S, Ouchi N, Kobayashi H, Okamoto Y, Ohashi K, et al. Adiponectin specifically increased tissue inhibitor of metalloproteinase-1 through interleukin-10 expression in human macrophages. *Circulation* (2004) 109(17):2046–9. doi: 10.1161/01.CIR.0000127953.98131.ED

67. Yokota T, Oritani K, Takahashi I, Ishikawa J, Matsuyama A, Ouchi N, et al. Adiponectin, a new member of the family of soluble defense collagens, negatively regulates the growth of myelomonocytic progenitors and the functions of macrophages. *Blood* (2000) 96(5):1723–32.

68. Murray RZ, Kay JG, Sangermani DG, Stow JL. A role for the phagosome in cytokine secretion. *Science* (2005) 310(5753):1492–5. doi: 10.1126/science.1120225

69. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet* (2011) 89(1):82–93. doi: 10.1016/j.ajhg.2011.05.029

70. Vértessy BG, Tóth J. Keeping uracil out of DNA: physiological role, structure and catalytic mechanism of dUTPases. *Acc Chem Res* (2009) 42(1):97–106. doi: 10.1021/ar800114w

71. Wilson PM, Labonte MJ, Lenz HJ, Mack PC, Ladner RD. Inhibition of dUTPase induces synthetic lethality with thymidylate synthase-targeted therapies in non-small cell lung cancer. *Mol Cancer Ther* (2012) 11(3):616–28. doi: 10.1158/1535-7163.MCT-11-0781

72. Krajewska M, Kim H, Shin E, Kennedy S, Duffy MJ, Wong YF, et al. Tumor-associated alterations in caspase-14 expression in epithelial malignancies. *Clin Cancer Res* (2005) 11(15):5462–71. doi: 10.1158/1078-0432.CCR-04-2527

73. Jackson RJ, Hellen CU, Pestova TV. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol* (2010) 11(2):113–27. doi: 10.1038/nrm2838

74. Soref CM, di YP, Hayden L, Zhao YH, Satre MA, Wu R. Characterization of a novel airway epithelial cell-specific short chain alcohol dehydrogenase/reductase gene whose expression is up-regulated by retinoids and is involved in the metabolism of retinol. *J Biol Chem* (2001) 276(26):24194–202. doi: 10.1074/jbc.M100332200

75. Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, Mcpherson R, Hobbs HH. Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* (2004) 305(5685):869–72. doi: 10.1126/science.1099870

76. Olsen I, Singhrao SK, Osmundsen H. Periodontitis, pathogenesis andprogression: miRNA-mediated cellular responses to. *J Oral Microbiol* (2017) 9(1):1333396.

77. Wahl SM, Allen JB, Costa GL, Wong HL, Dasch JR. Reversal of acute and chronic synovial inflammation by anti-transforming growth factor beta. *J Exp Med* (1993) 177(1):225–30. doi: 10.1084/jem.177.1.225

78. Zhou X, Han X, Wittfeldt A, Sun J, Liu C, Wang X, et al. Long non-coding RNA ANRIL regulates inflammatory responses as a novel component of NF-κB pathway. *RNA Biol* (2016) 13(1):98–108. doi: 10.1080/15476286.2015.1122164

79. He S, Gu W, Li Y, Zhu H. ANRIL/CDKN2B-AS shows two-stage clade-specific evolution and becomes conserved after transposon insertions in simians. *BMC Evol Biol* (2013) 13:247. doi: 10.1186/1471-2148-13-247

80. Gareus R, Kotsaki E, Xanthoulea S, van der Made I, Gijbels MJ, Kardakaris R, et al. Endothelial cell-specific NF-kappaB inhibition protects mice from atherosclerosis. *Cell Metab* (2008) 8(5):372–83. doi: 10.1016/j.cmet.2008.08.016

81. Bongartz T, Sutton AJ, Sweeting MJ, Buchan I, Matteson EL, Montori V. Anti-TNF antibody therapy in rheumatoid arthritis and the risk of serious infections and malignancies: systematic review and meta-analysis of rare harmful effects in randomized controlled trials. *JAMA* (2006) 295(19):2275–85. doi: 10.1001/jama.295.19.2275

Check for
updates

Common Methods for Performing Mendelian Randomization

Alexander Teumer [1,2]*

[1] Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany, [2] Partner Site Greifswald, Deutsches Zentrum für Herz-Kreislaufforschung (DZHK), Greifswald, Germany

Mendelian randomization (MR) is a framework for assessing causal inference using cross-sectional data in combination with genetic information. This paper summarizes statistical methods commonly applied and strait forward to use for conducting MR analyses including those taking advantage of the rich dataset of SNP-trait associations that were revealed in the last decade through large-scale genome-wide association studies. Using these data, powerful MR studies are possible. However, the causal estimate may be biased in case the assumptions of MR are violated. The source and the type of this bias are described while providing a summary of the mathematical formulas that should help estimating the magnitude and direction of the potential bias depending on the specific research setting. Finally, methods for relaxing the assumptions and for conducting sensitivity analyses are discussed. Future researches in the field of MR include the assessment of non-linear causal effects, and automatic detection of invalid instruments.

Keywords: mendelian randomization, causal inference, GWAS, bias, statistical methods

INTRODUCTION

Observational epidemiological studies made important contributions to our understanding of common diseases by identifying important risk factors. Although causal inference is of major interest as it builds a basis for intervention and prevention, it is difficult to perform using observational data from cross-sectional studies. Supposed causality was often revised e.g., by randomized controlled trials (RCTs) (1). Possible reasons for these contradicting findings include unobserved confounding, reverse causation and selection bias in the observational studies (2–4).

On the other hand, RCTs are often subject to long duration and ethical problems. Furthermore, confounding and selection bias is still a problem after the initiation of a RCT. This includes compliance problems or missing of follow-up information depending on treatment effect which may induce missing not at random problems.

During the last decade, huge efforts were undertaken searching for genetic risk factors underlying common traits and diseases. Genome-wide association studies (GWAS) revealed thousands of genetic associations predominantly based on single nucleotide polymorphisms (SNPs) including more than 950 related to cardiovascular diseases and measurements (by April 2018) and were made publically available (5). The effect sizes of these associations were often quite small (6–8), and thus their direct clinical relevance might be questioned. However, these genetic associations may help drawing causal inferences. This approach in which SNPs are used as instrumental variables (IVs) for specific exposures is called Mendelian randomization (MR) (9). By the Mendelian laws, alleles of SNPs segregate and are randomly inherited from parents to offspring. This principle can be seen analogously to the randomized treatment assignment in a RCT resulting in an unconfounded exposure-outcome relationship. Within an MR approach, the exposure represents a continuous or dichotomous risk factor of a disease, and

the outcome is the disease or a disease-related trait. These traits may e.g., be blood pressure defining hypertension, or estimated glomerular filtration rate (eGFR) defining the status of chronic kidney disease. Using the MR approach, causality between exposure and outcome can be tested. During recent years, the number of MR studies to assess causality increased substantially which includes also the field of cardiovascular diseases and nephrology (10–14). Furthermore, MR analyses revealed causal effects of blood lipids on coronary heart disease (15) as well as of alcohol consumption on cardiovascular traits (16). However, given the number of potential genetic instruments and statistical methods available nowadays, there is potential for assessing causality of many more traits by conducting successful MR analyses. Nevertheless, some important assumptions have to be fulfilled to be able to estimate an unconfounded and unbiased exposure-outcome relationship thus allowing drawing causal inference. This review describes the assumptions of MR and potential biases caused by violation of these assumptions, and provides an overview of commonly applied statistical methods for conducting MR analyses using individual level data as well as using GWAS meta-analyses results.

Estimation of the Causal Effect

The general aim of the MR approach is the estimation of a causal effect of an exposure X on an outcome Y using (one or more) genetic instruments Z for X (**Figure 1**). Basically, the causal effect will be obtained by two sequential steps. First, the exposure is estimated from its instruments. By using valid instruments, the estimated exposure will be independent of any confounders. In the second step, the outcome is regressed on this estimated exposure thus obtaining an unconfounded and therefore causal

effect estimate. The instrument Z is usually coded by 0, 1 and 2 per individual according to its number of coding (e.g., exposure increasing) alleles.

2-Stage Least Squares Estimator

Given a continuous outcome Y and assuming linear effects between X and Y without interaction, the causal estimate of the exposure X on Y can be estimated through a 2-stage least squares (2SLS) regression. This method performs both steps described before implicitly. In the first step, the exposure $\hat{X}$ which is independent of the confounders is estimated via the genotypes of the instruments by calculating the fitted values from the regression of X on Z. In the second step, the causal effect estimate $\beta_{\hat{X}Y}$ is obtained by regressing Y on $\hat{X}$. As both steps are performed in a single model instead of two separate regressions, the variation of both Z and $\hat{X}$ is taken into account which is required for obtaining correct standard errors (SE) of $\beta_{\hat{X}Y}$ (17). The 2SLS regression can be calculated by standard methods in statistical software packages like R (18) using the function *tsls* of the package SEM, or by the STATA software (https://www.stata.com/) using the command *ivregress*. The 2SLS was included in an MR of testosterone with cardiometabolic risk factors, but the single study analysis limited the statistical power substantially (19).

Ratio Estimator

Alternatively, the causal effect can be estimated by triangulation without the need of calculating $\beta_{\hat{X}Y}$ from the exposure-outcome association directly. The principle of this method is illustrated through **Figure 1**: the standard approach (including 2SLS) for

FIGURE 1 | Directed acyclic graph showing the effects β of the genetic instrument Z, the exposure X, the outcome Y and the (unobserved) confounder U for illustrating the Mendelian randomization (MR). The dashed line represents the estimated causal effect $\beta_{\hat{X}Y}$ using the instrumented exposure. The dotted lines show violations of the MR assumptions 2 (lower line) and 3 (upper line), and are marked by a red cross. The α_{ZY} represents the effect of the instrument that affects the outcome not via the exposure in case of violating the exclusion restriction assumption. In contrast to α_{ZY}, the gray line illustrates the SNP-outcome association with its effect β_{ZY} that is used to calculate the two-sample MR given a valid instrument.

obtaining the causal effect $\beta_{\hat{X}Y}$ follows the path from the instrument Z via X to Y. In this case, the direct effect β_{ZY} of the instrument on the outcome Y equals the product of effects underlying the path mediated by the exposure, i.e., $\beta_{ZY} = \beta_{ZX} \cdot \beta_{\hat{X}Y}$. By rearranging this equation, the causal effect can be estimated through dividing the effect of the IV on the outcome (β_{ZY}) by the effect of the IV on the exposure (β_{ZX}): $\beta_{\hat{X}Y} = \beta_{ZY}/\beta_{ZX}$. As the triangulation approach calculates the causal effect (and its SE for testing significant deviation from null) by the ratio of the two IV based effect estimates, it is also known as ratio estimate or Wald estimate. It is important for the computation that both IV based effect estimates refer to the same allele of the IV. Furthermore, the same requirements as for the 2SLS apply. The SE of $\beta_{\hat{X}Y}$ has to be estimated via the delta-method which is based on a Taylor series expansion, and can be approximated as (20):

$$var\left(\beta_{\hat{X}Y}\right) = var\left(\frac{\beta_{ZY}}{\beta_{ZX}}\right) \cong \frac{var(\beta_{ZY})}{\beta_{ZX}^2} + \frac{\beta_{ZY}^2}{\beta_{ZX}^4}\, var\left(\beta_{ZX}\right) - 2\,\frac{\beta_{ZY}}{\beta_{ZX}^3}\, cov\left(\beta_{ZY}, \beta_{ZX}\right)$$

$$SE\left(\beta_{\hat{X}Y}\right) = \sqrt{var\left(\beta_{\hat{X}Y}\right)},$$ where $cov\left(\beta_{ZY}, \beta_{ZX}\right)$ is the covariance of the two effect estimates. This term will vanish if the effect estimates are obtained from distinct samples. That concise approximation can be easily implemented for significance testing in statistical software packages like R or STATA.

In contrast to the 2SLS which has to be performed using data of a single sample (one-sample MR), different sample sets can be used for conducting the triangulation: the effect estimates of the IV on exposure X and outcome Y can be obtained from genetic association studies with either disjunct or overlapping samples (two-sample MR). By this means, genetic associations revealed through large-scale GWAS meta-analyses can be used as β_{ZX} and β_{ZY}. These association results are often publically available for a variety of traits.

The triangulation method can also be applied if the outcome Y is dichotomous, i.e., an indicator of a disease status. In this case, log-linear effects without interaction on Y and an approximately normal distribution of X are required. Causal effect estimates on the odds ratio (OR) scale can be calculated by performing a logistic regression analysis using the disease as outcome. This model was also applied in most GWAS. To estimate causal OR using triangulation, the rare disease assumption (i.e., prevalence <10%) has to be fulfilled. Alternatively, estimates of a causal risk ratio may be calculated using a log-linear model instead of a logistic regression (21). The SE of the $\beta_{\hat{X}Y}$ (i.e., the log causal OR) will be estimated by the same formula as applied in the case of a continuous outcome. An application of the ratio estimator is provided by the MR on cystatin c and cardiovascular disease (22).

Control Function Estimator

Another method for estimating the causal effect on a dichotomous outcome is provided through the control function estimator (21) which is a two-step approach. In the first step, the exposure X is regressed on the instruments Z. The residuals of the regression correspond to the non-instrumented part of the exposure and may therefore correlate with a (unobserved) confounder U of the exposure-outcome association. In the second step, a logistic regression of the outcome Y on X is performed, adding the residuals of the first step as a covariate to the model. By adding the residuals of the first step into the model, the effects of U on Y will be controlled. Thus, the effect of X on Y of the second regression corresponds to the causal effect estimate. In case a linear regression is conducted in the second step (i.e., for a continuous outcome), the control function estimator is equivalent to the 2SLS estimator (21). This type of MR was conducted for assessing the causal effect of blood lipids on coronary heart disease (15).

Assumptions of the Instrumental Variables

SNPs have several properties predisposing them for instruments of the exposure. The inherited alleles are not changed by a disease or trait and thus also do not change over time. The random inheritance of the SNP alleles makes the genotype distribution mostly independent from socio-economic and lifestyle factors (1, 23). Nevertheless, specific assumptions still need to be fulfilled to ensure the validity of the genetic variant as an instrument. There are three core assumptions for MR (24–26):

1. The genetic variant is associated with the exposure
2. The genetic variant is independent of the outcome given the exposure and all confounders (measured and unmeasured) of the exposure-outcome association
3. The genetic variant is independent of factors (measured and unmeasured) that confound the exposure-outcome relationship

The first condition is required because within the MR the (unconfounded) exposure will be estimated using the allele distribution of the IVs. This assumption can be easily tested, and is considered as fulfilled if the SNP-exposure association has an F-statistic >10 (21, 27).

The second assumption, which is also known as exclusion restriction, is equivalent to the condition that an IV does not have an effect on the outcome when the exposure remains fixed. In general, this assumption is hard to validate as there may be pleiotropic effects of SNPs or SNPs in linkage disequilibrium correlated with genes that have effects on the outcome independently of the exposure. Even without considering the linkage disequilibrium, using SNPs of the pleiotropic gene *GCKR* exemplarily as instruments for kidney function to assess a causal effect on blood pressure would result in an invalid IV as there are effects of *GCKR* on blood pressure likely that are independent of kidney function, e.g., by the known associations of *GCKR* with serum lipid levels. Another violation would occur if the sample consists of a population substructure with different allele distributions, and which is also associated with the outcome. In this case, the substructure would be a common cause of both SNP and outcome opening a pathway from SNP to outcome not mediated by the exposure. Several examples of different scenarios violating the exclusion restriction are provided in the work of Glymour et al. (24).

The third assumption is also hard to validate. Similar problems due to pleiotropy and population substructure as described in the exclusion restriction may occur but affecting confounders of the

exposure-outcome relationship instead of the outcome directly. In an example of assessing causality of kidney function with heart disease, using *GCKR* as an instrument would violate the third assumption because these SNPs are also associated with blood pressure being a confounder of the association of kidney function and heart disease.

Weak Instrument Bias

Until today, more than 50,000 SNP-trait associations were revealed by GWAS and are usually accessible through public repositories like the GWASCatalog (5). These SNPs can be considered as potential instruments for MR analyses. Because the majority of these SNPs explain only a small proportion (i.e. <1%) of the phenotypic variance, GWAS with sample sizes of more than 10,000 or 100,000 individuals were required to unravel these associations at the level of genome-wide significance. However, the small effect sizes of the SNPs on the exposure result in weak instruments when using smaller sample sizes (28). Weak instruments tend to lead estimated causal effects towards the observational association (27). The reason for this bias is originated in using finite sample sizes. Although the IVs are asymptotically independent of confounders, there might be still an association by chance in finite samples. Increasing the sample size or the strength of the instruments will reduce the weak instrument bias. To illustrate the origin and the effect of the bias, let β_{UX} and β_{UY} be the effects of the confounder U on the exposure and the outcome, respectively (**Figure 1**). Furthermore, let Δ_U be the (by chance) difference in U depending on the instrument Z. The estimated causal effect $\beta_{\hat{X}Y}$ can then be computed by the following sum of effects (27):

$\beta_{\hat{X}Y} = \beta_{causal} + \frac{\beta_{UY}\Delta_U}{\beta_{ZX}+\beta_{UX}\Delta_U}$, where as β_{causal} is the true causal effect, and the mean(Δ_U) =0 because Z is an instrument (assumption 3). This leads the bias term towards zero with increasing sample size resulting in $\beta_{\hat{X}Y} = \beta_{causal}$. The estimated causal effect is also close to the true causal effect in case the effect of the IV on the exposure β_{ZX} is relatively large compared to the by-chance difference in U on the exposure ($\beta_{UX}\Delta_U$). However, if β_{ZX} is small compared to $\beta_{UX}\Delta_U$ (in case of a weak instrument), the estimated causal effect will be biased towards the ratio of the effect of the confounder on the outcome and the effect of the confounder on the exposure, i.e., $\frac{\beta_{UY}}{\beta_{UX}}$.

Multiple Instruments Approach

Using multiple valid instruments will help to address the weak instrument bias. Adding multiple uncorrelated (linkage equilibrium) SNPs into a 2SLS model can increase the statistical power but might also increase the relative bias if weak instruments are added (28).

Alternatively, an allele score can be generated from the instruments and included as a single variable in the association model. This allele score Z is calculated per individual as the weighted or unweighted sum of the number of risk or trait increasing alleles Z_i of each SNP i, whereas the effect β_{Z_iX} of each SNP on the exposure X is used as weight: $Z = \beta_{Z_1X}Z_1 + \beta_{Z_2X}Z_2 + \cdots + \beta_{Z_kX}Z_k$. In case of an unweighted score where all β_{Z_iX} are set to 1, the allele score of an individual simplifies to the sum of its risk alleles. By using an allele score, the F-statistics increases because of the smaller degrees of freedom in the model. However, it has been shown that the unweighted score has lower power than adding multiple IVs into the 2SLS, but using an appropriately weighted allele score performs similarly. The causal effect is a little less biased when using a weighted allele score but might have a slightly lesser precision (and power) compared to the multiple IV 2SLS estimator. In general, effects obtained from external studies should be used as weights (28).

A third method for taking advantage of multiple IVs is to combine ratio estimates (triangulation) of single instruments using inverse variance weighting (29, 30). The method for combing the results is the same as used for meta-analyses, and is for example implemented in the R package *metafor*. Alternatively, the following simplification of this calculation can be used (31–33):

$$\beta_{\hat{X}Y} = \frac{\sum \beta_{ZX}\beta_{ZY}var\left(\beta_{ZY}\right)^{-1}}{\sum \beta_{ZX}^2 var\left(\beta_{ZY}\right)^{-1}}$$

with its approximated $SE\left(\beta_{\hat{X}Y}\right) = \sqrt{\frac{1}{\sum \beta_{ZX}^2 var(\beta_{ZY})^{-1}}}$, where the sum runs over the SNP specific estimates. This method is implemented in the R package *gtx*.

However, it is crucial that the effects of all IVs used in the calculation are corresponding to the allele referring to the same effect direction on the exposure (e.g., the trait increasing allele). In theory, problems of missing data may occur especially when using multiple IVs. Nowadays well established methods for imputing missing genotypes based on the linkage disequilibrium structure of the human genome are available to circumvent this problem (34–36).

Bias by Violation of the Assumptions 2 and 3

Importantly, valid instruments need to be included in the MR analyses. In case the assumptions are not fulfilled, different types of bias can occur leading to invalid causal effect estimates. Violation of the second assumption (the exclusion restriction) implies that there is at least a partial effect of the instrument on the outcome not mediated by the exposure, i.e. $\alpha_{ZY} \neq 0$ (**Figure 1**). Depending on the direction and strength of these pleiotropic effects, the causal effect will be over- or underestimated. As shown within the principle of triangulation, the estimated causal effect $\beta_{\hat{X}Y}$ is the sum of the true causal effect β_{causal} and a bias term: $\beta_{\hat{X}Y} = \beta_{causal} + \frac{\alpha_{ZY}}{\beta_{ZX}}$ (26). The bias increases due to larger pleiotropy (larger absolute α_{ZY} in the nominator) or weaker instruments (smaller absolute β_{ZX} in the denominator). Violation of assumption 3 leads to a bias similar to the weak instrument bias. In this case, the effect of the confounder U on exposure and outcome will not vary by chance but systematically because of the non-zero effect of the instrument Z on U. Thus, an increasing sample size will not remove the bias because mean(Δ_U) $\neq$ 0.

InSIDE Condition and Egger MR

Pleiotropic effects α_{ZY} of each IV will also be included in the model when applying the multiple instruments approach. However, in this scenario it is possible to substitute the exclusion restriction by a weaker assumption as explained below. If the ratio estimates of multiple instruments are combined via 2SLS or inverse variance weighting, equation (1) will result in

$$\beta_{\hat{X}Y} = \beta_{causal} + \frac{\sum \beta_{ZX}\alpha_{ZY}var(\beta_{ZY})^{-1}}{\sum \beta_{ZX}^2 var(\beta_{ZY})^{-1}}, \text{ where } \beta_{causal} \text{ equals the}$$

right side of (1) and $\dfrac{\sum \beta_{ZX}\alpha_{ZY}var(\beta_{ZY})^{-1}}{\sum \beta_{ZX}^2 var(\beta_{ZY})^{-1}}$ is a bias depending on

α_{ZY} and β_{ZX}. Thus, an unbiased causal effect will be obtained if the assumption 2 is true, i.e., all direct effects α_{ZY} of each IV on the outcome Y are zero. However, it will be sufficient for the bias term to equal zero if all pleiotropic effects α_{ZY} of all genetic IV cancel out. As shown below, this cancellation is sufficiently fulfilled if the correlation between direct genetic effects α_{ZY} on the outcome and their effects β_{ZX} on the exposure X (i.e., the strength of the IV) is zero. This independence between the genetic effects α_{ZY} and β_{ZX} is called InSIDE condition (Instrument Strength Independent of Direct Effect). If the InSIDE condition holds together with assumptions 1 and 3, an adaption of the Egger regression can be used to obtain a consistent causal estimate even for specific cases in which the exclusion restriction criteria is violated. The Egger regression for MR is an implementation of the meta-regression where the (total) SNP-outcome effect $\Gamma = \beta_{ZX}\beta_{causal} + \alpha_{ZY}$ for each SNP is regressed on the corresponding SNP-exposure effect β_{ZX}: $\mathbf{\Gamma} \sim \beta_{0E} + \beta_E\boldsymbol{\beta_{ZX}}$ where the slope β_E is the bias-reduced causal estimate (**Figure 2**). The principle behind this regression is that

Γ is proportional to the strength of the instrument β_{ZX} with the intercept $\beta_{0E} = 0$ for valid instruments, whereas under the InSIDE condition (i.e., α_{ZY} and β_{ZX} are uncorrelated) stronger instruments are expected to have a relatively small bias and thus are on average closer to the true causal effect than weak instruments. As the slope of the Egger MR can be calculated by the least squares estimator $\beta_E = \frac{cov(\Gamma,\beta_{ZX})}{var(\beta_{ZX})} = \beta_{causal} + \frac{cov(\beta_{ZX},\alpha_{ZY})}{var(\beta_{ZX})}$, the bias term will be zero if α_{ZY} and β_{ZX} are uncorrelated, which is the case under the InSIDE assumption. A non-zero intercept β_{0E} indicates an overall directional pleiotropy of the IVs (26).

Considering Statistical Power

The statistical power of an MR strongly depends on the proportion of variance of the exposure that is explained by the IV. The use of multiple IVs, either by direct inclusion or as an allele score in the model, may therefore increase the power as more variance of the exposure is explained. However, the validity of these instruments has to be ensured (37). Two-sample MR additionally provide a possibility to increase statistical power if published GWAS meta-analyses of both the exposure and the outcome are available. In this case, effect estimates based on large sample sizes of independent studies can be used to estimate the causal effect. Formulas for performing power calculations of MR using single instruments or allele scores are provided in the study of Burgess (37). Brion et al. (38) discusses the statistical power in case of single IV and continuous outcomes for 2SLS MR, and provide an online power calculator for both continuous and binary outcomes which is available at http://cnsgenomics.com/shiny/mRnd/. A tool for estimating statistical power of complex MR settings based on simulations is MR_predictor (39), whereas the PERL scripts required to run the estimator are available via GitHub.

Measurement Unit of the Causal Effect Estimates

When conducting two-sample MR, the causal effect corresponds to the unit of the outcome on a per unit change of the exposure that was used in the respective genetic association study of the IV with the corresponding trait (32). Some GWAS were meta-analyzed using the sample-size weighted z-score method (40) and thus do not provide effect estimates that can be directly included in a two-sample MR. However, it is possible to estimate the effect $\hat{\beta}$ for each SNP in Hardy-Weinberg equilibrium using its minor allele frequency MAF, its (large) GWAS sample size N, and its z-statistics z (which can be calculated from the inverse of the standard normal distribution using the association p-value and the corresponding effect direction) through the formula (41): $\hat{\beta} \approx z \cdot \frac{\sigma}{\sqrt{N \cdot 2 \cdot MAF \cdot (1-MAF)}}$, whereas the corresponding

$SE\left(\hat{\beta}\right) = \frac{\hat{\beta}}{z}$. The SD σ of the trait can be set to 1 for standardizing the phenotype (i.e., the effect corresponds to a change of one SD of the trait unit). If the outcome is a binary trait, e.g., a disease with prevalence p in the sample, then $\sigma = \sqrt{p \cdot (1-p)}$.
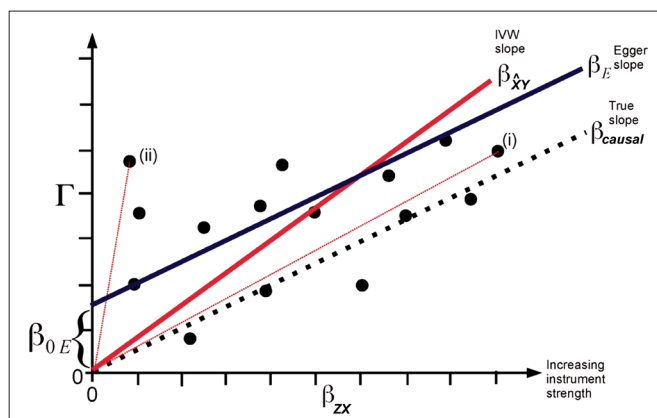

FIGURE 2 | Plot of the SNP-outcome (Γ) on the y-axis vs. the SNP-exposure (β_{ZX}) regression coefficients of potential genetic instruments (i.e., SNPs) of a Mendelian randomization analysis on the x-axis. The true causal effect represented by the slope β_{causal} is shown by a dotted line, the inverse variance weighted (IVW) causal estimate $\beta_{\hat{X}Y}$ by a red line, and the MR Egger regression estimate β_E by a dark blue line. The total SNP-outcome effect Γ is proportional to β_{ZX} for valid instruments. In case of invalid instruments but when the InSIDE assumption holds, stronger instruments are on average expected to be closer to the true causal effect (i) than weak instruments (ii). The intercept β_{0E} represents the overall directional pleiotropy of the instruments. The figure was adapted from the publication of Bowden et al., Int J Epidemiol. 2015;44(2):512–525 (26) (Creative Commons CC BY license).

DISCUSSION

MR provides a method for testing causality of different traits using cross-sectional data and genetics. Although large sample sizes are required to achieve sufficient statistical power for revealing causal effects, it is often possible to overcome this limitation by using the publically available genetic association results of large GWAS meta-analyses conducted during the last decade.

The statistical methods needed for conducting MR analyses are implemented in common statistical software frameworks. Additionally, the MRbase platform provides a possibility to conduct two-sample MR analyses both online and via the R package *TwoSampleMR*, including the methods discussed in this article (42). A detailed overview of different statistical methods for calculating MR is provided in the review of Burgess et al. (17).

However, it is important that the genetic associations that are used as instruments fulfil the MR assumptions to avoid calculation of biased or spurious causal estimates resulting in false causal inferences. Other than the required strong association of the genetic variant with the exposure, the remaining two assumptions are in general hard to validate.

This review emphasizes the bias that may occur by using invalid instruments, whereas the presented formulas should help estimating the magnitude and direction of this bias depending on the specific MR study that needs to be conducted. Using multiple instruments can help to test the violation of the MR assumptions which may occur due to pleiotropy and via SNPs in linkage disequilibrium (but not for a violation due to population stratification) (28), or to conduct sensitivity analyses (25). A strategy for assessing pleiotropy and population substructure specifically to MR analyses is discussed for example in the work of Lawler et al. (9). The Egger regression can be used as a multiple IV approach to relax the exclusion restriction criteria, and as a sensitivity analysis to test the robustness of the causal association (26). However, if the Egger MR-specific InSIDE assumption is violated, a biased causal estimate and an increased Type I error rate may occur (43). Thus, seeking for genetic variants that are valid IV should be performed as far as possible. Knowledge of the physiology or the biological pathways of the SNPs and their causal genes might be useful for selecting instruments.

The methods summarized in this review assume linear effects between exposure and outcome (or log-linear in case of a binary outcome) without effect modifications by the variables. Addressing these limitations is subject to future research. A method for successfully revealing non-linear causal effects was provided in an example for alcohol intake on cardiovascular traits, but this approach is restricted to additional assumptions and limitations (16). With respect to the presence of effect modifications, other statistical methods for conducting binary outcome MR like structural mean models or generalized method of moments make weaker assumptions but still not solve this issue completely (21). Finally, methods for automatically detecting invalid instruments (i.e., due to pleiotropy) are under development (44). Selection of valid instruments still remains a main challenge for automated causal inference.

AUTHOR CONTRIBUTIONS

AT designed and wrote the review.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

1. Davey Smith G, Ebrahim S. What can mendelian randomisation tell us about modifiable behavioural and environmental exposures? *BMJ* (2005) 330(7499):1076–9. doi: 10.1136/bmj.330.7499.1076

2. Lawlor DA, Davey Smith G, Kundu D, Bruckdorfer KR, Ebrahim S. Those confounded vitamins: what can we learn from the differences between observational versus randomised trial evidence? *Lancet* (2004) 363(9422):1724–7. doi: 10.1016/S0140-6736(04)16260-0

3. Lawlor DA, Davey Smith G, Ebrahim S. Commentary: the hormone replacement-coronary heart disease conundrum: is this the death of observational epidemiology? *Int J Epidemiol* (2004) 33(3):464–7. doi: 10.1093/ije/dyh124

4. Pilz S, Verheyen N, Grübler MR, Tomaschitz A, März W. Vitamin D and cardiovascular disease prevention. *Nat Rev Cardiol* (2016) 13(7):404–17. doi: 10.1038/nrcardio.2016.73

5. 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). *Nucleic Acids Res* (2017) 45(D1):D896–D901. doi: 10.1093/nar/gkw1133

6. Manolio TA. Genomewide association studies and assessment of the risk of disease. *N Engl J Med* (2010) 363(2):166–76. doi: 10.1056/NEJMra0905980

7. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. *Nature* (2009) 461(7265):747–53. doi: 10.1038/nature08494

8. Zuk O, Schaffner SF, Samocha K, do R, Hechter E, Kathiresan S, et al. Searching for missing heritability: designing rare variant association studies. *Proc Natl Acad Sci U S A* (2014) 111(4):E455–E464. . doi: 10.1073/pnas.1322563111

9. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat Med* (2008) 27(8):1133–63. doi: 10.1002/sim.3034

10. Sekula P, del Greco M F, Pattaro C, Köttgen A. Mendelian Randomization as an Approach to Assess Causality Using Observational Data. *J Am Soc Nephrol* (2016) 27(11):3253–65. doi: 10.1681/ASN.2016010098

11. Pattaro C. Genome-wide association studies of albuminuria: towards genetic stratification in diabetes? *J Nephrol* (2017) 0(0). doi: 10.1007/s40620-017-0437-3

12. Lanktree MB, Thériault S, Walsh M, Paré G, Cholesterol HDL. HDL Cholesterol, LDL Cholesterol, and Triglycerides as Risk Factors for CKD: A Mendelian Randomization Study. *Am J Kidney Dis* (2018) 71(2):166–72. doi: 10.1053/j.ajkd.2017.06.011

13. Ding EL, Song Y, Manson JE, Hunter DJ, Lee CC, Rifai N, et al. Sex hormone-binding globulin and risk of type 2 diabetes in women and men. *N Engl J Med* (2009) 361(12):1152–63. doi: 10.1056/NEJMoa0804381

14. van der Laan SW, Fall T, Soumaré A, Teumer A, Sedaghat S, Baumert J, et al. Cystatin C and Cardiovascular Disease: A Mendelian Randomization Study. *J Am Coll Cardiol* (2016) 68(9):934–45. doi: 10.1016/j.jacc.2016.05.092

15. Holmes MV, Asselbergs FW, Palmer TM, Drenos F, Lanktree MB, Nelson CP, et al. Mendelian randomization of blood lipids for coronary heart disease. *Eur Heart J* (2015) 36(9):539–50. doi: 10.1093/eurheartj/eht571

16. Silverwood RJ, Holmes MV, Dale CE, Lawlor DA, Whittaker JC, Smith GD, et al. Testing for non-linear causal effects using a binary genotype in a Mendelian randomization study: application to alcohol and cardiovascular traits. *Int J Epidemiol* (2014) 43(6):1781–90. doi: 10.1093/ije/dyu187

17. Burgess S, Small DS, Thompson SG. A review of instrumental variable estimators for Mendelian randomization. *Stat Methods Med Res* (2017) 26(5):2333–55. doi: 10.1177/0962280215597579

18. Core Team R. R: A Language and Environment for Statistical Computing. *https://www.r-project.org* (2016).

19. Haring R, Teumer A, Völker U, Dörr M, Nauck M, Biffar R, et al. Mendelian randomization suggests non-causal associations of testosterone with cardiometabolic risk factors and mortality. *Andrology* (2013) 1(1):17–23. doi: 10.1111/j.2047-2927.2012.00002.x

20. Thomas DC, Lawlor DA, Thompson JR. Re: Estimation of bias in nongenetic observational studies using "Mendelian triangulation" by Bautista et al. *Ann Epidemiol* (2007) 17(7):511–3. doi: 10.1016/j.annepidem.2006.12.005

21. Palmer TM, Sterne JA, Harbord RM, Lawlor DA, Sheehan NA, Meng S, et al. Instrumental variable estimation of causal risk ratios and causal odds ratios in Mendelian randomization analyses. *Am J Epidemiol* (2011) 173(12):1392–403. doi: 10.1093/aje/kwr026

22. van der Laan SW, Fall T, Soumaré A, Teumer A, Sedaghat S, Baumert J, et al. Cystatin C and Cardiovascular Disease. *J Am Coll Cardiol* (2016) 68(9):934–45. doi: 10.1016/j.jacc.2016.05.092

23. Hingorani A, Humphries S. Nature's randomised trials. *Lancet* (2005) 366(9501):1906–8. doi: 10.1016/S0140-6736(05)67767-7

24. Glymour MM, Tchetgen Tchetgen EJ, Robins JM. Credible Mendelian randomization studies: approaches for evaluating the instrumental variable assumptions. *Am J Epidemiol* (2012) 175(4):332–9. doi: 10.1093/aje/kwr323

25. Didelez V, Meng S, Sheehan NA. Assumptions of IV Methods for Observational Epidemiology. *Statistical Science* (2010) 25(1):22–40. doi: 10.1214/09-STS316

26. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* (2015) 44(2):512–. doi: 10.1093/ije/dyv080

27. Burgess S, Thompson SG CRP CHD Genetics Collaboration. Avoiding bias from weak instruments in Mendelian randomization studies. *Int J Epidemiol* (2011) 40(3):755–64. doi: 10.1093/ije/dyr036

28. Palmer TM, Lawlor DA, Harbord RM, Sheehan NA, Tobias JH, Timpson NJ, et al. Using multiple genetic variants as instrumental variables for modifiable risk factors. *Stat Methods Med Res* (2012) 21(3):223–42. doi: 10.1177/0962280210394459

29. Schmidt AF, Swerdlow DI, Holmes M V, Patel RS, Fairhurst-Hunter Z, Lyall DM. CSK9 genetic variants and risk of type 2 diabetes: a mendelian randomisation study. *lancet Diabetes Endocrinol* (2016).

30. Østergaard SD, Mukherjee S, Sharp SJ, Proitsi P, Lotta LA, Day F, et al. Associations between Potentially Modifiable Risk Factors and Alzheimer Disease: A Mendelian Randomization Study. *PLoS Med* (2015) 12(6):e1001841. doi: 10.1371/journal.pmed.1001841

31. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol* (2013) 37(7):658–65. doi: 10.1002/gepi.21758

32. Burgess S, Scott RA, Timpson NJ, Davey Smith G, Thompson SG EPIC-InterAct Consortium. Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. *Eur J Epidemiol* (2015) 30(7):543–52. doi: 10.1007/s10654-015-0011-z

33. Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, Chasman DI, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* (2011) 478(7367):103–9. doi: 10.1038/nature10405

34. Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, et al. A global reference for human genetic variation. *Nature* (2015) 526(7571):68–74. doi: 10.1038/nature15393

35. das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. *Nat Genet* (2016) 48(10):1284–7. doi: 10.1038/ng.3656

36. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* (2016) 48(10):1279–83. doi: 10.1038/ng.3643

37. Burgess S. Sample size and power calculations in Mendelian randomization with a single instrumental variable and a binary outcome. *Int J Epidemiol* (2014) 43(3):922–9. doi: 10.1093/ije/dyu005

38. Brion MJ, Shakhbazov K, Visscher PM. Calculating statistical power in Mendelian randomization studies. *Int J Epidemiol* (2013) 42(5):1497–501. doi: 10.1093/ije/dyt179

39. Voight BF. MR_predictor: a simulation engine for Mendelian Randomization studies. *Bioinformatics* (2014) 30(23):3432–4. doi: 10.1093/bioinformatics/btu564

40. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* (2010) 26(17):2190–1. doi: 10.1093/bioinformatics/btq340

41. Rietveld CA, Medland SE, Derringer J, Yang J, Esko T, Martin NW, et al. GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* (2013) 340(6139):1467–71. doi: 10.1126/science.1235488

42. Hemani G, Zheng J, Wade KH, Laurin C, Elsworth B, Burgess S. MR-Base: a platform for systematic causal inference across the phenome using billions of genetic associations. *bioRxiv* (2016) 078972.

43. Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the MR-Egger method. *Eur J Epidemiol* (2017) 32(5):377–89. doi: 10.1007/s10654-017-0255-x

44. Hemani G, Bowden J, Haycock PC, Zheng J, Davis O, Flach P. Automating Mendelian randomization through machine learning to construct a putative causal map of the human phenome. *bioRxiv* (2017) 173682. doi: doi:10.1101/173682

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

Translating GWAS Findings to Novel Therapeutic Targets for Coronary Artery Disease

Le Shu [1,2], Montgomery Blencowe [1] and Xia Yang [1,2,3,4,5]*

[1] Department of Integrative Biology and Physiology, University of California, Los Angeles, Los Angeles, CA, United States, [2] Molecular, Cellular, and Integrative Physiology Interdepartmental Program, University of California, Los Angeles, Los Angeles, CA, United States, [3] Bioinformatics Interdepartmental Program, University of California, Los Angeles, Los Angeles, CA, United States, [4] Institute for Quantitative and Computational Biosciences, University of California, Los Angeles, Los Angeles, CA, United States, [5] Molecular Biology Institute, University of California, Los Angeles, Los Angeles, CA, United States

The success of genome-wide association studies (GWAS) has significantly advanced our understanding of the etiology of coronary artery disease (CAD) and opens new opportunities to reinvigorate the stalling CAD drug development. However, there exists remarkable disconnection between the CAD GWAS findings and commercialized drugs. While this could implicate major untapped translational and therapeutic potentials in CAD GWAS, it also brings forward extensive technical challenges. In this review we summarize the motivation to leverage GWAS for drug discovery, outline the critical bottlenecks in the field, and highlight several promising strategies such as functional genomics and network-based approaches to enhance the translational value of CAD GWAS findings in driving novel therapeutics

Keywords: genome-wide association study, coronary artery disease, drug targets, multi-omics, functional genomics, networks

INTRODUCTION

Coronary artery disease (CAD) is a leading cause of mortality worldwide (1). CAD is well recognized as a complex disease with both genetic and environmental contributions (2). The heritability of CAD is estimated to be 40–50% (3), and the genetics of CAD plays an indispensable role in unraveling the pathogenic processes and ultimately facilitating the discovery of novel therapeutics. In the past decade, our understanding of the genetic architecture and mechanistic underpinnings for CAD has been substantially accelerated and broadened, primarily attributable to the successful global collaborative efforts in large-scale human genome-wide association studies (GWAS). These efforts have helped reveal hundreds of novel genetic variants demonstrating significant associations with CAD.

In contrast to the gratifying successes of GWAS, the development of CAD drugs has stagnated over the past decades, especially when compared to other therapeutic areas (4). What is particularly concerning is the fact that the drug development effort has been primarily concentrated on correcting previously established CAD risk factors such as lipid levels, coagulation factors, and hypertension, instead of targeting novel pathways revealed from recent studies (**Figure 1**). This decoupling between mechanistic discovery studies and drug development is striking. Therefore, it is of critical importance to form strategies that leverage the recent genetic discoveries from GWAS and other relevant efforts such as multi-dimensional data integration and systems genetics to allow for efficient identification of novel and reliable CAD drug targets. In this review, we summarize the state of CAD GWAS

FIGURE 1 | Summary of the reported CAD GWAS loci and important CAD drug discoveries. Candidate genes under GWAS loci identified to date were retrieved from the reported genes column in GWAS Catalog, organized by year. Only one candidate gene per locus was shown. GWAS loci that overlap with the targets of commercialized drugs were shown in red.
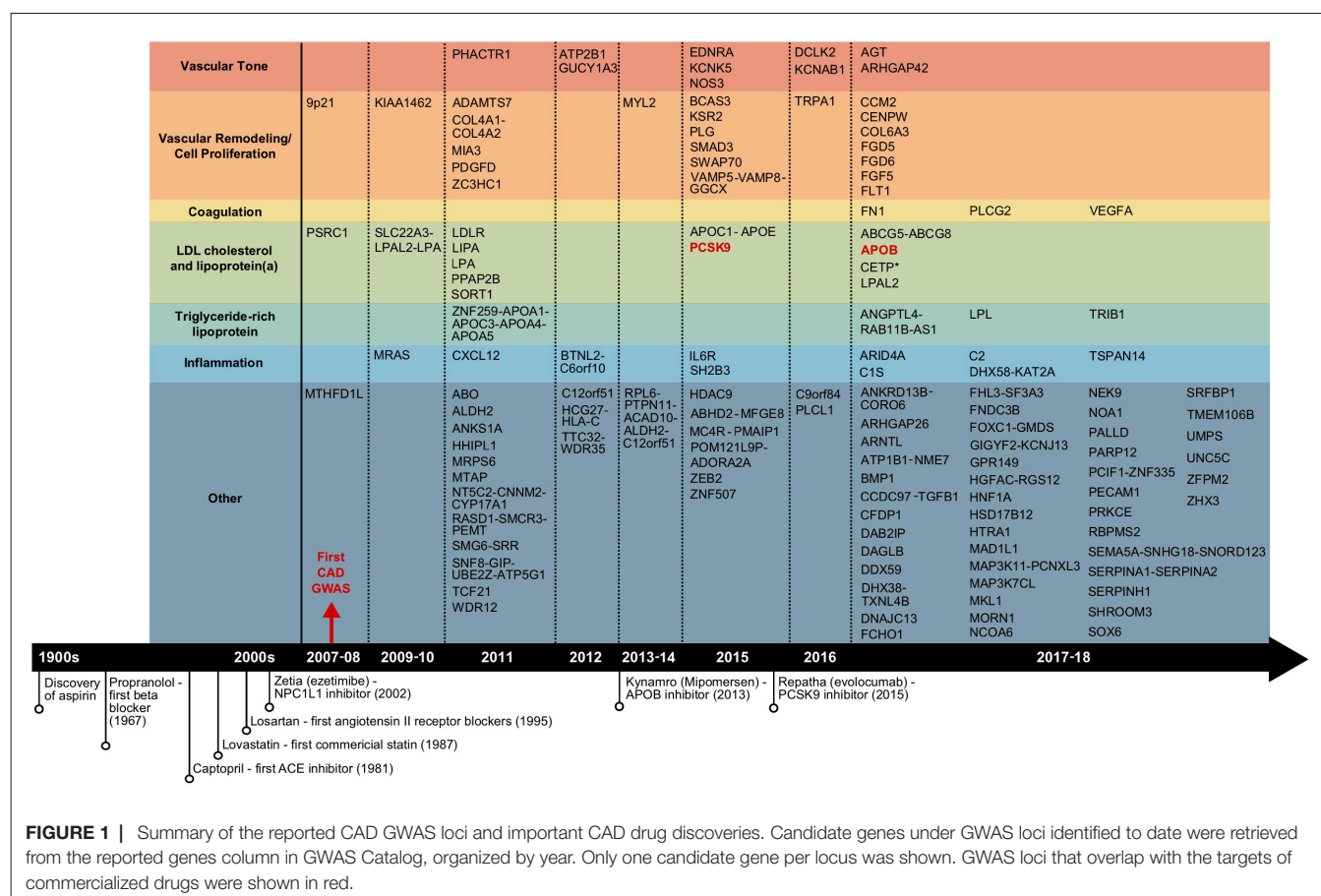
discovery, delineate the significant challenges of translating GWAS to drug targets, discuss successful examples of GWAS driven CAD drug target discovery, and outline promising strategies to further catalyze the translation of CAD GWAS into novel therapeutic options.

GWAS DISCOVERY FOR CAD AND ITS IMPLICATIONS FOR DRUG TARGET DISCOVERY

The completion of the human genome project, the rapid declining cost of genome sequencing, the rising feasibility of global multi-group collaborations, and the increasing accessibility of shared data repositories have collectively fueled the explosion of genetic studies of CAD, particularly GWAS. GWAS were typically designed to profile common known variants, often defined as variants with allele frequency ≥ 0.5% (5), on pre-designed microarrays containing primarily single nucleotide polymorphisms (SNPs). Since the first CAD GWAS in 2007 (6), over 18 GWAS studies have been carried out in the past decade, with the most recent and largest study involving 34,541 cases and 261,984 controls (7). These studies revealed a total of 163 genetic loci linked to CAD (8) (**Figure 1**), explaining 30–40% of CAD heritability (7, 9, 10). The swift pace of GWAS has greatly facilitated the comprehensive

construction of the CAD genetic landscape, and has led to rapid accumulation of potential causal variants and genes.

Overall, GWAS have played a key role in not only confirming classic CAD risk factors such as LDL cholesterol, hypertension, and coagulation, but also highlighting the causal roles of cellular proliferation and adhesion, extracellular matrix, and inflammation (**Figure 1**), which are processes related to the endothelial and smooth muscle cells in the vascular wall and the immune system (3, 11). Unfortunately, to date no novel CAD GWAS genes beyond a few involved in classic risk factors have been established as viable drug targets for CAD, a pattern that resonates for GWAS of most complex traits (**Figure 1**) (12). The disconnection between CAD GWAS findings and treatment targets is disappointing and has been criticized, but could also implicate major untapped opportunities (13). In particular, the causal variants and genes involved in the new causal pathways informed by GWAS have been encouraging early stage advances in uncovering novel therapeutic options targeting the vascular wall components, cell proliferation, and inflammation. For example, the *ADAMTS7* loci, coding for a metalloproteinase with thrombospondin motifs 7, was implicated for atherosclerotic progression through smooth muscle cell migration, a mechanism independent of classic CAD risk factors (14). Upon confirmation of its causal role in affecting atherosclerosis occurrence *in vivo* (15), development of the ADAMTS7 pharmacophore has progressed towards establishing inhibitors via virtual screening (16).

Tocilizumab, an anti-inflammatory agent blocking interleukin-6, was found to improve endothelial function (17). Antibodies targeting CD47, a key anti-phagocytic and tumorigenic molecule, were also shown to ameliorate atherosclerosis by stimulating efferocytosis (18).

Despite the potential promises, several factors could have complicated the extraction of therapeutic value from GWAS. First, the functional regulatory circuits from most variants to disease outcomes remain elusive. This is reflected by both the difficulty in pinpointing the causal variants and the corresponding target genes, especially for variants located in non-coding regions. In fact, the exact effector genes and functions for over 50% of the CAD GWAS loci are unclear. For example, the 9p21 locus was the strongest CAD locus but is located in a gene desert (6, 19, 20). Multiple follow-up studies have suggested several effectors for this locus, including the non-coding RNA ANRIL (21), *CDKN2A/CDKN2B* (22, 23), and interferon-gamma signaling (24). However, the detailed mechanism is still under debate after a decade of research (25). Moreover, even if a CAD variant is located within a gene-rich region, the most adjacent gene(s) may not be the functional candidate (26). Second, even if the candidate genes can be unequivocally determined, the functions of the genes are not necessarily well established, and extensive functional studies are required to derive a mechanistic understanding of how the candidate genes lead to CAD risks. Third, most common variants only confer weak to moderate CAD risk (<20% change in risk), most likely due to evolutionary pressure which selects against non-synonymous SNPs in disease genes involved in key physiological processes (12, 27–30). The prevalence of moderate/weak effect sizes of CAD risk variants makes prioritization of drug targets difficult. Lastly, it has been suspected that the top CAD risk variants identified so far predominantly inform on genes active in the early and slow phase of CAD development, whereas variants affecting late and rapid CAD phases tend to be missed by GWAS as these are likely more dependent on specific contexts such as particular environmental exposures or inflammatory states that are poorly controlled in most GWAS (31). Indeed, a recent study of Crohn's disease that focuses on disease course or prognosis using a within-cases design revealed loci that are completely different from those derived from case-control studies (32). This is also likely the case for CAD. Therefore, drug targets derived from CAD GWAS findings may not carry the expected efficacy to counteract CAD progression.

STRATEGIES TO FAST-FORWARD THE TRANSLATION OF GWAS TO TREATMENT TARGETS

To bypass the challenges facing the translation of GWAS findings to therapeutic targets as outlined above, a number of strategies have been designed and attempted. These efforts mainly focus on integrating GWAS hits with other data types that help inform on the functions of candidate genes, pathways, and networks, narrow down and prioritize the causal candidates, and leverage

the matching patterns between disease mechanisms and molecular patterns of drugs (**Figure 2**).

Use of Rare Variant Association Studies to Prioritize Targets

As discussed above, common variants uncovered from GWAS studies are numerous in number while carrying weak to subtle effect sizes, making it challenging to prioritize viable targets. Rare genetic variants (frequencies lower than 0.5%) that are associated with diseases, on the other hand, usually exhibit stronger perturbations in gene functions and are under stronger evolutionary pressure. Therefore, rare variants, especially those leading to loss-of-function, provide a natural setting mimicking human knockout cases to assess phenotypic and clinical consequences of variants, and their power in informing causal disease genes and drug targets has been long recognized (**Figure 2A**) (33, 34). Aggregation of rare mutations in 10 genes, including *APOA5* (35), *APOC3* (36), *ASGR1* (37), *ANGPTL3* (38), *ANGPTL4* (39), *LPA* (40), *LDLR* (35), *LPL* (41), *NPC1L1* (42), and *PCSK9* (43), has been linked to CAD risk through whole exome or whole genome sequencing-based studies. Out of the 10 genes, 5 (*ANGPTL4, APOC3, LPA, NPC1L1, PCSK9*) have been explored as drug targets for CAD (3). To date, the main success of this approach lies in the approval of PCSK9 antibodies by the FDA. Carriers with inactivating mutations on *PCSK9* were found to have markedly lower LDL cholesterol level and CAD risk, which led to the discovery of two FDA-approved monoclonal antibodies, Alirocumab and Evolucumab. However, the potential of drug discovery using CAD rare variants is also limited by both the small number of robust rare variants found so far, and their low cumulative contribution to CAD risk in the general population (9). Additionally, most of them are involved in the previously established pathways rather than novel mechanisms. Nevertheless, these rare variants provide compelling causal inference of the downstream genes and pathways in CAD pathogenesis, and they are more likely to be specific to a disease (broad effects could be detrimental in human knockouts) and have safer profiles, key components for success as drug targets. Future GWAS will likely evolve from SNP array design to whole genome sequencing to profile both common and rare variants (44), thus further expanding the pool of loss-of-function variants for drug target selection. Some of the novel rare variants may inform on novel causal mechanisms not captured by common variants, or converge on genes and pathways already informed by common variants thus serving to enhance the causal inference at a more functional level.

Functional Genomics to Identify and Prioritize Causal GWAS Genes

In contrast to rare coding mutations whose target genes and downstream mechanisms can be more readily uncovered through traditional functional studies, identifying the causal genes that are responsible for the observed link between GWAS risk variants and CAD is not an easy task. It is estimated that two thirds of the predicted target genes of GWAS locus are not the closest by proximity (45, 46), thus traditional proximity-
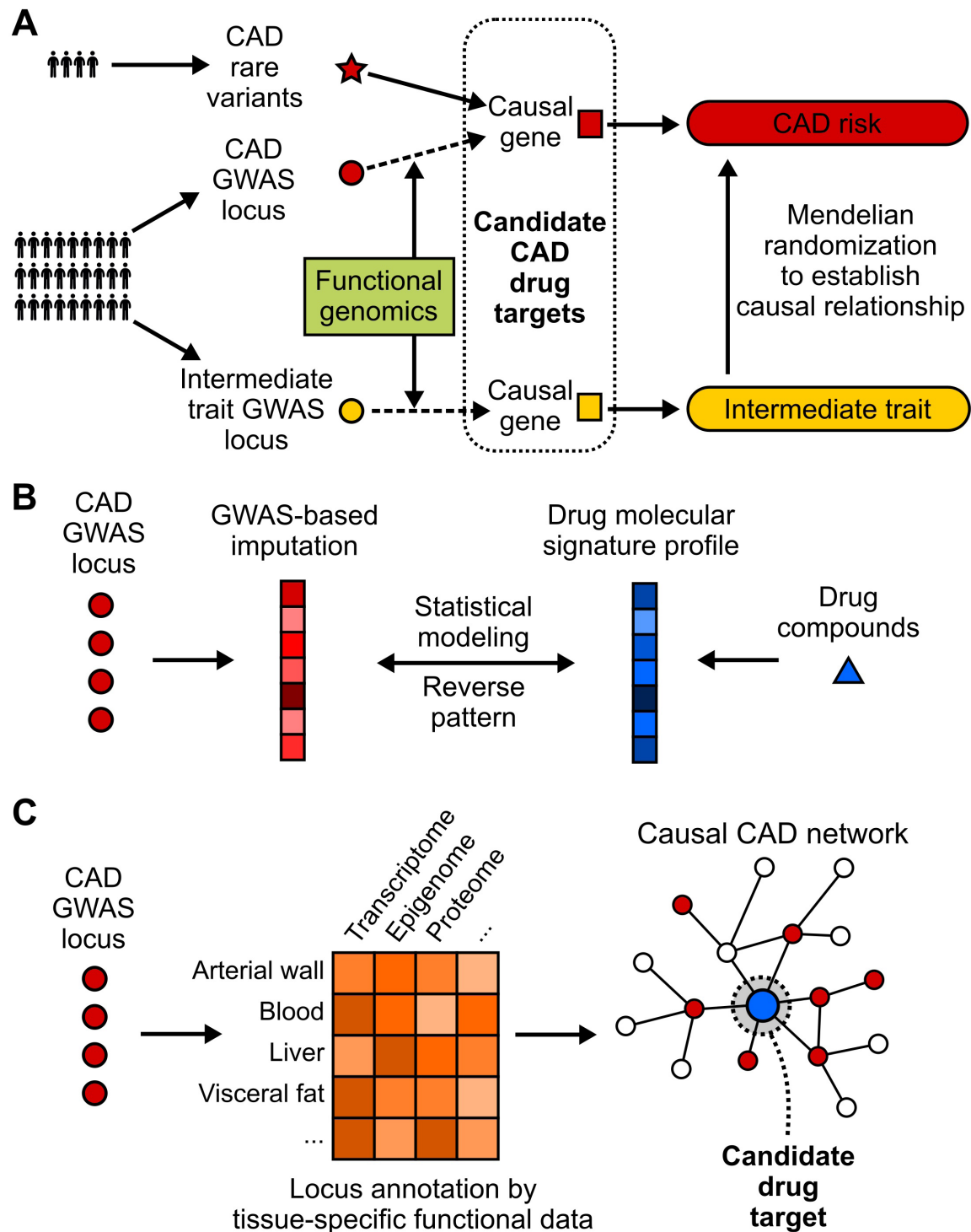
FIGURE 2 | Strategies to translate CAD GWAS into drug targets. **(A)** Identification of CAD causal genes as candidate drug targets by incorporating functional genomics, rare variants and Mendelian randomization. Loss-of-function rare variants can be linked to downstream genes. The connection between common variants and causal genes usually requires integration of functional genomics data. Mendelian randomization can further filter the drug target selection pool by incorporating causal intermediate traits. **(B)** A "target-less" approach to reposition existing drug compounds for CAD by evaluating the existence of opposite patterns between drug molecular profiles and GWAS imputed molecular profiles of disease. **(C)** Network-based approaches that model CAD GWAS data along with other omics data from CAD relevant tissues or cell types in the context of gene networks, which have the power to pinpoint key network regulators as candidate drug targets with more potent effects.

based locus mapping could introduce false interpretations that bias drug target selection. This challenge can be substantially alleviated by functional genomics tools that explore potential mechanisms linking causal variants to biological phenotypes (**Figure 2A**) (44). Supported by next-generation sequencing, typical functional molecular traits that may be characterized include expression quantitative trait locus (eQTL), non-coding RNA, transcription factor binding sites, epigenetic modification and chromatin interaction (26). The advance of gene editing technologies such as CRISPR/Cas has also significantly improved the efficiency of validation experiments (47). Recent functional genomics studies have substantially refined the candidate causal genes for CAD loci such as *SORT1* (48), *TRIB1* (49), *ADAMTS7* (15, 50), and *TCF21* (51, 52). Noteworthy, there have also been integrative functional genomics studies that combined genomics, epigenomics and transcriptomics profiling to prioritize causal variants and affected genes (7, 46, 53). For example, Miller et al. integrated Assay for Transposase Accessible Chromatin (ATAC-seq) and chromatin immunoprecipitation-sequencing (ChIP-seq) to unravel the cis-regulatory mechanisms in human coronary artery smooth muscle cells, and prioritized 64 variants over 7 candidate CAD loci including 9p21.3, *SMAD3, PDGFD, IL6R, BMP1, CCDC97/TGFB1* and *LMOD1* (53). Haitjema et al. also leveraged circular chromosome conformation capture sequencing (4C-seq) with RNA-seq and eQTL to identify 294 novel candidate CAD genes (54). These studies greatly contribute to the accumulation of viable treatment targets for follow-up drug development efforts.

Encouragingly, functional studies following GWAS are being further catalyzed by large-scale community efforts in establishing multi-cell or multi-tissue mapping of regulatory annotations. The advent of publicly available depositories such as GTEx (55), ENCODE (56) and Epigenome Roadmap (57) is gradually removing the hurdle to acquire multi-dimensional data resources necessary for the investigation of complex traits like CAD.

Mendelian Randomization (MR) to Facilitate Drug Target Selection

Previous successes in drug development for CAD have testified to the effectiveness of modulating intermediate causal risk factors such as circulating cholesterol levels and blood pressure in lowering CAD risk. Therefore, knowing the causal relationship between an intermediate phenotype that correlates with CAD status is of monumental importance as it can help prioritize biomarkers as intervention targets for CAD therapeutics (58, 59) (**Figure 2A**). The investigation of causal intermediate traits for CAD can be facilitated by MR, which utilizes genetic variants as instrumental variables to assess the causal relationship between exposure (e.g., LDL cholesterol, HDL cholesterol, weight-hip ratio) and outcome (CAD occurrence) (60). We are seeing both successful and ongoing efforts in developing drugs modulating LDL cholesterol, triglyceride-rich lipoproteins and lipoprotein (a) (3), whose causal relationships with CAD have been robustly verified in MR studies (61–63). On the contrary, MR studies revealed inconsistent relationship between HDL cholesterol and CAD (64–66). In concordance with this lack of robust support for the causality of HDL in CAD, substantial

obstacles have been met during the development of inhibitors for CETP (cholesteryl ester transfer protein), a gene harboring several loci associated with HDL cholesterol level (67). Three commercial CETP inhibitors, Dalcetrapib, Obicetrapib and Anacetrapib, all failed to achieve clinical efficacy during phase III clinical trials and were discontinued (68).

In addition to the well explored causal pathways such as cholesterol and blood pressure regulation, MR studies have informed several additional causal intermediate phenotypes, such as inflammation (69), uric acid (70), and iron status (71), that could serve as targets for future CAD drug development. By utilizing both summary-level GWAS statistics and UK Biobank data, a recent MR study demonstrated the causal association of waist-to-hip ratio adjusted for body mass index with coronary heart disease, thus providing new opportunities of intervening CAD risk by reducing abdominal obesity (72).

GWAS-Based "Target-Free" Drug Repositioning

Drug repurposing approaches could leverage known drugs used for other diseases that target the newly uncovered CAD causal genes and pathways to counteract CAD. For example, better understanding of CAD pathways involved in inflammation and cell cycle has promoted the repurposing of drugs targeting diseases such as rheumatoid arthritis (17) and cancer (18). On the other hand, given the challenging nature of identifying both the causal genes from GWAS and matching it with the target of drug compounds, "target-free" approaches have been developed which require no prior knowledge of targets for either drugs or GWAS variants and can simultaneously take many genetic loci into consideration (73) (**Figure 2B**). The fundamental concept behind these approaches is to impute gene expression profiles from GWAS summary statistics, compare the expression patterns against gene expression profiles of drugs, then prioritize top drug candidates whose profiles show reverse patterns compared with GWAS-imputed signatures. This approach is especially useful for repositioning existing drugs whose chemical properties and molecular responses have been well characterized and made accessible from public data repositories such as CMap (74) and its successor, the L1000 platform (75), as well as other chemoinformatic resources (76, 77).

To facilitate such efforts, the work by Gamazon et al. represents one of the first transcriptome imputation pipelines where disease relevant gene expression is estimated from a tissue-dependent model trained with personal genotype data and reference transcriptome (78). Gusev et al. and So et al. further developed summary GWAS statistics based transcriptome imputation methods, which relieved the requirement for individual genotype data (73, 79). In addition, inferring gene expression changes from GWAS enables researchers to assess transcriptome-wide associations with CAD that could yield novel candidate genes for functional and therapeutic investigation (45, 79). Although direct application of the "target-free" approaches for CAD is still under-explored, a computational framework has been developed to reposition existing drugs for psychiatry (73). The framework, built on a GWAS-based transcriptome imputation pipeline named MetaXcan (80), first imputed the gene expression

profiles of 10 brain regions for 7 psychiatric disorders based on GWAS and reference transcriptome data from GTEx (55). This disease transcriptome information was then used to match with drug-induced gene expression profiles from the CMap database (74) to prioritize drugs that showed opposite gene expression patterns compared to the disease patterns. These platforms are potentially translatable to CAD.

Network-Based Drug Discovery Approaches

The success of GWAS-driven drug target identification heavily relies on the fundamental assumption of how genetic risk variants eventually contribute to disease phenotype. An "omnigenic" model for the genetics of complex traits has been recently proposed (30). This provocative model objects the common belief that risk variants drive disease etiology through functional clustering in biological pathways, and emphasizes that all genes in disease-relevant cells could affect core disease processes through the coordination of gene regulatory networks.

Motivated by the gene network hypothesis, the CAD field has been actively investing on the development and application of systems genetics frameworks that integrate genetics and other data dimensions in the context of network topology to help prioritize candidate CAD genes (**Figure 2C**) (27–29, 81–84). The implementation of network-based target identification strategies poses several unique advantages over other methods. First, gene networks have the potential to comprehensively map the regulatory circuits under physiological or pathological conditions, thus improving the biological relevance of predicted targets. Second, gene networks serve as a natural platform for data integration, where GWAS and information from other omics space can be collectively leveraged to pinpoint network hotspots where key perturbation events likely happen. Third, gene networks enable the identification of essential disease genes, which is unlikely to tolerate high frequency loss-of-function variation at the population level and to be discovered by GWAS (85). Several methods have been developed to find network essential genes, or key drivers, by considering both network topology and external disease signatures (27, 81, 86). The validity of the predicted key drivers in driving CAD relevant traits has been well supported (27, 29, 82), and the key drivers have the potential to serve as novel drug targets with strong therapeutic effects due to their central importance in regulating the disease networks. For instance, Zhao et al. recently prioritized CAD key drivers and proposed plausible targets using network approaches (28). Extensive *in vitro* and *in vivo* gene perturbation experiments are required to evaluate the feasibility of using key driver genes as drug targets. If proven valid, network-based discoveries could provide exciting opportunities to formulate more focused and data-informed hypotheses for downstream therapeutic investigation. Nevertheless, it is important to caution that modulation of network key drivers may result in a lack of specificity and increase the risks for side effects due to their broad impact on numerous network genes.

One critical challenge for network-based CAD drug discovery is the availability of high-quality gene networks from CAD relevant cells, tissues, and subjects. Many existing networks are literature-based and lack tissue/cell specificity. Even for data-driven networks, data collection bias exists. For example, human network construction usually requires large numbers of clinical samples that are difficult to acquire, especially for samples from internal tissues. A major breakthrough is the establishment of the STARNET networks involving tissue-specific data from ~600 CAD patients (87). This resource, in combination with other networks generated from mouse models or non-disease human subjects, is invaluable for future CAD network studies. Coordinated efforts by the research community are needed to enhance the coverage of data-driven networks from CAD relevant tissues and cell types.

CONCLUSIONS AND FUTURE DIRECTIONS

GWAS has been highly successful in elucidating the genetic architecture of CAD and driving the discovery of novel biology. While confirming the genes and pathways targeted by classic CAD treatments, GWAS opens doors to a vast number of under-recognized candidates where future CAD drugs could originate. The field of GWAS-driven drug discovery is still at its infancy, and significant challenges remain. However, it is encouraging that numerous methodological advances have been made to address the bottlenecks, and application of these approaches is expected to facilitate future translational research in CAD.

Here we anticipate the following future directions that will help further advance the field. First, there is a need for broader collaboration to conduct large-scale functional genomics studies in human tissues and cell types that implement cutting-edge high-throughput profiling technologies over multiple omics to map the tissue- and cell-type specific regulatory circuit of GWAS loci. In particular, application of cell-type specific analyses at multi-omics levels will help address the functional heterogeneity in CAD relevant tissues, which will lead to refined understanding of disease etiology and lay a solid foundation for more accurate prediction of drugs that can counteract the specific pathogenic processes in the right cell types and tissues (88). The recent launch of the Human Cell Atlas project represents one of the first stepping stones towards this direction (89). Second, more efficient platforms are needed to facilitate sharing of summary-level GWAS data as well as databases and data repositories that curate multi-omics functional information. For example, in the neurological disorder field, there are emerging efforts like CommonMind (http://commonmind.org), PsychENCODE (90) and BrainSeq (91). Similar coordinated efforts by the CAD community will accelerate identification of CAD drug targets. Third, the translational value of GWAS data can be better exploited by the development of novel analytical pipelines that integrate multi-dimensional data from animals, humans, and chemoinformatic databases. Some of the recently developed analytical pipelines can integrate GWAS and functional genomics data for target prediction, and are directly applicable to CAD

(45, 79, 81, 92). Additional pipelines that couple disease datasets with drug footprints in a gene network framework will facilitate the identification of network regulators and pathways that can be accurately targeted. Lastly, gradual transition from initial target screening to GWAS-guided experimental validation of the predicted targets using a combination of *in vitro*, *in vivo*, and *in silico* methods will further the translational path.

AUTHOR CONTRIBUTIONS

LS, MB and XY drafted and edited the manuscript.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

1. Abubakar I, Tillmann T, Global BA. regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* (2015) 385(9963):117–71.
2. Khera AV, Emdin CA, Drake I, Natarajan P, Bick AG, Cook NR, et al. Genetic risk, adherence to a healthy lifestyle, and coronary disease. *N Engl J Med Overseas Ed* (2016) 375(24):2349–58. doi: 10.1056/NEJMoa1605086
3. Khera AV, Kathiresan S. Genetics of coronary artery disease: discovery, biology and clinical translation. *Nat Rev Genet* (2017) 18(6):331–44. doi: 10.1038/nrg.2016.160
4. Fordyce CB, Roe MT, Ahmad T, Libby P, Borer JS, Hiatt WR, et al. Cardiovascular drug development: is it dead or just hibernating? *J Am Coll Cardiol* (2015) 65(15):1567–82.
5. Zuk O, Schaffner SF, Samocha K, do R, Hechter E, Kathiresan S, et al. Searching for missing heritability: Designing rare variant association studies. *Proc Natl Acad Sci U S A* (2014) 111(4):E455–64. doi: 10.1073/pnas.1322563111
6. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* (2007) 447(7145):661–78. doi: 10.1038/nature05911
7. Van der Harst P, Verweij N. Identification of 64 novel genetic loci provides an expanded view on the genetic architecture of coronary artery disease novelty and significance. *Circ Res* (2018) 122(3):433–43. doi: 10.1161/CIRCRESAHA.117.312086
8. Erdmann J, Kessler T, Munoz Venegas L, Schunkert H. A decade of genome-wide association studies for coronary artery disease: the challenges ahead. *Cardiovasc Res* (2018) 385. doi: 10.1093/cvr/cvy084
9. Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, et al. A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet* (2015) 47(10):1121–30. doi: 10.1038/ng.3396
10. Klarin D, Zhu QM, Emdin CA, Chaffin M, Horner S, Mcmillan BJ, et al. Genetic analysis in UK Biobank links insulin resistance and transendothelial migration pathways to coronary artery disease. *Nat Genet* (2017) 49(9):1392–7. doi: 10.1038/ng.3914
11. Kessler T, Vilne B, Schunkert H. The impact of genome-wide association studies on the pathophysiology and therapy of cardiovascular disease. *EMBO Mol Med* (2016) 8(7):688–701. doi: 10.15252/emmm.201506174
12. Cao C, Moult J. GWAS and drug targets. *BMC Genomics* (2014) 15(Suppl 4):S5. doi: 10.1186/1471-2164-15-S4-S5
13. Schooling CM, Huang JV, Zhao JV, Kwok MK, Au Yeung SL, Lin SL. Disconnect between genes associated with ischemic heart disease and targets of ischemic heart disease treatments. *EBioMedicine* (2018) 28:311–5. doi: 10.1016/j.ebiom.2018.01.015
14. Reilly MP, Li M, He J, Ferguson JF, Stylianou IM, Mehta NN, et al. Identification of ADAMTS7 as a novel locus for coronary atherosclerosis and association of ABO with myocardial infarction in the presence of coronary atherosclerosis: two genome-wide association studies. *The Lancet* (2011) 377(9763):383–92. doi: 10.1016/S0140-6736(10)61996-4

15. Bauer RC, Tohyama J, Cui J, Cheng L, Yang J, Zhang X, et al. Knockout of *Adamts7*, a novel coronary artery disease locus in humans, reduces atherosclerosis in mice. *Circulation* (2015) 131(13):1202–13. doi: 10.1161/CIRCULATIONAHA.114.012669
16. Müller M, Kessler T, Schunkert H, Erdmann J, Tennstedt S. Classification of ADAMTS binding sites: The first step toward selective ADAMTS7 inhibitors. *Biochem Biophys Res Commun* (2016) 471(3):380–5. doi: 10.1016/j.bbrc.2016.02.025
17. Bacchiega BC, Bacchiega AB, Usnayo MJG, Bedirian R, Singh G, Pinheiro Gdarc. Interleukin 6 Inhibition and coronary artery disease in a high-risk population: A prospective community-based clinical study. *J Am Heart Assoc* (2017) 6(3):e005038. doi: 10.1161/JAHA.116.005038
18. Kojima Y, Volkmer J-P, Mckenna K, Civelek M, Lusis AJ, Miller CL, et al. CD47-blocking antibodies restore phagocytosis and prevent atherosclerosis. *Nature* (2016) 536(7614):86–90. doi: 10.1038/nature18935
19. 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(5830):1491–3. doi: 10.1126/science.1142842
20. 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(5830):1488–91. doi: 10.1126/science.1142447
21. 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
22. Kojima Y, Downing K, Kundu R, Miller C, Dewey F, Lancero H, et al. Cyclin-dependent kinase inhibitor 2B regulates efferocytosis and atherosclerosis. *J Clin Invest* (2014) 124(3):1083–97. doi: 10.1172/JCI70391
23. Kim JB, Deluna A, Mungrue IN, Vu C, Pouldar D, Civelek M, et al. Effect of 9p21.3 coronary artery disease locus neighboring genes on atherosclerosis in mice. *Circulation* (2012) 126(15):1896–906. doi: 10.1161/CIRCULATIONAHA.111.064881
24. Harismendy O, Notani D, Song X, Rahim NG, Tanasa B, Heintzman N, et al. 9p21 DNA variants associated with coronary artery disease impair interferon-γ signalling response. *Nature* (2011) 470(7333):264–8. doi: 10.1038/nature09753
25. Pjanic M, Miller CL, Wirka R, Kim JB, Direnzo DM, Quertermous T. Genetics and genomics of coronary artery disease. *Curr Cardiol Rep* (2016) 18(10):102. doi: 10.1007/s11886-016-0777-y
26. 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(4):586–606.
27. Mäkinen V-P, Civelek M, Meng Q, Zhang B, Zhu J, Levian C, et al. Integrative genomics reveals novel molecular pathways and gene networks for coronary artery disease. *PLoS Genet* (2014) 10(7):e1004502. doi: 10.1371/journal.pgen.1004502
28. Zhao Y, Chen J, Freudenberg JM, Meng Q, Rajpal DK, Yang X. Network-based identification and prioritization of key regulators of coronary artery disease loci. *Arterioscler Thromb Vasc Biol* (2016) 36(5):928–41. doi: 10.1161/ATVBAHA.115.306725

29. Shu L, Chan KHK, Zhang G, Huan T, Kurt Z, Zhao Y, et al. Shared genetic regulatory networks for cardiovascular disease and type 2 diabetes in multiple populations of diverse ethnicities in the United States. *PLoS Genet* (2017) 13(9):e1007040. doi: 10.1371/journal.pgen.1007040

30. Boyle EA, Li YI, Pritchard JK. An expanded view of complex traits: From polygenic to omnigenic. *Cell* (2017) 169(7):1177–86. doi: 10.1016/j.cell.2017.05.038

31. Björkegren JL, Kovacic JC, Dudley JT, Schadt EE. Genome-wide significant loci: how important are they?: systems genetics to understand heritability of coronary artery disease and other common complex disorders. *J Am Coll Cardiol* (2015) 65(8):830–45.

32. Lee JC, Biasci D, Roberts R, Gearry RB, Mansfield JC, Ahmad T, et al. Genome-wide association study identifies distinct genetic contributions to prognosis and susceptibility in Crohn's disease. *Nat Genet* (2017) 49(2):262–8. doi: 10.1038/ng.3755

33. Macarthur DG, Balasubramanian S, Frankish A, Huang N, Morris J, Walter K, et al. A systematic survey of loss-of-function variants in human protein-coding genes. *Science* (2012) 335(6070):823–8. doi: 10.1126/science.1215040

34. Bomba L, Walter K, Soranzo N. The impact of rare and low-frequency genetic variants in common disease. *Genome Biol* (2017) 18(1):77. doi: 10.1186/s13059-017-1212-4

35. do R, Stitziel NO, Won H-H, Jørgensen AB, Duga S, Angelica Merlini P, et al. Exome sequencing identifies rare LDLR and APOA5 alleles conferring risk for myocardial infarction. *Nature* (2015) 518(7537):102–6. doi: 10.1038/nature13917

36. Jørgensen AB, Frikke-Schmidt R, Nordestgaard BG, Tybjærg-Hansen A. Loss-of-function mutations in APOC3 and risk of ischemic vascular disease. *N Engl J Med Overseas Ed* (2014) 371(1):32–41. doi: 10.1056/NEJMoa1308027

37. Nioi P, Sigurdsson A, Thorleifsson G, Helgason H, Agustsdottir AB, Norddahl GL, et al. Variant ASGR1 associated with a reduced risk of coronary artery disease. *N Engl J Med Overseas Ed* (2016) 374(22):2131–41. doi: 10.1056/NEJMoa1508419

38. Stitziel NO, Khera AV, Wang X, Bierhals AJ, Vourakis AC, Sperry AE, et al. ANGPTL3 deficiency and protection against coronary artery disease. *J Am Coll Cardiol* (2017) 69(16):2054–63. doi: 10.1016/j.jacc.2017.02.030

39. Myocardial Infarction Genetics, C. ARDIoGRAM Exome Consortia Investigators, Stitziel NO, Stirrups KE, Masca NG, Erdmann J, et al. Coding Variation in ANGPTL4, LPL, and SVEP1 and the Risk of Coronary Disease. *N Engl J Med* (2016) 374(12):1134–44.

40. Emdin CA, Khera AV, Natarajan P, Klarin D, Won H-H, Peloso GM, et al. Phenotypic characterization of genetically lowered human lipoprotein(a) levels. *J Am Coll Cardiol* (2016) 68(25):2761–72. doi: 10.1016/j.jacc.2016.10.033

41. Khera AV, Won H-H, Peloso GM, O'Dushlaine C, Liu D, Stitziel NO, et al. Association of rare and common variation in the lipoprotein lipase gene with coronary artery disease. *JAMA* (2017) 317(9):937–46. doi: 10.1001/jama.2017.0972

42. Myocardial Infarction Genetics Consortium Investigators, Stitziel NO, Won HH, Morrison AC, Peloso GM, Do R, et al. Inactivating mutations in NPC1L1 and protection from coronary heart disease. *N Engl J Med* (2014) 371(22):2072–82. doi: 10.1056/NEJMoa1405386

43. Cohen JC, Boerwinkle E, Mosley TH, Hobbs HH. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *N Engl J Med Overseas Ed* (2006) 354(12):1264–72. doi: 10.1056/NEJMoa054013

44. Visscher PM, Wray NR, Zhang Q, Sklar P, Mccarthy MI, Brown MA, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am J Hum Genet* (2017) 101(1):5–22. doi: 10.1016/j.ajhg.2017.06.005

45. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet* (2016) 48(5):481–7. doi: 10.1038/ng.3538

46. Brænne I, Civelek M, Vilne B, Di Narzo A, Johnson AD, Zhao Y, et al. Prediction of causal candidate genes in coronary artery disease loci. *Arterioscler Thromb Vasc Biol* (2015) 35(10):2207–17. doi: 10.1161/ATVBAHA.115.306108

47. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* (2013) 339(6121):819–23. doi: 10.1126/science.1231143

48. Musunuru K, Strong A, Frank-Kamenetsky M, Lee NE, Ahfeldt T, Sachs KV, et al. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature* (2010) 466(7307):714–9. doi: 10.1038/nature09266

49. Douvris A, Soubeyrand S, Naing T, Martinuk A, Nikpay M, Williams A, et al. Functional analysis of the TRIB1 associated locus linked to plasma triglycerides and coronary artery disease. *J Am Heart Assoc* (2014) 3(3):e000884. doi: 10.1161/JAHA.114.000884

50. Pu X, Xiao Q, Kiechl S, Chan K, Ng FL, Gor S, et al. ADAMTS7 cleavage and vascular smooth muscle cell migration is affected by a coronary-artery-disease-associated variant. *The American Journal of Human Genetics* (2013) 92(3):366–74. doi: 10.1016/j.ajhg.2013.01.012

51. Miller CL, Haas U, Diaz R, Leeper NJ, Kundu RK, Patolla B, et al. Coronary heart disease-associated variation in TCF21 disrupts a miR-224 binding site and miRNA-mediated regulation. *PLoS Genet* (2014) 10(3):e1004263. doi: 10.1371/journal.pgen.1004263

52. Nurnberg ST, Cheng K, Raiesdana A, Kundu R, Miller CL, Kim JB, et al. Coronary artery disease associated transcription factor TCF21 regulates smooth muscle precursor cells that contribute to the fibrous cap. *PLoS Genet* (2015) 11(5):e1005155. doi: 10.1371/journal.pgen.1005155

53. Miller CL, Pjanic M, Wang T, Nguyen T, Cohain A, Lee JD, et al. Integrative functional genomics identifies regulatory mechanisms at coronary artery disease loci. *Nat Commun* (2016) 7:12092. doi: 10.1038/ncomms12092

54. Haitjema S, Meddens CA, van der Laan SW, Kofink D, Harakalova M, Tragante V, et al. Additional candidate genes for human atherosclerotic disease identified through annotation based on chromatin organization. *Circ Cardiovasc Genet* (2017) 10(2):e001664. doi: 10.1161/CIRCGENETICS.116.001664

55. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* (2013) 45(6):580–5. doi: 10.1038/ng.2653

56. Encode Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* (2012) 489(7414):57–74. doi: 10.1038/nature11247

57. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, et al. Integrative analysis of 111 reference human epigenomes. *Nature* (2015) 518(7539):317–30. doi: 10.1038/nature14248

58. Paternoster L, Tilling K, Davey Smith G. Genetic epidemiology and Mendelian randomization for informing disease therapeutics: Conceptual and methodological challenges. *PLoS Genet* (2017) 13(10):e1006944. doi: 10.1371/journal.pgen.1006944

59. Jansen H, Samani NJ, Schunkert H. Mendelian randomization studies in coronary artery disease. *Eur Heart J* (2014) 35(29):1917–24. doi: 10.1093/eurheartj/ehu208

60. Lawlor DA, Harbord RM, Sterne JAC, Timpson N, Davey Smith G. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Stat Med* (2008) 27(8):1133–63. doi: 10.1002/sim.3034

61. Varbo A, Benn M, Smith GD, Timpson NJ, Tybjaerg-Hansen A, Nordestgaard BG. Remnant cholesterol, low-density lipoprotein cholesterol, and blood pressure as mediators from obesity to ischemic heart disease. *Circ Res* (2014).

62. Do R, Willer CJ, Schmidt EM, Sengupta S, Gao C, Peloso GM, et al. Common variants associated with plasma triglycerides and risk for coronary artery disease. *Nat Genet* (2013) 45(11):1345–52. doi: 10.1038/ng.2795

63. Clarke R, Peden JF, Hopewell JC, Kyriakou T, Goel A, Heath SC, et al. Genetic variants associated with Lp(a) lipoprotein level and coronary disease. *N Engl J Med Overseas Ed* (2009) 361(26):2518–28. doi: 10.1056/NEJMoa0902604

64. Voight BF, Peloso GM, Orho-Melander M, Frikke-Schmidt R, Barbalic M, Jensen MK, et al. Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. *Lancet* (2012) 380(9841):572–80. doi: 10.1016/S0140-6736(12)60312-2

65. Holmes MV, Asselbergs FW, Palmer TM, Drenos F, Lanktree MB, Nelson CP, et al. Mendelian randomization of blood lipids for coronary heart disease. *Eur Heart J* (2015) 36(9):539–50. doi: 10.1093/eurheartj/eht571

66. Helgadottir A, Gretarsdottir S, Thorleifsson G, Hjartarson E, Sigurdsson A, Magnusdottir A, et al. Variants with large effects on blood lipids and the role of cholesterol and triglycerides in coronary disease. *Nat Genet* (2016) 48(6):634–9. doi: 10.1038/ng.3561

67. Kingwell BA, Chapman MJ, Kontush A, Miller NE. HDL-targeted therapies: progress, failures and future. *Nat Rev Drug Discov* (2014) 13(6):445–64. doi: 10.1038/nrd4279

68. Tall AR, Rader DJ. Trials and Tribulations of CETP Inhibitors. *Circ Res* (2018) 122(1):106–12. doi: 10.1161/CIRCRESAHA.117.311978

69. Rosenson RS, Koenig W. Mendelian randomization analyses for selection of therapeutic targets for cardiovascular disease prevention: a note of

circumspection. *Cardiovasc Drugs Ther* (2016) 30(1):65–74. doi: 10.1007/s10557-016-6642-9

70. Kleber ME, Delgado G, Grammer TB, Silbernagel G, Huang J, Kramer BK, et al. Uric acid and cardiovascular events: a mendelian randomization study. *Clin J Am Soc Nephrol* (2015) 26(11):2831–8. doi: 10.1681/ASN.2014070660

71. Gill D, del Greco MF, Walker AP, Srai SKS, Laffan MA, Minelli C. The effect of iron status on risk of coronary artery disease: a mendelian randomization study-brief report. *Arterioscler Thromb Vasc Biol* (2017) 37(9):1788–92.

72. Emdin CA, Khera AV, Natarajan P, Klarin D, Zekavat SM, Hsiao AJ, et al. Genetic association of waist-to-hip ratio with cardiometabolic traits, type 2 diabetes, and coronary heart disease. *JAMA* (2017) 317(6):626–34. doi: 10.1001/jama.2016.21042

73. So HC, Chau CK, Chiu WT, Ho KS, Lo CP, Yim SH, et al. Analysis of genome-wide association data highlights candidates for drug repositioning in psychiatry. *Nat Neurosci* (2017) 20(10):1342–9. doi: 10.1038/nn.4618

74. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, et al. The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* (2006) 313(5795):1929–35. doi: 10.1126/science.1132939

75. Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, et al. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* (2017) 171(6):e17):1437–52. doi: 10.1016/j.cell.2017.10.049

76. Breen G, Li Q, Roth BL, O'Donnell P, Didriksen M, Dolmetsch R, et al. Translating genome-wide association findings into new therapeutics for psychiatry. *Nat Neurosci* (2016) 19(11):1392–6. doi: 10.1038/nn.4411

77. Pritchard J-LE, O'Mara TA, Glubb DM. Enhancing the promise of drug repositioning through genetics. *Front Pharmacol* (2017) 8:896. doi: 10.3389/fphar.2017.00896

78. Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ, et al. A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet* (2015) 47(9):1091–8. doi: 10.1038/ng.3367

79. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BWJH, et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet* (2016) 48(3):245–52. doi: 10.1038/ng.3506

80. Barbeira A, Dickinson S, Torres J, Bonazzola R, Zheng J, Torstenson E. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics (2017). Available from: https://www.biorxiv.org/content/early/2017/10/03/045260.

81. Shu L, Zhao Y, Kurt Z, Byars SG, Tukiainen T, Kettunen J, et al. Mergeomics: multidimensional data integration to identify pathogenic perturbations to biological systems. *BMC Genomics* (2016) 17(1):874. doi: 10.1186/s12864-016-3198-9

82. Talukdar HA, ForoughiAsl H, Jain RK, Ermel R, Ruusalepp A, Franzén O, et al. Cross-Tissue Regulatory Gene Networks in Coronary Artery Disease. *Cell Syst* (2016) 2(3):196–208. doi: 10.1016/j.cels.2016.02.002

83. Lempiäinen H, Brænne I, Michoel T, Tragante V, Vilne B, Webb TR, et al. Network analysis of coronary artery disease risk genes elucidates disease mechanisms and druggable targets. *Sci Rep* (2018) 8(1):3434. doi: 10.1038/s41598-018-20721-6

84. Ghosh S, Vivar J, Nelson CP, Willenborg C, Segrè AV, Mäkinen V-P, et al. Systems genetics analysis of genome-wide association study reveals novel associations between key biological processes and coronary artery disease. *Arterioscler Thromb Vasc Biol* (2015) 35(7):1712–22. doi: 10.1161/ATVBAHA.115.305513

85. Bartha I, di Iulio J, Venter JC, Telenti A. Human gene essentiality. *Nat Rev Genet* (2018) 19(1):51–62. doi: 10.1038/nrg.2017.75

86. Zhang B, Zhu J, editors. "Identification of key causal regulators in gene networks," in *Proceedings of the World Congress on Engineering*. (2013).

87. Franzen O, Ermel R, Cohain A, Akers NK, di Narzo A, Talukdar HA, et al. Cardiometabolic risk loci share downstream cis- and trans-gene regulation across tissues and diseases. *Science* (2016) 353(6301):827–30. doi: 10.1126/science.aad6970

88. Heath JR, Ribas A, Mischel PS. Single-cell analysis tools for drug discovery and development. *Nat Rev Drug Discov* (2016) 15(3):204–16. doi: 10.1038/nrd.2015.16

89. Regev A, Teichmann SA, Lander ES, Amit I, Benoist C, Birney E, et al. The Human Cell Atlas. *eLife* (2017) 6. doi: 10.7554/eLife.27041

90. Akbarian S, Liu C, Knowles JA, Vaccarino FM, Farnham PJ, Crawford GE, et al. The PsychENCODE project. *Nat Neurosci* (2015) 18(12):1707–12. doi: 10.1038/nn.4156

91. BrainSeq AHBGC. BrainSeq: neurogenomics to drive novel target discovery for neuropsychiatric disorders. *Neuron* (2015) 88(6):1078–83. doi: 10.1016/j.neuron.2015.10.047

92. Pers TH, Karjalainen JM, Chan Y, Westra H-J, Wood AR, Yang J, et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun* (2015) 6(1):5890. doi: 10.1038/ncomms6890

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.

Genetics of Cardiovascular Disease: Fishing for Causality

Christoph Paone[1], Federica Diofano[1], Deung-Dae Park[1], Wolfgang Rottbauer[2] and Steffen Just[1]*

[1] Molecular Cardiology, Department of Internal Medicine II, University of Ulm, Ulm, Germany, [2] Department of Internal Medicine II, University of Ulm, Ulm, Germany

Cardiovascular disease (CVD) is still the leading cause of death in all western world countries and genetic predisposition in combination with traditional risk factors frequently mediates their manifestation. Genome-wide association (GWA) studies revealed numerous potentially disease modifying genetic loci often including several SNPs and associated genes. However, pure genetic association does not prove direct or indirect relevance of the modifier region on pathogenesis, nor does it define within the associated region the exact genetic driver of the disease. Therefore, the relevance of the identified genetic disease associations needs to be confirmed either in monogenic traits or in experimental *in vivo* model system by functional genomic studies. In this review, we focus on the use of functional genomic approaches such as gene knock-down or CRISPR/Cas9-mediated genome editing in the zebrafish model to validate disease-associated genomic loci and to identify novel cardiovascular disease genes. We summarize the benefits of the zebrafish for cardiovascular research and highlight examples demonstrating the successful combination of GWA studies and functional genomics in zebrafish to broaden our knowledge on the genetic and molecular underpinnings of cardiovascular diseases.

Keywords: genome-wide association study, zebrafish, functional genomics, CRISPR/Cas9, heart disease

INTRODUCTION

Cardiovascular disease (CVD) is the leading cause of mortality worldwide. CVD describes a class of diseases affecting the heart and blood vessels, such as cardiomyopathies, coronary artery disease, heart failure or arrhythmias. A variety of risk factors, such as smoking, obesity, hypertension or high cholesterol can be causative for CVD, however, it is understood that these traditional risk factors only contribute to a fraction of disease cases (1). Therefore, researchers also focus on the definition of the genetic basis of CVD to identify disease mechanisms independent of environmental risk factors. Recent advances in next-generation sequencing (NGS) techniques enable now an unbiased, whole-genome analysis of patients to identify disease-associated genetic alterations. One of these approaches comprises genome-wide association (GWA) studies (GWAS) that have emerged as a powerful tool to identify disease-related loci and have become a valuable candidate resource for disease causing genes and variants. A GWA study is a hypothesis-free approach utilizing the information of hundreds of thousands of genetic variants across the genome, so-called SNPs (single nucleotide polymorphisms), in large population samples. In this context, GWAS findings are purely genetic, but significant associations between SNPs and the disease are therefore excellent starting-points for detailed follow-up studies. More than 10,000 of such significant associations with

disorders and genomic traits were reported by GWA studies resulting in new insights into biology and molecular mechanisms of various diseases (2). Online platforms like the GWAS catalog provide researchers collected data of published GWA studies and enable the open-access view into these genome-wide analyses (3). The GWAS catalog comprises studies on a variety of diseases ranging from neurological disorders, various cancer types to cardiac diseases, such as cardiomyopathies or arrhythmias. All GWA studies rely on the exact definition of the disease phenotype in patients to obtain an as specific cohort as possible. The influence of a mixed cohort, secondary disease mechanisms or environmental variations might lead to non-significant or underestimated results. This could be particularly observed for GWA studies focusing on heart failure mechanisms (4). Although well-designed, some GWA studies still lack the clinical relevance due to missing causality of the candidate genes. In order to get a fast and reliable validation of GWAS hits, an adequate experimental model in follow-up studies is fundamental. Several model systems are available, ranging from cell culture to animal models and each model has its pros and cons depending on the respective disease mechanism. During the last decades, the zebrafish (*Danio rerio*) has emerged rapidly as a model organism in cardiovascular research. In this review, we will focus on the use of the zebrafish to investigate the pathomechanisms of heart diseases and discuss its suitability as an experimental tool to validate the disease-association of genes identified by GWA studies.

THE ZEBRAFISH: SMALL FISH, BIG IMPACT

Zebrafish possess a variety of features that are advantageous for the use as experimental model organism. Due to their small size (2–4 cm), zebrafish are easy to handle and one female can produce around 200 eggs per week. Zebrafish embryos develop externally and very rapidly to freely swimming and fed larvae within 5 days (5, 6). The zebrafish is an excellent system for microscopic applications as embryos are transparent and numerous transgenic fluorescent reporter lines are available or can easily be produced (7). Such reporter lines are widely used to image organ development and morphology as well as physiological parameters like membrane voltage or calcium transients (8, 9). Because of their suitability for imaging applications, zebrafish are also highly interesting for high-throughput small compound screens. This is enabled by already existing and continuously improving screening platforms e.g., for the automated detection of heartbeat, heartrate and fractional shortening in embryos or isolated hearts of adult zebrafish (10–13). Such set-ups facilitate rapid and high-throughput preclinical tests of large numbers of small molecules and help to identify novel therapeutic strategies (14–17).

Beside the mentioned general advantages, zebrafish exhibit characteristics making them appropriate to study heart development and disease (16, 18–21). Zebrafish heart development proceeds fast and results in a differentiated two-chambered heart within 48 hpf (22). In addition, zebrafish

embryos, in contrast to mammalian or avian embryos, are able to cover their oxygen demand by diffusion during the first days of development and are not dependent on blood circulation. This enables the investigation of gene knockouts or knockdowns, even if they lead to severe defects of the cardiovascular system (23).

On a genetic basis, humans and zebrafish share a 70% sequence similarity and 84% of human disease-causing genes can also be found in the zebrafish genome (24, 25). However, regarding the cardiovascular system, there are basic morphological differences as the zebrafish heart consists of only two heart chambers, one atrium and one ventricle. This is on the one hand advantageous as it displays a simplified experimental model, on the other hand, these anatomical difference may limit the translation of findings into the mammalian system (21). Unlike mammals, which develop a coronary system during embryogenesis, zebrafish show a vasculature on the heart surface starting at 1–2 month post hatching (26). This restricts the study of coronary artery disease (CAD) to adult zebrafish, although it is possible to analyze basic mechanisms of atherosclerotic lesion development also in the vasculature of zebrafish embryos (27). There are also several parameters of the zebrafish heart that are closer to the human situation than mammalian model organisms, such as the mouse (17). For example, the zebrafish heart rate of 120–180 bpm (beats per min) is comparable to the 60–100 bpm of the human heart, whereas the mouse heart beats 5 times faster. Furthermore, zebrafish ECG parameters are very similar to human values enabling a direct comparison and translation of experimental findings (20, 28).

In addition to the great benefits of the zebrafish in regard to organ development, physiology, handling and imaging, its suitability for genetic manipulation is another big advantage of the system (**Figure 1**). Here, we will give a compressed overview on the repertoire of zebrafish genetic tools and highlight examples, where they have been used to demonstrate the causality of genes or loci identified by GWAS.

ZEBRAFISH GENETIC SCREENS: A (SWIMMING) POOL OF DISEASE GENES

Before GWAS data became more and more accessible, candidate driven approaches have been very successful in identifying disease-associated mutations. Here, known molecular players and/or regulators of a specific disease-related pathway are screened in a cohort or hereditary trait to find an association with the pathological outcome. Forward genetic screens in zebrafish contributed a lot to these studies as a variety of genes responsible for cardiovascular defects were identified by zebrafish mutant lines arising from mutagenesis screens (18, 29). These screens, comparable to GWAS, have the advantage to be hypothesis-free approaches that identify genetic mutations via a randomly induced phenotype. The most prominent mutagenesis screens are based on alkylating agents like N-ethyl-N-nitrosurea (ENU), which give rise to point mutations leading to nonsense or missense mutations that affect the regulatory and coding region of genes (30–32).

FIGURE 1 | GWAS and zebrafish: a powerful combination for cardiovascular research. The benefits of the zebrafish as an experimental system can enormously help to validate the functional relevance of candidate genes identified by GWA studies. In addition, the system enables the *in vivo* analysis of underlying pathomechanisms and is highly suitable for high throughput screening applications. In summary, the combination of GWAS and the zebrafish experimental system has the potential to lead to improved and specific therapeutic approaches.

Although mainly recessive and single inherited mutations can be analyzed in such mutagenesis screens, the combination of zebrafish forward genetics followed by human genome analysis led to the identification of several disease-related genes. One example is the zebrafish mutant *main squeeze* (*msq*), which

harbors a mutation in the gene encoding ILK (Integrin-linked kinase). *Msq* mutants display progressive loss of ventricular contractility leading to heart failure (33). Another ILK mutant line, *lost contact* (*loc*), also displays a cardiomyopathy phenotype (34). After identifying ILK mutations as causative for the *loc*

mutant phenotype, Knöll et al. performed a mutation screen in the ILK gene of human cardiomyopathy patients. This screen revealed an ILK mutation that was associated with the disease and its disease causing effect could be again validated in *loc* mutants. These examples show that zebrafish can serve as (I) a resource for new candidate genes in heart failure through forward genetic screens as well as (II) a model organism to validate potential disease-causing mutations in reverse genetic analyses.

REVERSE GENETIC APPROACHES IN ZEBRAFISH

Reverse genetics can be regarded as targeted investigation of a gene of interest by increasing, reducing or silencing its expression. A diversity of reverse genetic tools can be applied in zebrafish, however, several characteristics of zebrafish genetics have to be kept in mind. Zebrafish underwent a whole-genome duplication event with the consequence that for many genes a partially redundant paralog is present (24, 35). In addition, there often exist several transcripts of the same gene and the knockdown or knockout of several genes might be necessary to model the loss-of-function phenotype of a human ortholog. Another aspect that needs to be considered is the genetic variation between and within zebrafish strains that might have an impact on the phenotype and the conclusion drawn from functional analyses (36).

An important and helpful resource for reverse genetic investigations is the zebrafish mutant project that provides a growing list of fish lines with a defined mutation in a specific gene (25). These mutations are induced by chemical mutagenesis, similar to the one used in forward genetic screens, and identified by high-throughput DNA genotyping, an approach called TILLING (targeting induced local lesions in genomes) (37, 38). If a desired and appropriate mutation is available, this open source platform might give scientists a fast access to a loss-of-function model that can be directly used for functional studies. The reverse genetic tools that can be applied in zebrafish are mainly (A) mRNA overexpression and (B) transgenesis, (C) Morpholino-modified antisense oligonucleotide (MO) mediated knockdown or (D) genome editing techniques such as ZFNs (zinc finger nucleases), TALENs (transcription activator-like effector nucleases) or the CRISPR/Cas9 system (clustered regulatory interspaced short palindromic repeats).

mRNA Overexpression

Injecting synthetic capped mRNA encoding the protein of interest into early embryonic stages is commonly used as a standard method to induce transient overexpression of genes or gene variants (e.g., SNPs/variants identified in GWAS or next-generation sequencing) for gain-of-function or loss-of-function studies. Thus, mRNA overexpression in zebrafish was used for example to analyze mutations in the *NEXN* (Nexilin) gene that were identified in human DCM (dilated cardiomyopathy) patients (39). When overexpressed in zebrafish embryos, these mutant *NEXN* variants induced a severe DCM phenotype showing the suitability of the method for fast and effective

testing of the impact of putative mutations. Even though mRNA overexpression is an effective way to elucidate functions of specific genes, its use is restricted to focus on early organ development and function because of limited stability of the injected mRNA.

Transgenesis

Transgenesis in zebrafish involves the insertion of foreign DNA into the genome and is often used to create reporter lines, in which a fluorescent reporter gene under the control of a specific promoter is used to label a particular tissue, organ or cell type. The most commonly used system to insert a transgene in the zebrafish germline is the *Tol2* system derived from medaka fish. This autonomously active *Tol2* element harbors a gene that encodes for a transposase mediating the transposition of the *Tol2* element into the genome (40). For transgenesis of zebrafish, the sequence or gene of interest needs to be flanked by 150–200 bp ends of the *Tol2* element. Injection of this construct together with i*n vitro* transcribed transposase mRNA leads to the highly efficient generation of transgenic F_1 offspring (41, 42). For zebrafish heart development and function a variety of transgenic lines are present, such as *cmlc2-* (*myosin light chain 7, myl7*) promoter driven reporter lines that specifically label cardiomyocytes of both heart chambers (43, 44). In addition, random insertional transgenesis of EGFP, so called enhancer trap, was shown to result in various reporter lines specifically labeling cardiac structures (45). A powerful combination is the use of transgenic lines in cell transplantation experiments that are widely-used in zebrafish embryos to investigate cell-autonomous mechanisms. With this approach, Sawamiphak and colleagues could, for example, analyze fusion events between cardiomyocytes during heart development that enable exchange of mRNA or proteins between individual cells (46).

Furthermore, stable transgenic expression of gene variants associated with heart diseases can serve as an appropriate *in vivo* model to study the underlying pathology. Huttner and coworkers, for example, showed that transgenic expression of the D1275N mutation of the human cardiac sodium channel (SCN5A), which is associated with cardiac abnormalities in humans, also leads to bradycardia and defects of the cardiac conduction-system in zebrafish (47). Further developments of transgenesis techniques in regard to tissue specificity or inducibility will broaden the possibilities for transgenesis in zebrafish and help to create improved experimental systems for cardiac research (48, 49).

Morpholino-Mediated Knockdown

Morpholinos (MO) are knockdown reagents that are very stable, resistant to nucleases and can be injected into 1-cell stage zebrafish embryos. Thus, they became a standard approach for gene knockdown in zebrafish (50, 51). In a variety of cardiovascular research studies, MO-mediated knockdowns were performed to analyze the disease association of a particular gene and/or to model specific pathological features. For example, knockdown of genes that are associated with DCM progression in humans also results in cardiomyopathy in the zebrafish (39, 52). However, phenotypes induced by MOs may be more severe than

those of the corresponding mutants. This discrepancy can be a result of genetic compensation in the mutant or due to off-target effects of the used MO (53–55). Therefore, proper control of MO specificity, efficiency and toxicity should be performed in all applications (51).

Genome Editing Techniques

Genome editing has evolved as a major strategy to disrupt the coding sequence of genes of interest leading to a loss-of function. During recent years, various CRISPR/Cas9, ZFN and TALEN approaches were developed and applied in zebrafish research to create gene knockouts. The detailed technical aspects are not the focus of this review, but are reviewed elsewhere (56, 57). ZFNs and TALEN approaches were successfully applied in zebrafish cardiovascular research studies (58, 59). For instance, ZFN-mediated knockout of *GATA2* results in severe defects in vascular organization highlighting the importance of this gene for cardiovascular development (60).

The discovery of CRISPR/Cas9 as a genome editing method declared a new era of reverse genetics (61, 62). The CRISPR/Cas9 system is the most efficient genome editing method for reverse genetics in zebrafish and exhibits, due to its simplicity and applicability, many advantages compared to ZFN and TALENs (63). The CRISPR/Cas9 system is a two-component complex composed of the Cas9 endonuclease, which induces DSBs (double-strand break) and a guide RNA (gRNA) recognizing specific DNA sequences (62). CRISPR/Cas9 is remarkably simple and adaptable due to its unique mechanism and therefore, is chosen as a major genome-editing tool among all the technologies present in the zebrafish field (64). Its suitability for reverse genetics in zebrafish, in regard to cardiovascular research, could be shown for example by a knockout of the large transcript *pr130* of the Protein Phosphatase 2 Regulatory subunit Bα (PPP2R3A) (65). Here, two *pr130* knockout lines demonstrated the importance of *pr130* for cardiac development and function and provide a suitable genetic model to study the underlying pathomechanisms. An aspect that needs to be considered in all genome editing approaches is the possible presence of off-target effects. Unbiased whole genome analyses of CRISPR/Cas9 off-target effects are still missing and researchers are most often restricted to the analysis of off-target genes that are predicted by computational approaches (66). By careful design and selection of gRNA sequences and the use of nuclease variants with high specificity the risk for off-target effects can be minimized. Additionally, continuous outcrossing of the mutation and the comparison of at least two independently produced knockout lines help to prevent misinterpretations of a genotype-phenotype connection.

Recently, a variety of improvements and new applications of the CRISPR/Cas9 system evolved contributing to the fast implementation of the method in many zebrafish laboratories. The classical targeted knock-out strategy involves the injection of gRNA and Cas9 (mRNA or protein) into 1-cell stage embryos and the screening for germline mutations in subsequent generations (67). Another strategy uses e.g., a catalytically dead Cas9 protein (dCas) lacking endonuclease activity to generate a DNA recognition complex that can specifically perturb transcriptional

elongation, RNA polymerase binding, or transcription factor binding (68). CRISPR/Cas9 can also be used to generate defined knock-in fish lines with integrated SNPs, stop codons, HA tags, loxP sites or fluorescent proteins (69, 70). The CRISPR/Cas9 toolbox is continuously growing and recent progress is achieved by using this method for tissue-specific blockage of gene function (71, 72) or by combining the strategy with optogenetic tools to have temporal control over Cas9 activity (73). In the context of cardiovascular research, these improvements will help to obtain heart-specific knockouts and to mimic the late onset and slow progression of many cardiomyopathy subtypes.

GWA STUDIES AND FUNCTIONAL GENOMICS IN ZEBRAFISH: A POWERFUL COMBINATION

The zebrafish functional genomics toolbox enables a defined analysis of theoretically any gene of interest *in vivo*. This makes the zebrafish a valuable experimental platform to validate putative disease causing genes that are identified by GWA studies. Indeed, a variety of genome-wide surveys, focusing on heart diseases, already used zebrafish to prove their initial findings. **Table 1** summarizes selected examples of genome-wide studies, for which the resulting candidate genes could be confirmed by zebrafish reverse genetics. An early GWA study in 2008 identified three co-segregating genes (*HBEGF*, *IK*, and *SRA1*) associated with DCM (87). For Heparin-binding EGF-like growth factor (*HBEGF*), the linkage to DCM progression was already known from mouse knockout studies (88). The DCM-association for the cytokine IK and the steroid receptor RNA activator1 SRA1 is a new connection arising from this study. The disease-relevance of these candidate genes could be verified in zebrafish embryos. The MO-mediated knockdown of all three genes, HBEGF, SRA1 and IK resulted in severe pericardial edema, accompanied by reduced fractional shortening (FS) of the ventricular chamber (87). Another study focused on the genetic basis of CAID (Chronic atrial and intestinal dysrhythmia) and found a linkage to the *SGOL1* gene (Shugosin-like 1) (86). The authors could show that *SGOL1* is expressed in the sinoatrial region and atrioventricular valves of the adult zebrafish heart. Consistent with its expression pattern, the knockdown of *SGOL1* in zebrafish embryos resulted in bradycardia confirming the involvement of SGOL1 in heart rhythm control (86). *KCNIP1* (potassium voltage-gated channel interacting protein 1) is another example of a gene that could be linked by whole genome analysis to heart disease, here atrial fibrillation (AF) (82). Interestingly, the reported mutation does not lead to a loss of function, but is suggested to increase KCNIP1 levels. The authors modeled this by the overexpression of *KCNIP1* in zebrafish and could show that increased KCNIP1 levels can result in transient atrial tachycardia and AF during high-rate pacing (82). Norton et al. (79) identified *BAG3* (Bcl-2 associated anthanogene 3) as a DCM-associated gene and could confirm its disease relevance by knocking-down *BAG3* in zebrafish embryos. BAG3-deficient fish showed severe pericardial edemas and a decreased fractional shortening as well as a reduced peak

TABLE 1 | Genome-Wide Studies using the zebrafish model to validate the causality of candidate genes.

Genome-wide study	Disease	Gene(s)	Zebrafish genetic tool	Zebrafish phenotype	References
(74): GWAS in 1,910 patients/3,630 controls	DCM	*HSPB7*	MO knockdown TALEN knockout	Heart looping defects Increased susceptibility to cardiomyopathy	(75, 76)
(77): GWAS in 1,590 patients/577 controls	HF	*HSPB7*			
(78): GWAS in 1,179 patients/1,108 controls	DCM	*BAG3*	MO knockdown	Cardiac and skeletal myopathy	(79, 80)
(79): WES in one multigeneration family	DCM	*BAG3*			
(81): WES in three multigeneration families	DCM	*FLNC*	MO knockdown	Abnormal cardiac function and structure	Same study (80)
(82): WES on three unrelated probands	AF	*KCNIP1*	Overexpression MO knockdown	Overexpression promotes AF	Same study
(83): Combined genetic and proteomic GWAS of 4 cohorts with 17,692 samples	LQTS	*VCL*	MO knockdown Gene-Trap CRISPR/Cas9	Defective cardiac repolarization Myocardial contractile dysfunction	Same study (84, 85)
(86): WES on three cases	CAID	*SGOL1*	MO knockdown	Bradycardia	Same study
(87): GWAS in 590 patients/732 controls	DCM	*HBEGF, IK, SRA1*	MO knockdown	Myocardial contractile dysfunction	Same study

Mentioned here are examples found by manual database analysis. The list might not be exhaustive.

AF, atrial fibrillation; CAID, Chronic atrial and intestinal dysrhythmia; DCM, dilated cardiomyopathy; GWAS, genome-wide association study; HF, heart failure; LQTS, long QT syndrome; MO, morpholino; WES, whole exon sequencing.

blood cell flow velocity (79). A second study independently identified also *BAG3* as a potential DCM-causing gene (78). In addition, the functional requirement of BAG3 for heart as well as skeletal muscle function was also confirmed by an independent MO-based analysis of several myopathy-related genes (80). Another gene linked to DCM that was identified by a whole genome study is Filamin C (*FLNC*) (81). The authors of this study also used MO-knockdown experiments to validate their findings. *FLNC* morphants exhibited dysmorphic or dilated heart chambers as well as impaired heart looping confirming the importance of FLNC for heart function and development (81). Lundby et al. (83) used a GWA approach combined with tissue-specific proteomics to analyze genes associated with LQTS (Long QT Syndrome) (83). They could identify Vinculin (*VCL*) as a disease-associated gene and could confirm its relevance by using a *VCL*-knockdown approach in zebrafish. In these experiments, the authors measured cardiac repolarization in isolated embryonic hearts using fluorescent probes and could observe an impaired repolarization response upon loss of VCL (83). Additionally, by using a gene-trap mutant zebrafish line as well as a CRISPR/Cas9 knockout line of *VCL*b, Cheng et al. could confirm its disease-relevance (85) and a MO-based knockdown of *VCL* in zebrafish in another independent approach also validated the role of Vinculin in heart function and structure (84).

Other studies did not use the zebrafish to validate their candidate genes, however, independent follow-up studies could confirm the disease-causing potential of some of them. For example, two independent GWA studies on DCM and heart failure identified SNPs near the *HSPB7* (Heat Shock Protein Family B Member 7) gene to be associated with the disease

(74, 77). Three years later, Rosenfeld et al. could confirm the requirement of HSPB7 in heart function and structure by MO-based knockdown experiments. HSPB7 depletion led to impaired cardiac morphogenesis due to defects in ventricular size, but also due to an early block of heart tube formation (75). By using a TALEN-mediated knockout of *HSPB7*, the same group recently showed that loss of *HSPB7* increases the susceptibility of adult mutant zebrafish for cardiomyopathy due to impaired protein homeostasis serving as another proof of the initial GWAS findings (76). In addition, this zebrafish study and most of the above mentioned, are not only validating candidate genes from GWA studies, they also allow a more detailed investigation of the underlying pathomechanisms and help to identify novel disease-associated pathways and protein networks.

CONCLUSION AND FUTURE PERSPECTIVES

The zebrafish has a variety of advantages to be combined with GWAS. Zebrafish are easy to keep, to handle and to image, show many physiological and genetic similarities to humans and are highly suitable for genetic manipulations. These features help to establish valid disease models and allow a plethora of follow-up studies.

Due to obvious differences in morphology and living environment, it should be clear that a simple translation of findings from the zebrafish system to humans is not always possible. In addition, the probably biggest difference and, also peculiarity of the zebrafish heart compared to mammals is the ability to regenerate injury (89). This may result in drawbacks

and limitations when comparing pathomechanisms in fish and humans, but also makes the zebrafish a highly interesting model to study the underlying mechanisms of regeneration (90, 91). It is important to mention that a disease-association of a particular gene that cannot be confirmed in zebrafish doesn't necessarily mean that it is not causative for the phenotype. Mechanisms like intrinsic repair processes or genetic compensation may hide a causative effect of a gene mutation. In such situations, other experimental models, like rodents or patient-derived iPSC (induced pluripotent stem cells) might lead to clearer results (92, 93). Nevertheless, all mentioned benefits make the zebrafish a valid and highly suitable model to investigate cardiovascular pathologies and to prove findings from GWAS. Many SNPs identified in GWA studies are located in non-coding regions of the genome and might affect for example enhancer or repressor binding, microRNA binding sites or chromatin structure. Also for these kinds of mutations, the zebrafish system can help to identify their *in vivo* relevance, biological role and the underlying pathomechanisms. Madelaine et al. (94), for example, could very recently confirm human disease-associated SNPs in CNEs (conserved non-exonic elements) by

using CRISPR/Cas9-mediated deletion of the respective non-coding locus (94).

We are sure that the fruitful synergism between GWAS and zebrafish in cardiac research will expand in the future and will lead to the identification of novel disease-causing genes and variants and help to screen for possible therapeutic strategies.

AUTHOR CONTRIBUTIONS

CP, FD, D-DP, WR, and SJ contributed substantially to the conception, drafting, and revision of the manuscript and approved the final version.

FUNDING

REFERENCES

1. Thanassoulis G, and Vasan RS. Genetic cardiovascular risk prediction. *Circulation* (2010) **122**:2323. doi: 10.1161/CIRCULATIONAHA.109.909309

2. Visscher PM, Wray NR, Zhang Q, Sklar P, and McCarthy MI, Brown MA, et al. 10 Years of GWAS discovery: biology, function, and translation. *Am J Hum Genet* (2017) **101**:5–22. doi: 10.1016/j.ajhg.2017.06.005

3. 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). (2017). *Nucleic Acids Res* **45**:D896–901. doi: 10.1093/nar/gkw1133

4. Rau CD, Lusis AJ, Wang Y. Genetics of common forms of heart failure: challenges and potential solutions. *Curr Opin Cardiol.* (2015) **30**:222–7. doi: 10.1097/HCO.0000000000000160

5. Cunliffe VT. Zebrafish: a practical approach. Edited by C. NÜSSLEIN-VOLHARD and R. DAHM. Oxford University Press. 2002. 322 pages. ISBN 0 19 963808 X. Genetical Research (2003) **82**:79. doi: 10.1017/S0016672303216384

6. Link BA, and Megason S G. (2008). *Zebrafish as a Model for Development. Sourcebook of Models for Biomedical Research*. P. M. Conn. Totowa, NJ: Humana Press.

7. Weber M, Huisken J. *In vivo* imaging of cardiac development and function in zebrafish using light sheet microscopy. *Swiss Med Wkly.* (2015). **145**:w14227. doi: 10.4414/smw.2015.14227

8. Russell JT. Imaging calcium signals *in vivo*: a powerful tool in physiology and pharmacology. *Br J Pharmacol* (2011) **163**:1605–25. doi: 10.1111/j.1476-5381.2010.00988.x

9. Hou JH, Kralj JM, Douglass AD, Engert F, Cohen AE. Simultaneous mapping of membrane voltage and calcium in zebrafish heart *in vivo* reveals chamber-specific developmental transitions in ionic currents. *Front Physiol* (2014) **5**:344. doi: 10.3389/fphys.2014.00344

10. Kitambi SS, Nilsson ES, Sekyrova P, Ibarra C, Tekeoh GN, Andäng M, et al. Small molecule screening platform for assessment of cardiovascular toxicity on adult zebrafish heart. *BMC Physiol* (2012) **12**:3. doi: 10.1186/1472-6793-12-3

11. Spomer W, Pfriem A, Alshut R, Just S, Pylatiuk C. High-throughput screening of zebrafish embryos using automated heart detection and imaging. *J Lab Auto* (2012) **17**:435–42. doi: 10.1177/2211068212464223

12. Pylatiuk C, Sanchez D, Mikut R, Alshut R, Reischl M, Hirth S, et al. Automatic zebrafish heartbeat detection and analysis for zebrafish embryos. *Zebrafish* (2014) **11**:379–83. doi: 10.1089/zeb.2014.1002

13. Nasrat S, Marcato D, Hirth S, Reischl M, Pylatiuk C. Semi-automated detection of fractional shortening in zebrafish embryo heart videos. *Curr Direct Biomed Eng.* (2016) **2**:233. doi: 10.1515/cdbme-2016-0052

14. Mathias JR, Saxena MT, Mumm JS. Advances in zebrafish chemical screening technologies. *Future Med Chem* (2012) **4**:1811–22. doi: 10.4155/fmc.12.115

15. Asnani A, Peterson RT. The zebrafish as a tool to identify novel therapies for human cardiovascular disease. *Dis Models Mech* (2014) **7**:763–7. doi: 10.1242/dmm.016170

16. Keßler M, Rottbauer W, and Just S. Recent progress in the use of zebrafish for novel cardiac drug discovery. *Expert Opin Drug Discov* (2015) **10**:1231–41. doi: 10.1517/17460441.2015.1078788

17. Gut P, Reischauer S, Stainier DY, Arnaout RR. Little Fish, Big Data: zebrafish as a model for cardiovascular and metabolic disease. *Physiol Rev* (2017) **97**:889–938. doi: 10.1152/physrev.00038.2016

18. Dahme T, Katus H, A., and Rottbauer W. Fishing for the genetic basis of cardiovascular disease. *Dis Models Mech.* (2009) **2**:18–22. doi: 10.1242/dmm.000687

19. Bakkers J. Zebrafish as a model to study cardiac development and human cardiac disease. *Cardiovasc Res* (2011) **91**:279–88. doi: 10.1093/cvr/cvr098

20. Pott A, Rottbauer W, and Just S. Functional genomics in zebrafish as a tool to identify novel antiarrhythmic targets. *Curr Med Chem.* (2014) **21**:1320–29. doi: 10.2174/0929867321666131227130218

21. Brown RD, Samsa AL, Qian L, Liu J. Advances in the study of heart development and disease using zebrafish. *J Cardiovasc Dev Dis.* (2016) **3**:2. doi: 10.3390/jcdd3020013

22. Kimmel CB, Ballard WW, Kimmel SR, Ullmann B, Schilling T F. Stages of embryonic development of the zebrafish. *Dev Dyn.* (1995) **203**:253–310. doi: 10.1002/aja.1002030302

23. Stainier DYR. Zebrafish genetics and vertebrate heart formation. *Nat Rev Genet.* (2001) **2**:39. doi: 10.1038/35047564

24. Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M, et al. The zebrafish reference genome sequence and its relationship to the human genome. *Nature* (2013) **496**:498–503. doi: 10.1038/nature12111

25. Kettleborough RNW, Busch-Nentwich EM, Harvey SA, Dooley CM, de Bruijn E, van Eeden F, et al. (2013). A systematic genome-wide analysis of zebrafish protein-coding gene function. *Nature* **496**:494–7. doi: 10.1038/nature11992

26. Harrison MRM, Bussmann J, Huang Y, Zhao L, Osorio A, Burns CG, et al. Chemokine guided angiogenesis directs coronary vasculature formation in zebrafish. *Dev Cell* (2015) 33:442–54. doi: 10.1016/j.devcel.2015.04.001

27. Fang L, Liu C, Miller Y I. Zebrafish models of dyslipidemia: relevance to atherosclerosis and angiogenesis. *Transl Res.* (2014) 163:99–108. doi: 10.1016/j.trsl.2013.09.004

28. Wang LW, Huttner IG, Santiago CF, Kesteven SH, Yu ZY, Feneley M, P., et al. Standardized echocardiographic assessment of cardiac function in normal adult zebrafish and heart disease models. *Dis Model Mech.* (2017) 10:63–76. doi: 10.1242/dmm.026989

29. Stainier DY, Fouquet B, Chen JN, Warren KS, Weinstein BM, Meiler SE, et al. Mutations affecting the formation and function of the cardiovascular system in the zebrafish embryo. *Development* (1996) 123:285.

30. Driever W, Solnica-Krezel L, Schier AF, Neuhauss SC, Malicki J, Stemple DL, et al. A genetic screen for mutations affecting embryogenesis in zebrafish. *Development* (1996) 123:37.

31. Haffter P, Granato M, Brand M, Mullins MC, Hammerschmidt M, Kane DA., et al. The identification of genes with unique and essential functions in the development of the zebrafish, *Danio rerio*. *Development* (1996) 123:1.

32. Patton EE, Zon LI. The art and design of genetic screens: zebrafish. *Nat Rev Genet.* (2001) 2:956–66. doi: 10.1038/35103567

33. Bendig G, Grimmler M, Huttner IG, Wessels G, Dahme T, Just S, et al. Integrin-linked kinase, a novel component of the cardiac mechanical stretch sensor, controls contractility in the zebrafish heart. *Genes Dev.* (2006) 20:2361–72. doi: 10.1101/gad.1448306

34. Knöll R, Postel R, Wang J, Krätzner R, Hennecke G, Vacaru AM, et al. Laminin-α4 and integrin-linked kinase mutations cause human cardiomyopathy via simultaneous defects in cardiomyocytes and endothelial cells. *Circulation* (2007) 116:515. doi: 10.1161/CIRCULATIONAHA.107.689984

35. Meyer A, Schartl M. Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. *Curr Opin Cell Biol.* (1999) 11:699–704. doi: 10.1016/S0955-0674(99)00039-3

36. Brown KH, Dobrinski KP, Lee AS, Gokcumen O, Mills RE, Shi X, et al. Extensive genetic diversity and substructuring among zebrafish strains revealed through copy number variant analysis. *Proc Natl Acad Sci USA.* (2012) 109:529–54. doi: 10.1073/pnas.1112163109

37. Wienholds E, van Eeden F, Kosters M, Mudde J, Plasterk RHA, Cuppen E. Efficient target-selected mutagenesis in zebrafish. *Genome Res.* (2003) 13:2700–07. doi: 10.1101/gr.1725103

38. Stemple DL. TILLING — a high-throughput harvest for functional genomics. *Nature Rev Genet.* (2004) 5:145. doi: 10.1038/nrg1273

39. Hassel D, Dahme T, Erdmann J, Meder B, Huge A, Stoll M, et al. Nexilin mutations destabilize cardiac Z-disks and lead to dilated cardiomyopathy. *Nat Med.* (2009) 15:1281–8. doi: 10.1038/nm.2037

40. Kawakami K, Koga A, Hori H, Shima A. Excision of the Tol2 transposable element of the medaka fish, *Oryzias latipes*, in zebrafish, *Danio rerio. Gene* (1998) 225:17–22. doi: 10.1016/S0378-1119(98)00537-X

41. Kawakami K, Shima A, Kawakami N. Identification of a functional transposase of the Tol2 element, an Ac-like element from the Japanese medaka fish, and its transposition in the zebrafish germ lineage. *Proc Natl Acad Sci USA* (2000) 97:11403–8. doi: 10.1073/pnas.97.21.11403

42. Suster ML, Kikuta H, Urasaki A, Asakawa K, Kawakami K. (2009). *Transgenesis in Zebrafish with the Tol2 Transposon System. Transgenesis Techniques: Principles and Protocols*. E. J. Cartwright. Totowa, NJ: Humana Press

43. Rottbauer W, Saurin AJ, Lickert H, Shen X, Burns CG, Wo ZG, et al. Reptin and pontin antagonistically regulate heart growth in zebrafish embryos. *Cell* (2002) 111:661–672. doi: 10.1016/S0092-8674(02)01112-1

44. Rottbauer W, Wessels G, Dahme T, Just S, Trano N, Hassel D, et al. Cardiac myosin light chain-2. *Circ Res.* (2006) 99:323. doi: 10.1161/01.RES.0000234807.16034.fe

45. Poon KL, Liebling M, Kondrychyn I, Garcia-Lecea M, and Korzh V. Zebrafish cardiac enhancer trap lines: new tools for *in vivo* studies of cardiovascular development and disease. *Dev Dyn.* (2010) 239:914–26. doi: 10.1002/dvdy.22203

46. Sawamiphak S, Kontarakis Z, Filosa A, Reischauer S, Stainier DYR. Transient cardiomyocyte fusion regulates cardiac development in zebrafish. *Nat Commun.* (2017) 8:1525. doi: 10.1038/s41467-017-01555-8

47. Huttner IG, Trivedi G, Jacoby A, Mann SA, Vandenberg JI, Fatkin D. A transgenic zebrafish model of a human cardiac sodium channel mutation exhibits bradycardia, conduction-system abnormalities and early death. *J Mol Cell Cardiol.* (2013) 61:123–32. doi: 10.1016/j.yjmcc.2013.06.005

48. Knopf F, Schnabel K, Haase C, Pfeifer K, Anastassiadis K, Weidinger G. Dually inducible TetON systems for tissue-specific conditional gene expression in zebrafish. *Proc Natl Acad Sci USA.* (2010) 107:19933–8. doi: 10.1073/pnas.1007799107

49. Akerberg AA, Stewart S, and Stankunas K. Spatial and temporal control of transgene expression in zebrafish. *PLoS ONE* (2014) 9:e92217. doi: 10.1371/journal.pone.0092217

50. Nasevicius A, Ekker SC. Effective targeted gene 'knockdown' in zebrafish. *Nat Genet.* (2000) 26:216–20. doi: 10.1038/79951

51. Stainier DYR, Raz E, Lawson ND, Ekker SC, Burdine RD, Eisen JS, et al. Guidelines for morpholino use in zebrafish. *PLoS Genet.* (2017) 13:e1007000. doi: 10.1371/journal.pgen.1007000

52. Vogel B, Meder B, Just S, Laufer C, Berger I, Weber S, et al. *In-vivo* characterization of human dilated cardiomyopathy genes in zebrafish. *Biochem Biophys Res Commun.* (2009) 390:516–22. doi: 10.1016/j.bbrc.2009.09.129

53. Kok FO, Shin M, Ni CW, Gupta A, Grosse AS, van Impel A, et al. Reverse genetic screening reveals poor correlation between morpholino-induced and mutant phenotypes in zebrafish. *Dev Cell* (2015) 32:97–108. doi: 10.1016/j.devcel.2014.11.018

54. Novodvorsky P, Watson O, Gray C, Wilkinson RN, Reeve S, Smythe C, et al. klf2ash317 mutant zebrafish do not recapitulate morpholino-induced vascular and haematopoietic phenotypes. *PLoS ONE* (2015) 10:e0141611. doi: 10.1371/journal.pone.0141611

55. Rossi A, Kontarakis Z, Gerri C, Nolte H, Hölper S, Krüger M, et al. Genetic compensation induced by deleterious mutations but not gene knockdowns. *Nature* (2015) 524:230–3. doi: 10.1038/nature14580

56. Koester R, Sassen WA. A molecular toolbox for genetic manipulation of zebrafish. *Adv Genom Genet.* (2015) 5:151–163. doi: 10.2147/AGG.S57585

57. Sertori R, Trengove M, Basheer F, Ward AC, Liongue C. Genome editing in zebrafish: a practical overview. *Brief Funct Genomics* (2016) 15:322–30. doi: 10.1093/bfgp/elv051

58. Huang P, Xiao A, Zhou M, Zhu Z, Lin S, and Zhang B. Heritable gene targeting in zebrafish using customized TALENs. *Nat Biotechnol.* (2011) 29:699–700. doi: 10.1038/nbt.1939

59. Zu Y, Tong X, Wang Z, Liu D, Pan R, Li Z, et al. TALEN-mediated precise genome modification by homologous recombination in zebrafish. *Nat Methods* (2013) 10:329–31. doi: 10.1038/nmeth.2374

60. Zhu C, Smith T, McNulty J, Rayla AL, Lakshmanan A, Siekmann AF, et al. Evaluation and application of modularly assembled zinc-finger nucleases in zebrafish. *Development* (2011) 138:4555–64. doi: 10.1242/dev.066779

61. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* (2012) 337:816–21. doi: 10.1126/science.1225829

62. Wright AV, Nunez JK, Doudna JA. Biology and applications of CRISPR systems: harnessing nature's toolbox for genome engineering. *Cell* (2016) 164:29–44. doi: 10.1016/j.cell.2015.12.035

63. Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* (2014) 157:1262–78. doi: 10.1016/j.cell.2014.05.010

64. Li M, Zhao L, Page-McCaw PS, Chen W. Zebrafish genome engineering using the CRISPR-Cas9 system. *Trends Genet.* (2016) 32:815–27. doi: 10.1016/j.tig.2016.10.005

65. Yang J, Li Z, Gan X, Zhai G, Gao J, Xiong C, et al. Deletion of Pr130 interrupts cardiac development in zebrafish. *Int J Mol Sci.* (2016) 17:1746. doi: 10.3390/ijms17111746

66. Zischewski J, Fischer R, and Bortesi L. Detection of on-target and off-target mutations generated by CRISPR/Cas9 and other sequence-specific nucleases. *Biotechnol Adv.* (2017) 35:95–104. doi: 10.1016/j.biotechadv.2016.12.003

67. Gagnon JA, Valen E, Thyme SB, Huang P, Akhmetova L, Pauli A, et al. Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and

large-scale assessment of single-guide RNAs. *PLoS ONE* (2014) **9**:e98186. doi: 10.1371/journal.pone.0098186

68. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* (2013) **152**:1173–83. doi: 10.1016/j.cell.2013.02.022

69. Li J, Zhang BB, Ren YG, Gu SY, Xiang YH, Du JL. Intron targeting-mediated and endogenous gene integrity-maintaining knockin in zebrafish using the CRISPR/Cas9 system. *Cell Res.* (2015) **25**:634–7. doi: 10.1038/cr.2015.43

70. Hoshijima K, Jurynec MJ, Grunwald DJ. Precise editing of the zebrafish genome made simple and efficient. *Dev Cell* (2016) **36**:654–67. doi: 10.1016/j.devcel.2016.02.015

71. Ablain J, Durand EM, Yang S, Zhou Y, Zon LI. A CRISPR/Cas9 vector system for tissue-specific gene disruption in zebrafish. *Dev Cell* (2015) **32**:756–64. doi: 10.1016/j.devcel.2015.01.032

72. Wild R, Klems A, Takamiya M, Hayashi Y, Strähle U, Ando K, et al. Neuronal sFlt1 and Vegfaa determine venous sprouting and spinal cord vascularization. *Nat Commun.* (2017) **8**:13991. doi: 10.1038/ncomms13991

73. Reade A, Motta-Mena LB, Gardner KH, Stainier DY, Weiner OD, and Woo S. TAEL: a zebrafish-optimized optogenetic gene expression system with fine spatial and temporal control. *Development* (2017) **144**:345–55. doi: 10.1242/dev.139238

74. Stark K, Esslinger UB, Reinhard W, Petrov G, Winkler T, Komajda M, et al. Genetic association study identifies HSPB7 as a risk gene for idiopathic dilated cardiomyopathy. *PLoS Genet.* (2010) **6**:e1001167. doi: 10.1371/journal.pgen.1001167

75. Rosenfeld GE, Mercer EJ, Mason CE, Evans T. Small heat shock proteins Hspb7 and Hspb12 regulate early steps of cardiac morphogenesis. *Dev Biol.* (2013) **381**:389–400. doi: 10.1016/j.ydbio.2013.06.025

76. Mercer EJ, Lin YF, Cohen-Gould L, Evans T. Hspb7 is a cardioprotective chaperone facilitating sarcomeric proteostasis. *Dev Biol.* (2018) **435**:41–55. doi: 10.1016/j.ydbio.2018.01.005

77. Cappola TP, Li M, He J, Ky B, Gilmore J, Qu L, et al. Common variants in HSPB7 and FRMD4B associated with advanced heart failure. *Circulation* (2010) **3**:147–54. doi: 10.1161/CIRCGENETICS.109.898395

78. Villard E, Perret C, Gary F, Proust C, Dilanian G, Hengstenberg C, et al. A genome-wide association study identifies two loci associated with heart failure due to dilated cardiomyopathy. *Eur Heart J.* (2011) **32**:1065–76. doi: 10.1093/eurheartj/ehr105

79. Norton N, Li D, Mark Rieder J, Jill Siegfried D, Rampersaud E, Züchner S, et al. Genome-wide studies of copy number variation and exome sequencing identify rare variants in BAG3 as a cause of dilated cardiomyopathy. *Am J Hum Genet.* (2011) **88**:273–82. doi: 10.1016/j.ajhg.2011.01.016

80. Bührdel JB, Hirth S, Keßler M, Westphal S, Forster M, Manta L, et al. *In vivo* characterization of human myofibrillar myopathy genes in zebrafish. *Biochem Biophys Res Commun.* (2015) **461**:217–23. doi: 10.1016/j.bbrc.2015.03.149

81. Begay RL, Tharp CA, Martin A, Graw SL, Sinagra G, Miani D, et al. FLNC gene splice mutations cause dilated cardiomyopathy. *JACC* (2016) **1**:344–59. doi: 10.1016/j.jacbts.2016.05.004

82. Tsai TC, Hsieh CS, Chang SN, Chuang EY, Ueng KC, Tsai CF, et al. Genome-wide screening identifies a KCNIP1 copy number variant as a genetic predictor for atrial fibrillation. *Nat Commun.* (2016) **7**:10190. doi: 10.1038/ncomms10190

83. Lundby A, Rossin EJ, Steffensen AB, Rav Acha M, Newton-Cheh C, Pfeufer A, et al. Annotation of loci from genome-wide association studies using tissue-specific quantitative interaction proteomics. *Nat Methods* (2014) **11**:868–74. doi: 10.1038/nmeth.2997

84. Hirth S, Bühler A, Bührdel JB, Rudeck S, Dahme T, Rottbauer W, et al. Paxillin and focal adhesion kinase (FAK) regulate cardiac contractility in the zebrafish heart. *PLoS ONE* (2016) **11**:e0150323. doi: 10.1371/journal.pone.0150323

85. Cheng F, Miao L, Wu Q, Gong X, Xiong J, and Zhang J. Vinculin b deficiency causes epicardial hyperplasia and coronary vessel disorganization in zebrafish. *Development* (2016) **143**:3522–31. doi: 10.1242/dev.132936

86. Chetaille P, Preuss C, Burkhard S, Côté JM, Houde C, Castilloux J, et al. Mutations in SGOL1 cause a novel cohesinopathy affecting heart and gut rhythm. *Nat Genet.* (2014) **46**:1245–9. doi: 10.1038/ng.3113

87. Friedrichs F, Zugck C, Rauch GJ, Ivandic B, Weichenhan D, Muller-Bardorff M, et al. HBEGF, SRA1, and IK: Three cosegregating genes as determinants of cardiomyopathy. *Genome Res.* (2009) **19**:395–403. doi: 10.1101/gr.076653.108

88. Iwamoto R, Yamazaki S, Asakura M, Takashima S, Hasuwa H, Miyado K, et al. Heparin-binding EGF-like growth factor and ErbB signaling is essential for heart function. *Proc Natl Acad Sci USA.* (2003) **100**:3221–6. doi: 10.1073/pnas.0537588100

89. Poss KD, Wilson LG, Keating MT. Heart regeneration in zebrafish. *Science* (2002) **298**:2188. doi: 10.1126/science.1077857

90. Sehring IM, Jahn C, Weidinger G. Zebrafish fin and heart: what's special about regeneration? *Curr Opin Genet Dev.* (2016) **40**:48–56. doi: 10.1016/j.gde.2016.05.011

91. González-Rosa JM, Burns CE, Burns CG. Zebrafish heart regeneration: 15 years of discoveries. *Regeneration* (2017) **4**:105–23. doi: 10.1002/reg2.83

92. Cox RD, Church C D. Mouse models and the interpretation of human GWAS in type 2 diabetes and obesity. *Dis Model Mech.* (2011) **4**:155–64. doi: 10.1242/dmm.000414

93. Warren CR, O'Sullivan JF, Friesen M, Becker CE, Zhang X, Liu P, et al. Induced pluripotent stem cell differentiation enables functional validation of GWAS variants in metabolic disease. *Cell Stem Cell* (2017) **20**:547–57.e547. doi: 10.1016/j.stem.2017.01.010

94. Madelaine R, Notwell JH, Skariah G, Halluin C, Chen CC, Bejerano G, et al. A screen for deeply conserved non-coding GWAS SNPs uncovers a MIR-9-2 functional mutation associated to retinal vasculature defects in human. *Nucleic Acids Res.* (2018) **46**:3517–31. doi: 10.1093/nar/gky166

Using Gene Expression to Annotate Cardiovascular GWAS Loci

Matthias Heinig [1,2]*

[1] Institute of Computational Biology, Helmholtz Zentrum München German Research Center for Environmental Health, Neuherberg, Germany, [2] Department of Informatics, Technical University of Munich, Munich, Germany

Genetic variants at hundreds of loci associated with cardiovascular phenotypes have been identified by genome wide association studies. Most of these variants are located in intronic or intergenic regions rendering the functional and mechanistic follow up difficult. These non-protein-coding regions harbor regulatory sequences. Thus the study of genetic variants associated with transcription—so called expression quantitative trait loci—has emerged as a promising approach to identify regulatory sequence variants. The genes and pathways they control constitute candidate causal drivers at cardiovascular risk loci. This review provides an overview of the expression quantitative trait loci resources available for cardiovascular genetics research and the most commonly used approaches for candidate gene identification.

Keywords: eQTL, expression quantitative trait loci, genome wide association study, GWAS, cardiovascular disease

BACKGROUND

The ultimate goal of any genetic association analysis is to identify genetic variation linked to variation of a phenotype and to elucidate the molecular mechanisms, which are altered by the sequence variation. Genome wide association studies have been tremendously successful in identifying thousands of disease-associated loci as documented by the steady growth of the continuously updated GWAS catalog (1). This progress has also highlighted hundreds of loci associated with cardiovascular phenotypes: the current GWAS catalog (2) lists 249 distinct chromosomal regions associated with coronary artery disease with candidate genes and pathways at many loci summarized in Klarin et al. (3), 138/115 with diastolic/systolic blood pressure, 109 with QT interval, to name just the top three cardiovascular phenotypes. Follow up analysis of these loci aim to establish the causal mechanisms underlying the statistical associations. In classical family based linkage studies typically identifying rare variants with very large effect sizes, the causal variants are typically located in the protein sequence and have a strong impact on protein function (4), for instance truncating mutations in the sarcomeric protein TTN cause dilated cardiomyopathy (5–8). In GWAS however, the identification of causal variants proved to be very challenging, since the vast majority of these disease-associated variants is located either in introns of genes or in intergenic regions (2). Therefore the classical approach of identifying the variant with strongest impact on protein function, such as gained stop codons is not sufficient.

Recent large-scale efforts have annotated a plethora of functional regulatory elements such as enhancers residing in the non-protein-coding part of the genome (9, 10). Therefore an alternative mechanism might be that disease-associated regulatory variants alter the sequence and function of such regulatory elements. Indeed a systematic analysis of the location of disease-associated variants showed that they preferentially reside in regulatory elements (11, 12). Since regulatory elements are highly tissue specific, this information can even be used to identify the disease-relevant

tissues (11, 12). These results from localization analysis are highly suggestive that disease-associated variants alter regulatory elements. It now remains to be shown that they indeed are altered and to identify the respective target gene whose transcription is controlled by the regulatory element.

Integrated analysis of the genetics of gene expression provides an elegant way of directly assessing the consequences of putative regulatory sequence variants on transcription. In this study design (13), a population cohort is characterized for their genome wide patterns of genetic variation and also for genome wide gene expression. Gene expression levels are treated as quantitative traits and systematically tested for associations between sequence variants and gene expression. Significant associations are called expression quantitative trait loci (eQTL). These eQTL not only identify putative regulatory variants, but also their target genes as the gene whose expression is associated with the variant (14, 15). Biological information processing and regulation is not limited to transcription, so this approach has also been generalized toward other intermediate molecular traits such as DNA methylation (16, 17), open chromatin (18), histone modifications (19–21), gene, exon and transcript expression levels (22–26) translation and protein levels (27) as well as metabolites (28, 29). In particular the information from the epigenome can be used to identify regulatory variants, and to characterize their role in disease (11, 18, 21, 27).

eQTL RESOURCES FOR CARDIOVASCULAR GENETICS

Regulatory elements and also the effects of variants on those elements can be highly tissue specific, therefore it is key to investigate the tissue relevant for the disease (11, 12, 25, 30). Because biopsies of tissues relevant for cardiovascular diseases, in particular of the heart are very difficult to obtain from humans, it is not surprising, that early applications of eQTL analysis to identify candidate genes for cardiovascular phenotypes were reported in animal models (31). To understand the regulatory impact of sequence variants in humans, samples of disease relevant tissues are often obtained during surgery, from organ donors or from post-mortem sections. As a consequence of these practical considerations, the transcriptome data might be confounded by differences in tissue composition (32) or ischemic time of post-mortem samples (25). Therefore additional care has to be taken in data analysis accounting for observed and hidden confounders (33). Current reviews provide an overview of recent human eQTL studies (15, 34). The most comprehensive study to date is the Genotype tissue expression (GTEx) project, which aims to characterize regulatory sequence variants across 44 distinct tissues from post-mortem sections (26). This includes cardiac tissues: left ventricle, atrial appendage; vascular tissues: aorta, tibial artery, coronary artery; as well as metabolic tissues: liver, subcutaneous and viscelar adipose tissue (**Table 1**). In terms of sample size and coverage of tissues of interest, the eQTL data generated in the STARNET consortium is currently the most comprehensive

TABLE 1 | Recent cardiovascular eQTL resources.

References	Tissue	Sample size	Population
(35)	Left Atrial wall	62	European
(32)	Left Ventricle	205	European
(36)	Left Atria	329	European/African American
(37)	Left Ventricle	129	European
(26)	Atrial Appendage	264	European/African American
(26)	Left Ventricle	272	European/African American
(26)	Aorta	267	European/African American
(26)	Tibial artery	388	European/African American
(26)	Coronary artery	152	European/African American
(26)	Adipose—Subcutaneous	385	European/African American
(26)	Adipose—Visceral	313	European/African American
(26)	Liver	153	European/African American
(38)	Mammary artery	600	European
(38)	Atherosclerotic aortic root	600	European
(38)	Visceral abdominal fat	600	European
(38)	Skeletal muscle	600	European
(38)	Liver	600	European

resource (38). It focuses on vascular and metabolic tissues in patients with coronary artery disease. It has been shown that eQTL are sometimes dependent on the disease context (32). This observation is also supported by the finding that more eQTLs associated with disease SNP can be found in diseased populations (38). Formation of atherosclerotic plaques is an inflammatory process, therefore also immune cells such as monocytes or macrophages are considered disease relevant tissues and have been extensively profiled (39). Since the disease relevant tissues are not always known a priori efforts are currently underway to establish cohorts of induced pluripotent stem cell that can potentially be differentiated into any cell type for genetic mapping (40). These eQTL projects are complemented by large scale projects aimed at creating a reference map of regulatory elements across an exhaustive set of 111 human cell types and tissues (10) by annotation with epigenetic markers of regulatory elements and recent developments of sequencing based methods (e.g., Hi-C) to study chromosomal architecture (41) in a wide variety of human tissues (42) including heart, liver and aorta. These techniques can identify promoter—enhancer interactions and have already been used successfully to identify IRX3 as the causal gene underlying an obesity GWAS hit located in the intron of the FTO gene (43).

CANDIDATE IDENTIFICATION STRATEGIES

cis eQTL Candidate Genes

Overlapping eQTL and GWAS SNPs is the most straightforward approach to identify candidate genes for GWAS hits. If a GWAS SNP is also an eQTL for a close by gene or in tight LD with an eQTL, it is conceivable that the SNP indeed affects a regulatory element controlling the expression the respective gene. These genes are typically called *cis*-eQTL when the distance between gene and variant is not further than 500 kb−1 Mb, as opposed to *trans*-eQTL, where the distances are greater or the variant and gene are located on different chromosomes. Cardiovascular candidate genes such as SORT1 (44) and LIPA (45) have been identified as *cis*-eQTL. It has been demonstrated that these candidate genes frequently are not the genes located closest to the GWAS SNP for heart related traits (32) and also more generally for any GWAS trait (25, 26). Nowadays, this candidate annotation approach is becoming a standard analysis included in many GWAS papers and can be performed conveniently using the online software FUMA (46). For instance a recent GWAS on CAD (47) identified eQTL for 196 genes at 97 of the 161 CAD loci found in the analysis from GTEx and other eQTL data bases. This result already demonstrates one caveat of the approach: several candidate genes might emerge for a locus and might be inconsistent between tissues or GWAS variants might also associate with eQTL by chance (26). In this particular example 36 loci have unique candidate genes and additional 24 loci have candidate genes detected consistently across tissues, so 60 loci can be annotated confidently. Overall a highly significant enrichment of trait associated SNPs can be observed among eQTLs as demonstrated for heart related traits (32). Less frequently also trans-eQTL are considered for the annotation of GWAS SNPs, as they do not readily provide a clear mechanistic explanation. Nevertheless, it has been shown in a systematic analysis of GWAS variants, that they frequently also associate with expression levels of genes distant to the GWAS locus (48).
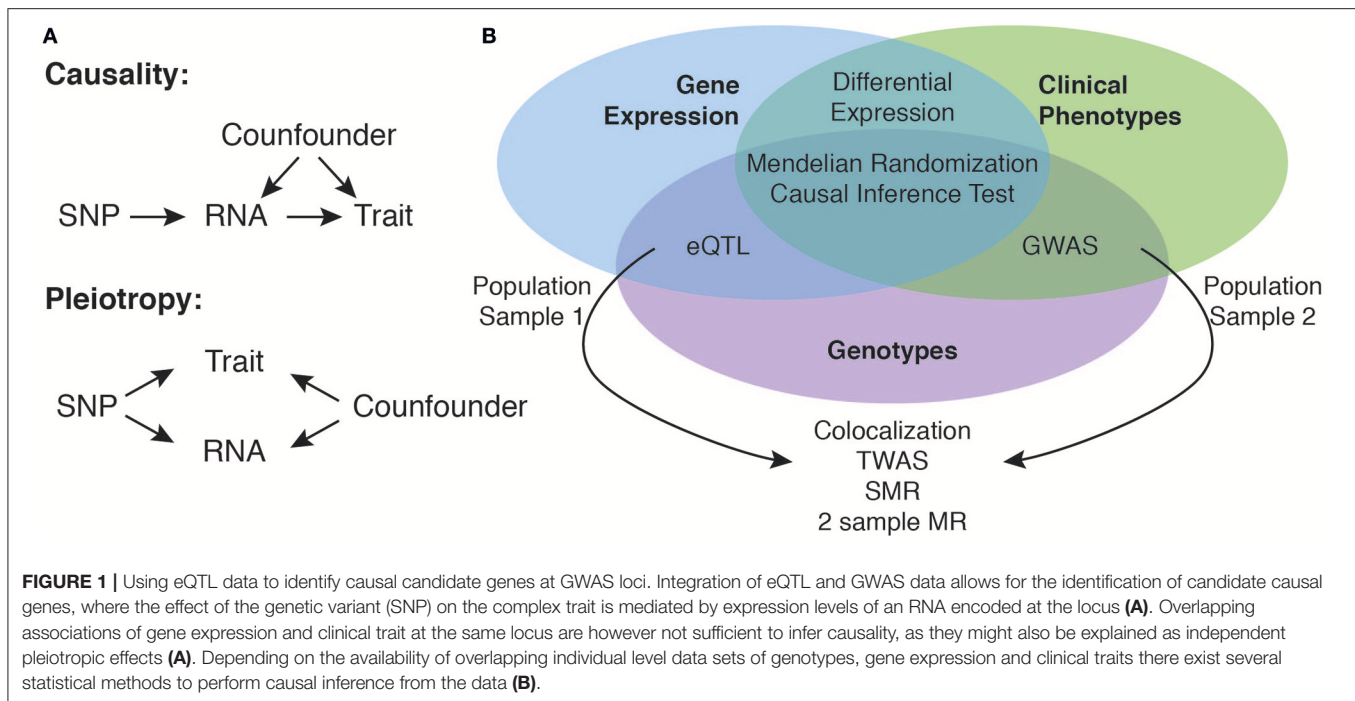
An important limitation of the overlap-based strategy is that it cannot be used to establish causality. Strictly speaking the experimental design does only allow inferring causality in a statistical sense. In genetic associations the direction of causality is always fixed (**Figure 1A**). To establish a causal chain between genetic variation, gene expression and the disease phenotype in the strict sense, an interventional experiment would be required, where all other confounding factors that could determine the phenotype are fixed and only the gene expression level would be manipulated to test an effect on the phenotype. If gene expression is indeed causal for the phenotype, any change of the gene expression necessarily would cause a change in the phenotype. In the concept of Mendelian randomization (MR) one is considering a genetic variant as instrumental variable controlling the levels of gene expression and observes its effect on the phenotypic outcome (49). In analogy to randomized control trials, individuals get assigned to a group based on their genotype. Because the direction of causality between genetic variant and gene expression is fixed and the genetic variant is robustly associated with expression levels, one group will receive a higher dose of gene expression. Assuming that the genotype is independent of confounding factors (**Figure 1A**) changes in phenotypic outcome can be attributed to the changes in gene expression.

Classically, MR and similar approaches to statistically establish causality (50, 51) require to measure all variables in the same population (**Figure 1B**). This is often not feasible, as gene expression profiling in each and every disease cohort is prohibitively expensive. In practice GWAS SNPs and eQTLs are identified in separate populations. Because of data privacy regulations, often a researcher only has access to the full individual level data of one population and the summary statistics of the other population. Depending on which full data set is available there exist several methods allowing to directly integrate the measured data with summary statistics (52–55). A Bayesian co-localization approach based on summary statistics (56) is testing whether the co-localization of two association signals is compatible with a common underlying causal variant and has been successfully applied to blood lipid traits and liver eQTL. An alternative approach is to impute gene expression levels (57) into a GWAS population (54, 58) using eQTL summary statistics from an eQTL reference population. Subsequently the imputed gene expression can be correlated to the disease phenotype to identify candidate genes (54, 58). Alternatively the transcriptome wide association study (TWAS) method (54) and other methods (Barbeira et al. in review) can also work completely without individual level data by indirectly associating expression and phenotype using eQTL and GWAS summary statistics and the LD structure between SNPs. The TWAS approach showed superior power compared to colocalization analysis and simple overlap based analysis in cases where the causal variants are not directly observed, or when multiple causal variants affecting expression and phenotype exist. Consistent with other candidate identification strategies, analysis of obesity related traits with TWAS showed that 66% of identified trait associated genes were not the closest gene (54). Summary data-based Mendelian Randomization (SMR) is a method that can be used if only summary statistics are available from both eQTL and GWAS results. The method makes use of standard two-sample MR (59) to identify causal or pleiotropic effects of sequence variants on gene expression and phenotypes and distinguishes this situation from overlapping independent causal variants in LD using a test on multiple SNPs (55). Similar to results from TWAS analyses, the application of this method to five common diseases showed that only 60% of the identified candidate genes are the closest gene to the GWAS SNP.

Network Based Analysis

Genes are not acting in isolation, but rather form functionally related pathways and networks. Pathways are usually defined based on curated prior knowledge about well-studied processes such as biochemical reactions and signaling pathways (KEGG, Reactome, GO). Pathways can be represented as sets of genes of the same process or as networks preserving the topological

FIGURE 1 | Using eQTL data to identify causal candidate genes at GWAS loci. Integration of eQTL and GWAS data allows for the identification of candidate causal genes, where the effect of the genetic variant (SNP) on the complex trait is mediated by expression levels of an RNA encoded at the locus **(A)**. Overlapping associations of gene expression and clinical trait at the same locus are however not sufficient to infer causality, as they might also be explained as independent pleiotropic effects **(A)**. Depending on the availability of overlapping individual level data sets of genotypes, gene expression and clinical traits there exist several statistical methods to perform causal inference from the data **(B)**.

information which genes are connected to one another, for instance by catalyzing adjacent steps in a metabolic pathway. Alternatively, networks can be derived from high-throughput experiments such as transcriptome profiling (co-expression network) or protein-protein interaction (PPI) screening (PPI network). Pathways and networks defined either from prior knowledge or from data can subsequently be used for the interpretation of disease associations derived from GWAS. Representing pathways as sets of genes, one can ask, whether a set of genes shows higher evidence of association to disease than random gene sets of the same size. Because GWAS test individual SNPs and not genes, a mapping between SNPs and genes is required, for instance based on genomic positions. Methods such as SNP set enrichment analysis (60, 61) can then be used to test the statistical significance of the association between gene sets and the GWAS results by comparing the distribution of GWAS *P*-values of SNPs within the pathway to a background distribution. These methods have been applied to show the association between CAD and pathways for lipid metabolism, coagulation, immunity (62).

Since eQTL experiments require transcriptome profiling in large cohorts, it is natural to use this data to define data driven gene co-expression networks and gene sets, so called co-expression modules. These gene sets are then annotated according to their gene function or cell type specificity and then related to disease via GWAS results using SNP set enrichment analysis. The link between genes and SNPs can naturally be established via cis-eQTLs of the genes of a co-expression module. This approach was also used in the CAD study mentioned above (62). It is important to note that co-expression modules are not necessarily fully overlapping

with biochemical pathways although they might represent the same disease process. For instance the modules might contain transcriptional regulators and parts of a biochemical process that they control.

Network topology of co-expression networks is often used to prioritize candidate genes based on the assumption, that genes with many network connections (so called hubs) are more important (38, 62–65). A study investigating shared molecular networks and their drivers between cardiovascular diseases and type 2 Diabetes applied this strategy (64). Knockout mice for selected key driver genes show indeed metabolic phenotypes and gene expression changes in the network neighborhood of the key drivers. Similarly several studies on CAD identified key driver genes and provided evidence for their functional implication in mouse (65) and *in vitro* studies (62, 65).

CONCLUSIONS

eQTL data provides first leads toward uncovering the mechanisms underlying the statistical associations observed between genetic loci and common cardiovascular diseases. Major challenges for a broad applicability of this approach need to be overcome. First, regulatory elements and therefore also the regulatory impact of sequence variation is highly cell type specific. The GTEx project is addressing this challenge by providing a large scale cross tissue eQTL data base. However, not all conceivable tissues and cell types can be systematically analyzed. In particular transient developmental stages might leave a lasting phenotypic footprint. Induced pluripotent stem cells from cohorts offer an elegant solution (40) as they can potentially be differentiated into any cell type or developmental

stage (Nguyen et al. in review) and studied for eQTLs. A second challenge is posed by variability of the genetic effects on expression between different cells making up a tissue and even between cells of the same cell type. eQTL mapping based on single cell transcriptomic data is becoming feasible (66) and can be used to quantify and map the genetic determinants of cell to cell variability of gene expression. Lastly the grand challenge is to move from correlation or co-localization toward causation. Clearly this is the most difficult task and requires on top of rigorous statistical approaches such as MR also experimental validation.

AUTHOR CONTRIBUTIONS

FUNDING

REFERENCES

1. 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). *Nucleic Acids Res.* (2017) **45**:D896–901. doi: 10.1093/nar/gkw1133

2. Burdett T, Hall PN, Hastings E, Hindorff LA, Junkins H, Klemm A, MacArthur J, et al. *The NHGRI-EBI Catalog of Published Genome-Wide Association Studies.* (2018). Available online at: www.ebi.ac.uk/gwas

3. Klarin D, Zhu QM, Emdin CA, Chaffin M, Horner S, McMillan, B. J, et al. Genetic analysis in UK biobank links insulin resistance and transendothelial migration pathways to coronary artery disease. *Nat Genet.* (2017) **49**:1392–97. doi: 10.1038/ng.3914

4. Timpson NJ, Greenwood CMT, Soranzo N, Lawson DJ, Richards JB, et al. Genetic architecture: the shape of the genetic contribution to human traits and disease. *Nat Rev Genet.* (2018) **19**:110–24. doi: 10.1038/nrg.2017.101

5. Siu BL, Niimura H, Osborne JA, Fatkin D, MacRae C, Solomon S, et al. Familial dilated cardiomyopathy locus maps to chromosome 2q31. *Circulation* (1999) **99**:1022–26. doi: 10.1161/01.CIR.99.8.1022

6. Gerull B, Gramlich M, Atherton J, McNabb M, Trombitás K, Sasse-Klaassen S, et al. Mutations of TTN, encoding the giant muscle filament titin, cause familial dilated cardiomyopathy. *Nat Genet.* (2002) **30**:201–4. doi: 10.1038/ng815

7. Herman DS, Lam L, Taylor MRG, Wang L, Teekakirikul P, Christodoulou D, et al. Truncations of titin causing dilated cardiomyopathy. *New Eng J Med.* (2012) **366**:619–28. doi: 10.1056/NEJMoa1110186

8. Roberts AM, Ware JS, Herman DS, Schafer S, Baksi J, Bick AG, et al. Integrated allelic, transcriptional, and phenomic dissection of the cardiac effects of titin truncations in health and disease. *Sci Trans Med.* (2015) **7**:270ra6–270ra6. doi: 10.1126/scitranslmed.3010134

9. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* (2012) **489**:57–74. doi: 10.1038/nature11247

10. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen, A. et al. Integrative analysis of 111 reference human epigenomes. *Nature* (2015) **518**:317–30. doi: 10.1038/nature14248

11. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory, DNA. *Science* (2012) **337**:1190–5. doi: 10.1126/science.1222794

12. Farh KK, Marson A, Zhu J, Kleinewietfeld M, Housley, W. J, Beik S, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* (2015) **518**:337–43. doi: 10.1038/nature13835

13. Jansen RC, and Nap JP. Genetical genomics: the added value from segregation. *Trends Genet.* (2001) **17**:388–91. doi: 10.1016/S0168-9525(01)02310-1

14. Civelek M, and Lusis AJ. Systems genetics approaches to understand complex traits. *Nat Rev Genet.* (2014) **15**:34–48. doi: 10.1038/nrg3575

15. Albert FW, Kruglyak L. The role of regulatory variation in complex traits and disease. *Nat Rev Genet.* (2015) **16**:197–212. doi: 10.1038/nrg3891

16. Banovich NE, Lan X, McVicker G, van de Geijn B, Degner JF, Blischak JD, et al. Methylation QTLs are associated with coordinated changes in transcription factor binding, histone modifications, and gene expression levels. *PLoS Genet.* (2014) **10**:e1004663. doi: 10.1371/journal.pgen.1004663

17. Lemire M, Zaidi SHE, Ban M, Ge B, Aïssi D, Germain M, et al. Long-range epigenetic regulation is conferred by genetic variation located at thousands of independent loci. *Nat Commun.* (2015) **6**:6326. doi: 10.1038/ncomms7326

18. Degner JF, Pai AA, Pique-Regi R, Veyrieras JB, Gaffney DJ, Pickrell JK, et al. DNase I sensitivity QTLs are a major determinant of human expression variation. *Nature* (2012) **482**:390–94. doi: 10.1038/nature10808

19. Waszak SM, Delaneau O, Gschwind AR, Kilpinen H, Raghav SK, Witwicki RM, et al. Population variation and genetic control of modular chromatin architecture in humans. *Cell* (2015) **162**:1039–50. doi: 10.1016/j.cell.2015.08.001

20. Grubert F, Zaugg JB, Kasowski M, Ursu O, Spacek DV, Martin AR, et al. Genetic control of chromatin states in humans involves local and distal chromosomal interactions. *Cell* (2015) **162**:1051–65. doi: 10.1016/j.cell.2015.07.048

21. del Rosario RC, Cruz-Herrera R, Poschmann J, Rouam SL, Png E, Khor CC, et al. Sensitive detection of chromatin-altering polymorphisms reveals autoimmune disease mechanisms. *Nat Methods* (2015) **12**:458–64. doi: 10.1038/nmeth.3326

22. Montgomery SB, Sammeth M, Arcelus MG, Lach RP, Ingle C, Nisbett J, et al. Transcriptome genetics using second generation sequencing in a caucasian population. *Nature* (2010) **464**:773–77. doi: 10.1038/nature08903

23. Pickrell JK, Marioni JC, Pai AA, Degner JF, Engelhardt BE, Nkadori E, et al. Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* (2010) **464**:768–72. doi: 10.1038/nature08872

24. Lappalainen T, Sammeth M, Friedländer MR, 't Hoen PA, Monlong J, Rivas MA, et al. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* (2013) **501**:506–11. doi: 10.1038/nature 12531

25. GTEx Consortium, Ardlie KG, Wright FA, Dermitzakis ET. Human Genomics. the Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* (2015) **348**:648–60. doi: 10.1126/science.1262110

26. Battle A, Brown CD, Engelhardt BE, Montgomery SB. Genetic effects on gene expression across human tissues. *Nature* (2017) **550**:204–13. doi: 10.1038/nature24277

27. Li YI, van de Geijn B, Raj A, Knowles DA, Petti AA, Golan D, et al. RNA splicing is a primary link between genetic variation and disease. *Science* (2016) **352**:600–4. doi: 10.1126/science.aad9417

28. Suhre K, Shin SY, Petersen AK, Mohney RP, Meredith D, Wägele B, et al. Human metabolic individuality in biomedical and pharmaceutical research. *Nature* (2011) **477**:54–60. doi: 10.1038/nature10354

29. Shin SY, Fauman EB, Petersen AK, Krumsiek J, Santos R, Huang, J, et al. An atlas of genetic influences on human blood metabolites. *Nat Genet.* (2014) **46**:543–50. doi: 10.1038/ng.2982

30. Grundberg E, Small KS, Hedman ÅK, Nica AC, Buil A, Keildson S, et al. Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat Genet.* (2012) **44**:1084–89. doi: 10.1038/ng.2394

31. Monti JJ, Fischer SP, Heinig M, Schulz H. Soluble epoxide hydrolase is a susceptibility factor for heart failure in a rat model of human disease. *Nature* (2008) **40**:529–37. doi: 10.1038/ng.129

32. Heinig M, Adriaens ME, Schafer S, van Deutekom HWM, Lodder EM, Ware JS, et al. Natural genetic variation of the cardiac transcriptome in non-diseased donors and patients with dilated cardiomyopathy. *Genome Biol.* (2017) **18**:170. doi: 10.1186/s13059-017-1286-z

33. Stegle O, Parts L, Durbin R, Winn J. A bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. *PLoS Comput Biol.* (2010) **6**:e1000770. doi: 10.1371/journal.pcbi.1000770

34. Vandiedonck C. Genetic association of molecular traits: a help to identify causative variants in complex diseases. *Clin Genet.* (2018) **93**:520–32. doi: 10.1111/cge.13187

35. Sigurdsson MI, Saddic L, Heydarpour M, Chang T-W, Shekar P, Aranki S, et al. Post-operative atrial fibrillation examined using whole-genome RNA sequencing in human left atrial tissue. *BMC Med Genomics* (2017) **10**:25. doi: 10.1186/s12920-017-0270-5

36. Christophersen IE, Rienstra M, Roselli C, Yin X, Geelhoed B, Barnard J, et al. Large-scale analyses of common and rare variants identify 12 new loci associated with atrial fibrillation. *Nat. Genet.* (2017) **49**:946–52. doi: 10.1038/ng.3843

37. Koopmann TT, Adriaens ME, Moerland PD, Marsman RF, Westerveld ML, Lal S, et al. (2014). Genome-wide identification of expression quantitative trait loci (eQTLs) in human heart. *PLoS ONE* **9**:e97380. doi: 10.1371/journal.pone.0097380

38. Franzén O, Ermel R, Cohain A, Akers NK, Di Narzo A, Talukdar HA, et al. Cardiometabolic risk loci share downstream cis- and trans-gene regulation across tissues and diseases. *Science* (2016) **353**:827–30. doi: 10.1126/science.aad6970

39. Zeller T, Wild P, Szymczak S, Rotival M, Schillert A, Castagne R, et al. Genetics and beyond–the transcriptome of human monocytes and disease susceptibility. *PLoS ONE* (2010) **5**:e10693. doi: 10.1371/journal.pone.0010693

40. Kilpinen H, Goncalves A, Leha A, Afzal V, Alasoo K, Ashford S, et al. Common genetic variation drives molecular heterogeneity in human iPSCs. *Nature* (2017) **546**:370–75. doi: 10.1038/nature22403

41. Davies JO, Oudelaar AM, Higgs DR, Hughes JR. How best to identify chromosomal interactions: a comparison of approaches. *Nat Methods* (2017) **14**:125–34. doi: 10.1038/nmeth.4146

42. Schmitt AD, Hu M, Jung I, Xu Z, Qiu Y, Tan CL, et al. A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Rep.* (2016) **17**:2042–59. doi: 10.1016/j.celrep.2016.10.061

43. Smemo S, Tena JJ, Kim KH, Gamazon ER, Sakabe NJ, Gómez-Marín, C, et al. Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature* (2016) **507**:371–75. doi: 10.1038/nature13138

44. Musunuru K, Strong A, Frank-Kamenetsky M, Lee NE, Ahfeldt T, Sachs KV, et al. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature* (2010) **466**:714–19. doi: 10.1038/nature09266

45. Wild PS, Zeller T, Schillert A, Szymczak S, Sinning CR, Deiseroth A, et al. A genome-wide association study identifies LIPA as a susceptibility gene for coronary artery disease. *Circ Cardiovasc Genet.* (2011) **4**:403–12. doi: 10.1161/CIRCGENETICS.110.958728

46. Watanabe K, Taskesen E, van Bochoven A, Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun.* (2017) **8**:1826. doi: 10.1038/s41467-017-01261-5

47. van der Harst P, Verweij N. Identification of 64 novel genetic loci provides an expanded view on the genetic architecture of coronary artery diseasenovelty and significance. *Circ Res.* (2018) **122**:433–43. doi: 10.1161/CIRCRESAHA.117.312086

48. Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen, J, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet.* (2013) **45**:1238–43. doi: 10.1038/ng.2756

49. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet.* (2014) **23**:R89–98. doi: 10.1093/hmg/ddu328

50. Schadt EE, Lamb J, Yang X, Zhu J, Edwards S, Guhathakurta D, et al. An integrative genomics approach to infer causal associations between gene expression and disease. *Nat Genet.* (2005) **37**:710–17. doi: 10.1038/ng1589

51. Millstein J, Zhang B, Zhu J, Schadt EE. Disentangling molecular relationships with a causal inference test. *BMC Genet.* (2009) **10**:23. doi: 10.1186/1471-2156-10-23

52. Pickrell JK, Berisa T, Liu JZ, Ségurel L, Tung JY, Hinds DA. Detection and interpretation of shared genetic influences on 42 human traits. *Nat Genet.* (2016) **48**:709–17. doi: 10.1038/ng.3570

53. Hormozdiari F, van de Bunt M, Segrè AV, Li X, Joo JWJ, Bilow M, et al. Colocalization of GWAS and eQTL signals detects target genes. *Am J Hum Genet.* (2016) **99**:1245–60. doi: 10.1016/j.ajhg.2016.10.003

54. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx, Brenda W, et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet.* (2016) **48**:245–52. doi: 10.1038/ng.3506

55. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet.* (2016) **48**:481–87. doi: 10.1038/ng.3538

56. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian Test for Colocalisation Between Pairs of Genetic Association Studies Using Summary Statistics. *PLoS Genet.* (2014) **10**:e1004383. doi: 10.1371/journal.pgen.1004383

57. Manor O, Segal E. Robust prediction of expression differences among human individuals using only genotype information. *PLoS Genet.* (2013) **9**:e1003396. doi: 10.1371/journal.pgen.1003396

58. Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ, et al. A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet.* (2015) **47**:1091–98. doi: 10.1038/ng.3367

59. Pierce BL, Burgess S Efficient design for mendelian randomization studies: subsample and 2-sample instrumental variable estimators. *Am J Epidemiol.* (2013) **178**:1177–84. doi: 10.1093/aje/kwt084

60. Zhong H, Beaulaurier J, Lum PY, Molony C, Yang X, Macneil DJ, et al. Liver and adipose expression associated SNPs are enriched for association to type 2 diabetes. *PLoS Genet.* (2010) **6**:e1000932. doi: 10.1371/journal.pgen.1000932

61. Zhong H, Yang X, Kaplan LM, Molony C, Schadt EE. Integrating pathway analysis and genetics of gene expression for genome-wide association studies. *Am J Hum Genet.* (2010) **86**:581–91. doi: 10.1016/j.ajhg.2010.02.020

62. Mäkinen VP, Civelek M, Meng Q, Zhang B, Zhu J, Levian C, et al. Integrative genomics reveals novel molecular pathways and gene networks for coronary artery disease. *PLoS Genet.* (2014) **10**:e1004502. doi: 10.1371/journal.pgen.1004502

63. Wang IM, Zhang B, Yang X, Zhu J, Stepaniants S, Zhang C, et al. Systems analysis of eleven rodent disease models reveals an inflammatome signature and key drivers. *Mol Syst Biol.* (2012) **8**:594. doi: 10.1038/msb.2012.24.

64. Shu L, Chan KHK, Zhang G, Huan T, Kurt Z, Zhao Y, et al. Shared genetic regulatory networks for cardiovascular disease and type 2 diabetes in multiple populations of diverse ethnicities in the United States. *PLoS Genet.* (2017) **13**:e1007040. doi: 10.1371/journal.pgen.1007040

65. Talukdar HA, Foroughi Asl H, Jain RK, Ermel R, Ruusalepp A, Franzén, O, et al. Cross-tissue regulatory gene networks in coronary artery disease. *Cell Syst.* (2016) **2**:196–208. doi: 10.1016/j.cels.2016.02.002

66. Kang HM, Subramaniam M, Targ S, Nguyen M, Maliskova L, McCarthy E, et al. Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat Biotech.* (2018) **36**:89–94. doi: 10.1038/nbt.4042

frontiers
in Cardiovascular Medicine

Scientific Contributions of Population-Based Studies to Cardiovascular Epidemiology in the GWAS Era

Wolfgang Lieb [1]* and Ramachandran S. Vasan [2,3]

[1] Institute of Epidemiology, Kiel University, Kiel, Germany, [2] Framingham Heart Study (FHS), Framingham, MA, United States, [3] Section of Preventive Medicine and Epidemiology, Boston University School of Medicine, Boston, MA, United States

Longitudinal, well phenotyped, population-based cohort studies offer unique research opportunities in the context of genome-wide association studies (GWAS), including GWAS for new-onset (incident) cardiovascular disease (CVD) events, the assessment of gene x lifestyle interactions, and evaluating the incremental predictive utility of genetic information in apparently healthy individuals. Furthermore, comprehensively phenotyped community-dwelling samples have contributed to GWAS of numerous traits that reflect normal organ function (e.g., cardiac structure and systolic and diastolic function) and for many traits along the CVD continuum (e.g., risk factors, circulating biomarkers, and subclinical disease traits). These GWAS have heretofore identified many genetic loci implicated in normal organ function and different stages of the CVD continuum. Finally, population-based cohort studies have made important contributions to Mendelian Randomization analyses, a statistical approach that uses genetic information to assess observed associations between cardiovascular traits and clinical CVD outcomes for potential causality.

Keywords: GWAS (genome-wide association study), population, genetic variation, genetic predisposition to disease, risk prediction

WHAT ARE KEY FEATURES OF POPULATION-BASED COHORT STUDIES?

As a brief introduction, we would like to highlight important design features of population-based studies. As opposed to hospital-based referral samples, population-based epidemiological studies examine community-dwelling or random samples from the general population. As such, study participants are not selected based on a given disease, but rather to represent the general population of the areas sampled, so that observations from such a sample are generalizable to the underlying source population. It has to be kept in mind, though, that the response rate of some landmark cohort studies is rather low [e.g., 5,5% for the UK Biobank (1)], which increases the potential for selection bias (2). Furthermore, most population-based studies are longitudinal studies that are re-examining their participants every few years so that repeated measures of several traits are available and trajectories over time (and their genetic underpinning) can be assessed, as opposed to analyses of single occasion measurements of select traits in typical referral samples. Thus, population-based cohort studies include many individuals free of the disease of interest at the beginning of the study, but who might develop the condition of interest over the course of the study. Therefore, population-based cohort studies are ideal

to study risk factors and intermediate traits for the development of chronic disease conditions and to estimate measures of disease incidence (3, 4).

Third, many population-based cohort studies perform deep physiological/clinical and molecular phenotyping of their study participants (5). For example, comprehensive physiological, biochemical, subclinical, and clinical measurements are obtained on the participants using highly standardized methods. Similarly, clinical endpoints are adjudicated in a comprehensive and highly standardized process, which enhances the accuracy and validity of endpoint data from population-based cohort studies. The molecular characterization may include the assessment of common and rare genetic variation and other OMICs measurements, such as epigenomics, transcriptomics, lipidomics, proteomics, and metabolomics (5). These key features of population-based studies allow specific research questions to be addressed in the context of genome-wide association studies (GWAS). For example, the detailed phenotyping allows comprehensive adjustments and mediation analyses in order to delineate whether an observed association between a genetic variant and cardiovascular outcomes is independent of traditional risk factors and whether traditional risk factors or biomarkers might mediate the observed association. Overall, population-based studies have made a substantial contribution to scientific discoveries in the GWAS era. A few illustrative highlights of such findings from cohort studies are described below.

REFERENCE SAMPLE FOR GENETIC-EPIDEMIOLOGICAL ANALYSES

Since many community-dwelling samples are representative of the general population, population-based studies have served as reference ("control") samples for many genetic case-control analyses. In essence, genetic case-control studies compare allelic frequencies of genetic variants in prevalent cases (patients who have the disease of interest when they are sampled) and controls. Ideally, the control sample captures the distribution of the exposure (in this case, the allele frequencies of putative genetic variants) in the source population from which the cases were derived (6). Therefore, population-based studies have provided controls for genetic case-control studies of a broad spectrum of traits, including myocardial infarction (MI)/coronary artery disease (CAD) (7), stroke (8, 9), and dilated cardiomyopathy (10). Importantly, as detailed below, GWAS might reveal different results depending on whether prevalent or incident cases are being analyzed.

GWAS ANALYSES FOR A BROAD SPECTRUM OF PHENOTYPIC TRAITS AND BIOMARKERS ALONG THE CARDIOVASCULAR DISEASE CONTINUUM

The broad and highly standardized phenotyping of their study participants has allowed many different contributions of

population-based cohort studies to GWAS. Specifically, researchers from population-based studies have performed and contributed to numerous GWAS for traits along the cardiovascular disease continuum, including traditional CVD risk factors [e.g., lipids (11), blood pressure (12, 13), and glycemic traits (14)], circulating cardiovascular biomarkers [e.g., B-type natriuretic peptide (BNP) (15), C-reactive protein (16), troponin (17), aldosterone, renin concentration, renin activity (18), adipokines (19), and fibrinogen levels (20)], and subclinical cardiovascular disease traits [such as indices of left ventricular structure and function (21, 22), carotid intima media thickness [IMT] (23), and coronary artery calcification (24)]. Of note, cardiac function can be assessed by different modalities, including e.g., ECG, echocardiography, MRI/CT and circulating biomarkers; and genome-wide genetic analyses have been conducted for various of these traits, including ECG parameters (25), echocardiographic traits (21, 22) and MRI measures of cardiac structure and function (26), as well as relevant biomarkers (15).

It is important to keep in mind that community-based samples (as opposed to clinical samples with established disease) include many individuals free of CVD at the time of inclusion in the study so that population-based cohort studies offer great opportunities to study the development of cardiovascular disease conditions over the adult life course (27), including very early (clinically asymptomatic) stages of the disease process and the genetic underpinning of these early stages. Thus, the above-mentioned GWAS have described to what extent different stages along the CVD continuum are associated with genetic variation and which genes might be involved.

Furthermore, given the large proportion of apparently healthy individuals in population-based cohort studies (as opposed to clinical samples), these studies conducted GWAS of many traits that reflect relatively normal organ function, including biomarkers of cardiac structure and systolic and diastolic function (21, 22). These studies provided important insights how physiological organ function is influenced by genetic variation, and how organ dysfunction might contribute to different disease processes (21, 22).

ASSESSMENT OF GENE X LIFESTYLE INTERACTIONS

It is an important and growing area of research to quantify the contribution of genes and of different lifestyle factors (and their interactions) to inter-individual variation in cardiovascular risk factor levels and disease risk. Since well phenotyped cohort studies usually have comprehensive genetic data and detailed lifestyle information available, population-based studies represent an ideal setting to study gene x lifestyle interactions. The interaction of a genetic risk score (based on 50 SNPs) and a lifestyle score (including information on smoking, obesity, physical activity, and diet) on the incidence of CAD has been analyzed in several large community-based cohorts (28). Key observations from these analyses were that (i) both scores, the genetic risk score and the lifestyle score, were independently associated with the risk of incident CVD and that (ii) a favorable lifestyle was associated with an almost 50% reduction in the relative risk for CAD, as compared to those with an unfavorable lifestyle

profile (28). This reduction in the relative risk of CAD by a favorable lifestyle was observed in individuals with high genetic risk, but also in individuals with low and intermediate genetic risk (28). Very similar observations were made in more than 270.000 participants of the UK Biobank, when a polygenic risk score, representing 314 BP-associated loci, as well as a slightly different lifestyle score (including information on body mass index, healthy diet, sedentary lifestyle, alcohol consumption, smoking, and urinary sodium excretion levels) were related to different BP traits and to incident CVD (29). Both, the genetic risk score as well as the lifestyle score were associated with BP traits and incident CVD. Importantly, a favorable lifestyle as compared to an unfavorable lifestyle was associated with substantially lower average BP values in all categories of genetic risk (low, intermediate, high) and with an about 30% lower relative risk for incident CVD (29).

The same lifestyle score as in Reference (28) was used in a sample of young women (aged 25 to 40 years) from the Dutch Lifelines cohort to assess the contribution of rare and common genetic variation and of lifestyle factors to very low (≤1st age- and sex-specific percentile) and very high (≥99% age- and sex-specific percentile) levels of LDL-C. The study revealed that about two thirds of the women with very low LDL-C levels had a likely genetic cause (either a relevant mutation in an established gene for monogenic hypocholesterolemia or a very low polygenic risk score), whereas the lifestyle score (28) was not statistically significantly associated with low LDL-C concentrations (30). In cases with hypercholesterolemia, however, an unfavorable lifestyle seems to be more relevant. Only about 40% of the women had a genetic cause (relevant mutations in genes for monogenic familial hypercholesterolemia) or predisposition (high polygenic risk score) for high LDL-C; and of the women without genetic cause for hypercholesterolemia, more than half of women displayed an unfavorable lifestyle profile (30).

Community-based studies have also been involved in studying uncommon loss-of-function variants that may offer insights into function of variants. For example, (gain-of-function) mutations in the PCSK9 (proprotein convertase subtilisin/kexin type 9) serine protease gene were initially identified in families with autosomal dominant hypercholesterolemia (31). Subsequently, loss-of-function mutations were reported in individuals with low circulating low-density lipoprotein (LDL) cholesterol levels (32). Analyses in population-based studies revealed that low-frequency sequence variants in the *PCSK9* gene and a *PCSK9* genetic score were associated with lower circulating LDL cholesterol levels and reduced risk of cardiovascular events in the general population (33, 34). Recently, PCSK9 inhibitors have been tested in randomized controlled trials (35).

THE GENETIC UNDERPINNING OF CHANGE IN CARDIOVASCULAR TRAITS OVER THE LIFE COURSE

Due to the availability of repeated measures over time, cohort studies are also suitable to explore the genetic underpinning of changes in cardiovascular risk factors over time, and of the progression of subclinical CVD traits longitudinally. For example, a GWAS for carotid IMT measured at different time points over a 10-year period

has recently been published (36). Furthermore, several researchers assessed the association of risk factor-associated genetic variants with trajectories of the respective risk factor over the life course. For example, BMI-associated genetic variants have been related to repeated measures of BMI over time (37). Interestingly, BMI in childhood and adulthood were associated with different sets of single nucleotide polymorphisms (SNPs) (37), respectively, consistent with the concept that genetic effects on risk factors might be age-dependent. In line with this concept, genetic linkage analyses for BMI provided evidence for age-dependent effects of select genetic loci (38).

On a parallel note, a genetic risk score consisting of 29 SNPs was not only associated with blood pressure and hypertension prevalence at baseline, but also with new-onset hypertension and change in blood pressure over the life course in a large Swedish cohort study (39).

GWAS FOR INCIDENT DISEASE CONDITIONS

The longitudinal character of population-based cohort studies allows genetic variation to be studied in relation to disease incidence. For example, population-based cohort studies have facilitated GWAS for incident heart failure (40), incident stroke (41) and incident MI/ coronary heart disease (CHD) (3). Interestingly, GWAS for incident MI/CHD (3) reported partially discrepant results as compared to GWAS using prevalent CAD cases (7). As an example, the chromosome 9p21 locus – consistently replicated in case-control GWAS for CAD/ MI (7, 42) – provided only modest evidence for association in a GWAS for incident MI/CHD within the CHARGE consortium (3). Of note, the CHARGE consortium (Cohorts for Heart and Aging Research in Genomic Epidemiology) was founded to coordinate joint GWAS analyses of several traits in large population-based cohort studies and to provide opportunities for mutual replication efforts (43).

It is well known that analyses based on prevalent disease cases and those based on incident cases might reveal different results if the association between an exposure and the disease outcome differs by disease severity or disease duration (a phenomenon referred to as prevalence-incidence bias) (44). In order to be included in a case-control study as prevalent MI/CAD case, MI patients have to survive the acute event until they are sampled. Given that MI is still associated with substantial case fatality (45, 46), case-control studies are likely enriched for MI/CAD survivors with rather long survival (3). Thus, alleles associated with prevalent CAD in case-control analyses could be related to the risk of developing the CAD event, but could also be related to the chances of surviving the acute CAD event. In line with this concept, the CAD risk allele at the 9p21 locus was associated with longer survival after MI in several population-based cohorts within CHARGE (3).

IMPACT OF GENETIC VARIATION ON RISK PREDICTION

Furthermore, community-based prospective cohorts allow assessing whether genetic information improves risk prediction models beyond traditional risk factors. It was, indeed, one of the main motivations

of the human genome project to use genetic information to predict disease risks in healthy individuals and to predict the response to a given therapy among patients. Several analyses conducted in various population-based cohorts assessed whether genetic variation – e.g., in an aggregated form as risk scores – improved performance measures of risk prediction models for a first CVD event, including discrimination, calibration, and reclassification (47–50). Although the results from individual studies vary, in most cases, the genetic risk scores displayed clear statistically significant associations with CVD endpoints, but improvements in discrimination (e.g., C-statistics; integrated discrimination improvement) and reclassification (e.g., net reclassification index) were more modest (47, 48) and some studies did not provide evidence for improvement in these performance metrices beyond traditional risk factors (49, 50).

MENDELIAN RANDOMIZATION ANALYSES FOR CARDIOVASCULAR TRAITS

Genetic information in population-based cohort studies has also been used to assess causality between cardiovascular risk factors or circulating biomarkers and cardiovascular outcomes (incident CVD events) using instrumental variable analyses, a statistical approach referred to as Mendelian Randomization (MR) (51–53). This term, MR, refers to the random assortment of alleles of a given locus at meiosis (51, 52). Thus, if a genetic locus (or a genetic risk score) is strongly associated with circulating biomarker levels or with risk factor levels, individuals are "randomized" to genetically determined high or low biomarker/risk factor levels (51, 52, 54). If the biomarker/risk factor is causally related to CVD, this difference in genetically determined higher or lower biomarker/risk factor levels should translate into corresponding quantitative differences in disease risk (51, 52, 54). Therefore, in addition to the association between the genetic variant and the risk factor/biomarker of interest, MR analyses also assess the associations between the risk factor/biomarker and incident CVD as well as between the genetic variant and incident CVD (52); the two latter analyses are facilitated by population-based cohort studies. By using genetic information as instrumental variable for the biomarker/risk factor of interest, MR analyses try to avoid two important limitations of observational studies, reverse causality and confounding (54, 55). Using MR analyses in population-based

samples, several traits along the CVD continuum and biomarkers have been tested for potentially causal relations with incident CVD, including high-density lipoprotein (HDL) cholesterol (53), C-reactive protein (56), lipoprotein(a) (57), and many others. It has to be kept in mind, though, that instrumental variable analyses can be affected by different types of selection bias. For example, such analyses might be biased, if a genetic variant is related to mortality, and MR analyses are conducted in an elderly sample (58, 59).

CONCLUSION

Population-based studies have substantially improved our understanding of the genetic architecture of normal and abnormal organ function, CVD risk factors, circulating biomarkers, subclinical disease, and overt CVD traits over the life course. Furthermore, they were essential in exploring gene x lifestyle interactions and in evaluating genetic variation in the context of risk prediction models for incident CVD. In addition, population-based cohort studies provided great opportunities to conduct GWAS for incident CVD events, such as MI, stroke and heart failure, and thereby, to overcome classic limitations of case-control GWAS including prevalence-incidence bias. Finally, population-based cohort studies used genetic information as instrumental variables to assess whether cardiovascular risk factors or biomarkers are causally related to clinical CVD (Mendelian Randomization analyses).

AUTHOR CONTRIBUTIONS

WL and RV wrote the article together.

FUNDING

REFERENCES

1. Ganna A, Ingelsson E. 5 year mortality predictors in 498,103 UK Biobank participants: a prospective population-based study. *Lancet* (2015) 386(9993):533–40. doi: 10.1016/S0140-6736(15)60175-1
2. Riedel-Heller SG, Schork A, Matschinger H, Angermeyer MC. Recruitment procedures and their impact on the prevalence of dementia. Results from the leipzig longitudinal study of the aged (LEILA75+). *Neuroepidemiology* (2000) 19(3):130–40. doi: 10.1159/000026248
3. Dehghan A, Bis JC, White CC, Smith AV, Morrison AC, Cupples LA, et al. Genome-wide association study for incident myocardial infarction and coronary heart disease in prospective cohort studies: the CHARGE consortium. *PLoS One* (2016) 11(3):e0144997. doi: 10.1371/journal.pone.0144997
4. Oleckno WA. "Cohort Studies". In: *Epidemiology - Concepts and Methods.* United States: Waveland Press (2008). p. 315–70.

5. Wijmenga C, Zhernakova A. The importance of cohort studies in the post-GWAS era. *Nat Genet* (2018) 50(3):322–8. doi: 10.1038/s41588-018-0066-3
6. Rothman K. "Case-Control Studies". *Epidemiology - An Introduction*. United Kingdom: Oxford University Press (2002). 73 p.
7. Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, Mayer B, et al. Genomewide association analysis of coronary artery disease. *N Engl J Med* (2007) 357(5):443–53. doi: 10.1056/NEJMoa072366
8. Cheng YC, Stanne TM, Giese AK, Ho WK, Traylor M, Amouyel P, et al. Genome-wide association analysis of young-onset stroke identifies a locus on chromosome 10q25 near HABP2. *Stroke* (2016) 47(2):307–16. doi: 10.1161/STROKEAHA.115.011328
9. 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(4):524–. doi: 10.1038/s41588-018-0058-3

10. Meder B, Rühle F, Weis T, Homuth G, Keller A, Franke J, et al. A genome-wide association study identifies 6p21 as novel risk locus for dilated cardiomyopathy. *Eur Heart J* (2014) 35(16):1069–77. doi: 10.1093/eurheartj/eht251

11. 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(7307):707–13. doi: 10.1038/nature09270

12. Newton-Cheh C, Johnson T, Gateva V, Tobin MD, Bochud M, Coin L, et al. Genome-wide association study identifies eight loci associated with blood pressure. *Nat Genet* (2009) 41(6):666–76. doi: 10.1038/ng.361

13. Warren HR, Evangelou E, Cabrera CP, Gao H, Ren M, Mifsud B, et al. Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. *Nat Genet* (2017) 49(3):403–15. doi: 10.1038/ng.3768

14. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet* (2010) 42(2):105–16. doi: 10.1038/ng.520

15. Musani SK, Fox ER, Kraja A, Bidulescu A, Lieb W, Lin H, et al. Genome-wide association analysis of plasma B-type natriuretic peptide in blacks: the Jackson heart study. *Circ Cardiovasc Genet* (2015) 8(1):122–30. doi: 10.1161/CIRCGENETICS.114.000900

16. Dehghan A, Dupuis J, Barbalic M, Bis JC, Eiriksdottir G, Lu C, et al. Meta-analysis of genome-wide association studies in >80 000 subjects identifies multiple loci for C-reactive protein levels. *Circulation* (2011) 123(7):731–8. doi: 10.1161/CIRCULATIONAHA.110.948570

17. Yu B, Barbalic M, Brautbar A, Nambi V, Hoogeveen RC, Tang W, et al. Association of genome-wide variation with highly sensitive cardiac troponin-T levels in European Americans and Blacks: a meta-analysis from atherosclerosis risk in communities and cardiovascular health studies. *Circ Cardiovasc Genet* (2013) 6(1):82–8. doi: 10.1161/CIRCGENETICS.112.963058

18. Lieb W, Chen MH, Teumer A, de Boer RA, Lin H, Fox ER, et al. Genome-wide meta-analyses of plasma renin activity and concentration reveal association with the kininogen 1 and prekallikrein genes. *Circ Cardiovasc Genet* (2015) 8(1):131–40. doi: 10.1161/CIRCGENETICS.114.000613

19. Kilpeläinen TO, Carli JF, Skowronski AA, Sun Q, Kriebel J, Feitosa MF, et al. Genome-wide meta-analysis uncovers novel loci influencing circulating leptin levels. *Nat Commun* (2016) 7:10494. doi: 10.1038/ncomms10494

20. Dehghan A, Yang Q, Peters A, Basu S, Bis JC, Rudnicka AR, et al. Association of novel genetic Loci with circulating fibrinogen levels: a genome-wide association study in 6 population-based cohorts. *Circ Cardiovasc Genet* (2009) 2(2):125–33. doi: 10.1161/CIRCGENETICS.108.825224

21. Vasan RS, Glazer NL, Felix JF, Lieb W, Wild PS, Felix SB, et al. Genetic variants associated with cardiac structure and function: a meta-analysis and replication of genome-wide association data. *JAMA* (2009) 302(2):168–78. doi: 10.1001/jama.2009.978-a

22. Wild PS, Felix JF, Schillert A, Teumer A, Chen MH, Leening MJG, et al. Large-scale genome-wide analysis identifies genetic variants associated with cardiac structure and function. *J Clin Invest* (2017) 127(5):1798–812. doi: 10.1172/JCI84840

23. Bis JC, Kavousi M, Franceschini N, Isaacs A, Abecasis GR, Schminke U, et al. Meta-analysis of genome-wide association studies from the CHARGE consortium identifies common variants associated with carotid intima media thickness and plaque. *Nat Genet* (2011) 43(10):940–7. doi: 10.1038/ng.920

24. O'Donnell CJ, Kavousi M, Smith AV, Kardia SL, Feitosa MF, Hwang SJ, et al. Genome-wide association study for coronary artery calcification with follow-up in myocardial infarction. *Circulation* (2011) 124(25):2855–64. doi: 10.1161/CIRCULATIONAHA.110.974899

25. Eijgelsheim M, Newton-Cheh C, Sotoodehnia N, de Bakker PI, Müller M, Morrison AC, et al. Genome-wide association analysis identifies multiple loci related to resting heart rate. *Hum Mol Genet* (2010) 19(19):3885–94. doi: 10.1093/hmg/ddq303

26. Fox ER, Musani SK, Barbalic M, Lin H, Yu B, Ogunyankin KO, et al. Genome-wide association study of cardiac structure and systolic function in African Americans: the candidate gene association Resource (CARe) study. *Circ Cardiovasc Genet* (2013) 6(1):37–46. doi: 10.1161/CIRCGENETICS.111.962365

27. Vasan RS, Kannel WB. Strategies for cardiovascular risk assessment and prevention over the life course: progress amid imperfections. *Circulation* (2009) 120(5):360–3. doi: 10.1161/CIRCULATIONAHA.109.881995

28. Khera AV, Emdin CA, Drake I, Natarajan P, Bick AG, Cook NR, et al. Genetic risk, adherence to a healthy lifestyle, and coronary disease. *N Engl J Med* (2016) 375(24):2349–58. doi: 10.1056/NEJMoa1605086

29. Pazoki R, Dehghan A, Evangelou E, Warren H, Gao H, Caulfield M, et al. Genetic predisposition to high blood pressure and lifestyle factors: associations with midlife blood pressure levels and cardiovascular events. *Circulation* (2018) 137(7):653–61. doi: 10.1161/CIRCULATIONAHA.117.030898

30. Balder JW, Rimbert A, Zhang X, Viel M, Kanninga R, van Dijk F, et al. Genetics, lifestyle, and low-density lipoprotein cholesterol in young and apparently healthy women. *Circulation* (2018) 137(8):820–31. doi: 10.1161/CIRCULATIONAHA.117.032479

31. Abifadel M, Varret M, Rabès JP, Allard D, Ouguerram K, Devillers M, et al. Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nat Genet* (2003) 34(2):154–6. doi: 10.1038/ng1161

32. Cohen J, Pertsemlidis A, Kotowski IK, Graham R, Garcia CK, Hobbs HH. Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. *Nat Genet* (2005) 37(2):161–5. doi: 10.1038/ng1509

33. Cohen JC, Boerwinkle E, Mosley TH, Hobbs HH. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *N Engl J Med* (2006) 354(12):1264–72. doi: 10.1056/NEJMoa054013

34. Ference BA, Robinson JG, Brook RD, Catapano AL, Chapman MJ, Neff DR, et al. Variation in PCSK9 and HMGCR and risk of cardiovascular disease and diabetes. *N Engl J Med* (2016) 375(22):2144–53. doi: 10.1056/NEJMoa1604304

35. Sabatine MS, Giugliano RP, Keech AC, Honarpour N, Wiviott SD, Murphy SA, et al. Evolocumab and clinical outcomes in patients with cardiovascular disease. *N Engl J Med* (2017) 376(18):1713–22. doi: 10.1056/NEJMoa1615664

36. Xie G, Myint PK, Voora D, Laskowitz DT, Shi P, Ren F, et al. Genome-wide association study on progression of carotid artery intima media thickness over 10 years in a Chinese cohort. *Atherosclerosis* (2015) 243(1):30–7. doi: 10.1016/j.atherosclerosis.2015.08.034

37. Mei H, Chen W, Jiang F, He J, Srinivasan S, Smith EN, et al. Longitudinal replication studies of GWAS risk SNPs influencing body mass index over the course of childhood and adulthood. *PLoS One* (2012) 7(2):e31470. doi: 10.1371/journal.pone.0031470

38. Atwood LD, Heard-Costa NL, Fox CS, Jaquish CE, Cupples LA. Sex and age specific effects of chromosomal regions linked to body mass index in the framingham study. *BMC Genet* (2006) 7:7. doi: 10.1186/1471-2156-7-7

39. Fava C, Sjögren M, Montagnana M, Danese E, Almgren P, Engström G, et al. Prediction of blood pressure changes over time and incidence of hypertension by a genetic risk score in Swedes. *Hypertension* (2013) 61(2):319–26. doi: 10.1161/HYPERTENSIONAHA.112.202655

40. Smith NL, Felix JF, Morrison AC, Demissie S, Glazer NL, Loehr LR, et al. Association of genome-wide variation with the risk of incident heart failure in adults of European and African ancestry: a prospective meta-analysis from the cohorts for heart and aging research in genomic epidemiology (CHARGE) consortium. *Circ Cardiovasc Genet* (2010) 3(3):256–66. doi: 10.1161/CIRCGENETICS.109.895763

41. Ikram MA, Seshadri S, Bis JC, Fornage M, Destefano AL, Aulchenko YS, et al. Genomewide association studies of stroke. *N Engl J Med* (2009) 360(17):1718–28. doi: 10.1056/NEJMoa0900094

42. Schunkert H, Götz A, Braund P, McGinnis R, Tregouet DA, Mangino M, et al. Repeated replication and a prospective meta-analysis of the association between chromosome 9p21.3 and coronary artery disease. *Circulation* (2008) 117(13):1675–84. doi: 10.1161/CIRCULATIONAHA.107.730614

43. Psaty BM, O'Donnell CJ, Gudnason V, Lunetta KL, Folsom AR, Rotter JI, et al. Cohorts for heart and aging research in genomic epidemiology (CHARGE) consortium: design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circ Cardiovasc Genet* (2009) 2(1):73–80. doi: 10.1161/CIRCGENETICS.108.829747

44. Oleckno WA. "Assessing the Accuracy of Epidemiologic Studies". In: Oleckno WA, editor. *Epidemiology: Concepts and Methods*. United States: Waveland Press (2008). p. 195–231.

45. Abildstrom SZ, Rasmussen S, Rosén M, Madsen M. Trends in incidence and case fatality rates of acute myocardial infarction in Denmark and Sweden. *Heart* (2003) 89(5):507–11. doi: 10.1136/heart.89.5.507

46. Lehto HR, Lehto S, Havulinna AS, Ketonen M, Lehtonen A, Kesäniemi YA, et al. Sex differences in short- and long-term case-fatality of myocardial

infarction. *Eur J Epidemiol* (2011) 26(11):851–61. doi: 10.1007/s10654-011-9601-6

47. Ripatti S, Tikkanen E, Orho-Melander M, Havulinna AS, Silander K, Sharma A, et al. A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses. *Lancet* (2010) 376(9750):1393–400. doi: 10.1016/S0140-6736(10)61267-6

48. Hughes MF, Saarela O, Stritzke J, Kee F, Silander K, Klopp N, et al. Genetic markers enhance coronary risk prediction in men: the MORGAM prospective cohorts. *PLoS One* (2012) 7(7):e40922. doi: 10.1371/journal.pone.0040922

49. Paynter NP, Chasman DI, Paré G, Buring JE, Cook NR, Miletich JP, et al. Association between a literature-based genetic risk score and cardiovascular events in women. *JAMA* (2010) 303(7):631–7. doi: 10.1001/jama.2010.119

50. Brautbar A, Pompeii LA, Dehghan A, Ngwa JS, Nambi V, Virani SS, et al. A genetic risk score based on direct associations with coronary heart disease improves coronary heart disease risk prediction in the atherosclerosis risk in communities (ARIC), but not in the Rotterdam and Framingham offspring, studies. *Atherosclerosis* (2012) 223(2):421–6. doi: 10.1016/j.atherosclerosis.2012.05.035

51. Emdin CA, Khera AV, Kathiresan S. Mendelian Randomization. *JAMA* (2017) 318(19):1925–6. doi: 10.1001/jama.2017.17219

52. Swerdlow DI, Kuchenbaecker KB, Shah S, Sofat R, Holmes MV, White J, et al. Selecting instruments for Mendelian randomization in the wake of genome-wide association studies. *Int J Epidemiol* (2016) 45(5):1600–16. doi: 10.1093/ije/dyw088

53. Voight BF, Peloso GM, Orho-Melander M, Frikke-Schmidt R, Barbalic M, Jensen MK, et al. Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. *Lancet* (2012) 380(9841):572–80. doi: 10.1016/S0140-6736(12)60312-2

54. Jansen H, Samani NJ, Schunkert H. Mendelian randomization studies in coronary artery disease. *Eur Heart J* (2014) 35(29):1917–24. doi: 10.1093/eurheartj/ehu208

55. Schunkert H, Samani NJ. Elevated C-reactive protein in atherosclerosis-chicken or egg? *N Engl J Med* (2008) 359(18):1953–5. doi: 10.1056/NEJMe0807235

56. Zacho J, Tybjaerg-Hansen A, Jensen JS, Grande P, Sillesen H, Nordestgaard BG. Genetically elevated C-reactive protein and ischemic vascular disease. *N Engl J Med* (2008) 359(18):1897–908. doi: 10.1056/NEJMoa0707402

57. Kamstrup PR, Tybjaerg-Hansen A, Steffensen R, Nordestgaard BG. Genetically elevated lipoprotein(a) and increased risk of myocardial infarction. *JAMA* (2009) 301(22):2331–9. doi: 10.1001/jama.2009.801

58. Vansteelandt S, Dukes O, Martinussen T. Survivor bias in Mendelian randomization analysis. *Biostatistics* (2017). doi: 10.1093/biostatistics/kxx050

59. Boef AG, Le Cessie S, Dekkers OM. Mendelian randomization studies in the elderly. *Epidemiology* (2015) 26(2):e15–16. doi: 10.1097/EDE.0000000000000243

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.

GWAS Reveal Targets in Vessel Wall Pathways to Treat Coronary Artery Disease

Adam W. Turner[1], Doris Wong[1,2], Caitlin N. Dreisbach[1,3] and Clint L. Miller[1,2,3,4*]

[1] Center for Public Health Genomics, University of Virginia, Charlottesville, VA, United States, [2] Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, VA, United States, [3] Data Science Institute, University of Virginia, Charlottesville, VA, United States, [4] Department of Public Health Sciences, University of Virginia, Charlottesville, VA, United States

Coronary artery disease (CAD) is the leading cause of mortality worldwide and poses a considerable public health burden. Recent genome-wide association studies (GWAS) have revealed >100 genetic loci associated with CAD susceptibility in humans. While a number of these loci harbor gene targets of currently approved therapies, such as statins and PCSK9 inhibitors, the majority of the annotated genes at these loci encode for proteins involved in vessel wall function with no known drugs available. Importantly many of the associated genes linked to vascular (smooth muscle, endothelial, and macrophage) cell processes are now organized into distinct functional pathways, e.g., vasodilation, growth factor responses, extracellular matrix and plaque remodeling, and inflammation. In this mini-review, we highlight the most recently identified loci that have predicted roles in the vessel wall and provide genetic context for pre-existing therapies as well as new drug targets informed from GWAS. With the development of new modalities to target these pathways, (e.g., antisense oligonucleotides, CRISPR/Cas9, and RNA interference) as well as the computational frameworks to prioritize or reposition therapeutics, there is great opportunity to close the gap from initial genetic discovery to clinical translation for many patients affected by this common disease.

Keywords: genome-wide association study (GWAS), coronary artery disease (CAD), drug targets, smooth muscle cells, vascular wall

INTRODUCTION

Coronary artery disease (CAD) is a maladaptive inflammatory disease of the coronary artery vessel wall that remains one of the leading causes of death worldwide. It involves numerous cell types (smooth muscle cells, endothelial cells, and macrophages) and often manifests in myocardial infarction. Development of CAD is due to a combination of genetic and environmental factors. Early twin studies indicated CAD heritability was ~40-60% (1, 2). Linkage and family-based studies identified genes with now well-established roles in disease pathogenesis, such as the LDL receptor (LDLR) (3), apolipoprotein B (apoB) (4), and proprotein convertase subtilisin/kexin type 9 (PCSK9) (5).

In 2007 the first genome-wide association studies (GWAS) of CAD published the association of the 9p21 locus with both CAD and myocardial infarction (MI) (6–8). The 9p21 locus remains the most robust locus in the genome with respect to CAD association. Many more CAD loci have been discovered in subsequent GWAS over the past decade, leading to the formation of the CARDIoGRAM (9) and C4D (10) consortia and resulting meta-analyses (11–15). The most recent GWAS meta-analysis for CAD has ~300,000 combined cases and controls and identified almost 100 independent loci reaching genome-wide significance ($p < 5 \times 10^{-8}$), and over 300 loci significant at a 5% false discovery rate.

Despite the discovery of many new loci associated with CAD, the current challenges are to validate the causal genes and pathways at CAD loci and to translate this knowledge into new therapies. In this mini-review, we highlight recent GWAS identified non-lipid genes and pathways (with an emphasis on vessel wall pathways) that have the potential to accelerate new treatments for CAD (**Figure 1**). In addition, we provide some genetic perspective on currently approved and future therapies, as well as the use of genetic risk scores (GRS) to identify high risk patients who may require these novel treatments to augment traditional lipid-lowering therapy.

CAD GWAS GENES AND PATHWAYS

Vessel Wall Signaling

Once atherogenic lipoproteins have crossed the endothelium and are taken up by macrophage-derived foam cells, there is a subsequent cascade of complex signaling events in the vessel wall. This involves a tightly orchestrated interplay of vascular smooth muscle cells, endothelial cells, macrophages, cytokines, and extracellular matrix proteins. Reactome pathway gene-set enrichment analysis carried out by the CARDIoGRAM consortium indicated that CAD genes were enriched for pathways involved in NO/cGMP signaling, TGFβ/SMAD signaling, PDGF signaling, extracellular matrix (ECM) integrity/organization, and innate immunity (16). Further integrative analyses of CARDIoGRAM summary data, tissue-specific regulatory networks and gene expression data have revealed interactions across CAD-relevant pathways as well as potential druggable targets such as LUM and STAT3, which serve as key regulators of vessel wall biology (17). Assuming that the genes in these pathways are the most likely causal genes at the associated loci, these results argue that vascular wall pathways have comparable associations to the well-established lipid and lipoprotein mediated pathways (16). In fact up to 75% of the 95 CAD loci (15) appear to be associated independently of classical risk factors. This observation suggests that these risk factors are intrinsic to dysregulated processes in the vessel wall.

NO/cGMP Signaling

NO/cGMP signaling is fundamental to diverse cardiovascular physiological responses and emerging evidence suggests that activation of this pathway is defective in the setting of atherosclerosis and CAD. Nitric oxide (NO) is an important gas that is synthesized by endothelial nitric oxide synthase (eNOS),

which upon activation results in paracrine signaling through the myoendothelial junction to smooth muscle cells, subsequent activation of soluble guanylate cyclase, cGMP production, and cGMP-dependent protein kinase (protein kinase G; PKG) mediated phosphorylation of downstream targets involved in vasodilation. The 1000 Genomes based CARDIoGRAMplusC4D (12) and recent UK Biobank-CARDIoGRAMplusC4D meta-analysis (15) identified an association for rs3918226 at *NOS3,* the gene which encodes eNOS, implicating a role in endothelial dysfunction. An intronic variant rs7692387 in *GUCY1A3,* encoding the alpha1-subunit of sGC, was associated with CAD (11), while another variant rs13139571 was associated with systolic (SBP) and diastolic (DBP) blood pressure (18). Recent functional studies identified a mechanism by which the non-risk allele at rs7692387 preferentially binds the ZEB1 transcription factor leading to increased *GUCY1A3* expression and sGC levels, which correlated with reduced atherosclerosis severity in mice (19). Other members of this pathway that have been linked to CAD include recently identified *PDE5A* (rs7678555) (15) and *PDE3A,* previously associated at 5% FDR (11), suggesting alterations in vascular wall signaling could be rescued with existing therapies (e.g., sildenafil, cilostazol).

TGFβ and PDGF Signaling

The CARDIoGRAM GWAS studies have implicated several components of the transforming growth factor beta (TGFβ) signaling pathway in CAD. Activated TGFβ receptor I phosphorylates receptor-regulated SMAD proteins (SMAD3 or SMAD2). These are transcriptional mediators of TGFβ signaling that along with SMAD4 translocate to the nucleus to regulate transcription of TGFβ target genes. The *TGFβ1* and *SMAD3* genes are both associated with CAD in addition to bone morphogenic protein 1 (*BMP1*), a member of the TGF beta superfamily (20). Mechanistic studies have implicated a functional intronic SNP in *SMAD3* (rs17293632) that disrupts binding of the AP-1 transcription factor complex underlying this association (21, 22). The genetic association of rs36096196 at the *SKI* locus suggests a role for SKI, a co-repressor of SMAD3/SMAD2 signaling in CAD (23).

The rs150512726 SNP [proxy for the recently reported SNP rs142695226 (15)] results in a 3 amino acid deletion in the integrin beta 5 (*ITGB5*) protein. ITGB5 has been shown to play a role in activation of the latent TGFβ precursor protein outside the cell (24). The TGFβ pathway also regulates gene expression at the 9p21 locus. SNPs at this locus disrupt TEAD factor binding and the TEAD3-dependent TGF beta induction of p16 in human aortic smooth muscle cells (25).

The CARDIoGRAM studies have also identified SNPs at the platelet-derived growth factor D (*PDGFD*) locus associated with CAD at genome-wide significance. This PDGF mediated pathway may involve many other risk-associated genes. Preliminary work by our group has provided evidence of cross-talk with smooth muscle cell enriched pathways using genome-wide profiling of these cells. For example, the expression of *TCF21*, a transcription factor which determines the fate of epicardial progenitor cells during development, is increased in individuals carrying the risk alleles, rs121902987 and rs12524865 (26). Its expression
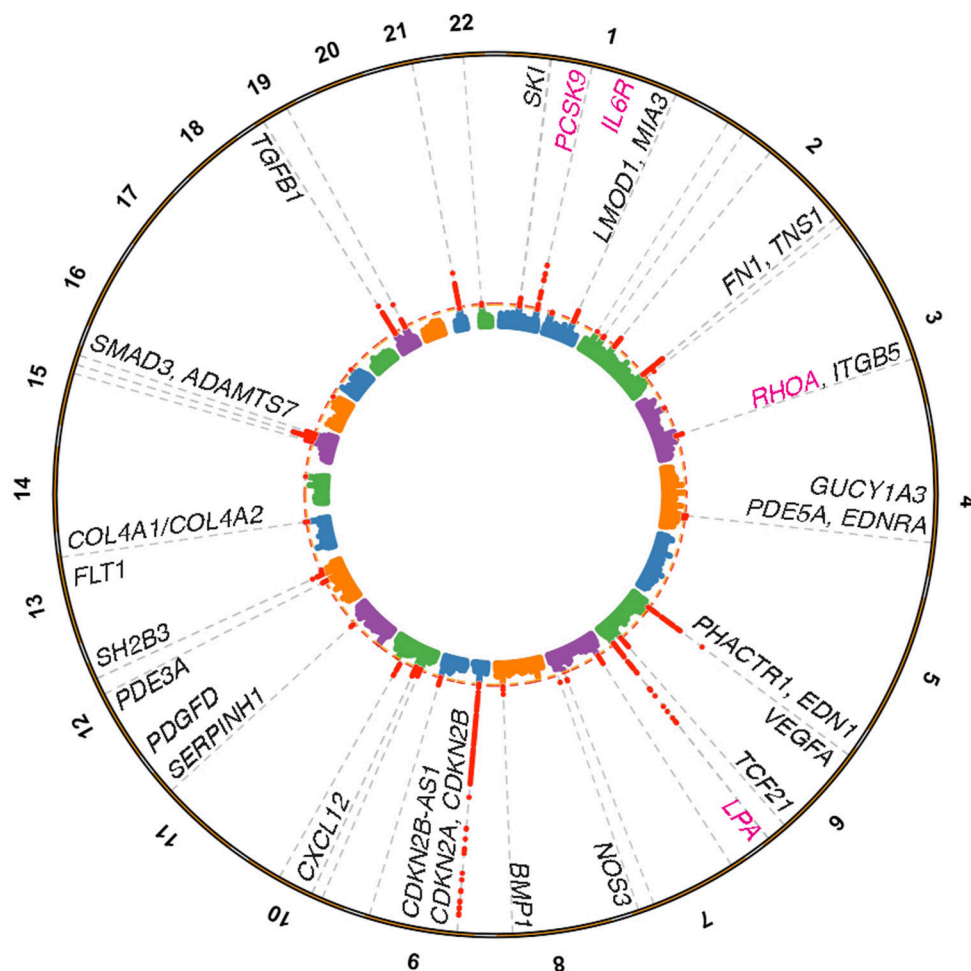
FIGURE 1 | Coronary artery disease loci harboring genes linked to vessel wall functions. Manhattan plot depicting genome-wide significant loci identified from the Nelson et al. (15) meta-analysis for CAD based on a 5% false discovery rate using the pilot UK Biobank data. Loci were annotated through a combination of gene expression, epigenomic features, eQTL, and literature based searches. Vessel wall gene symbols are shown above associations in red, with dotted orange line representing $P = 5.0 \times 10^{-8}$. Pink gene symbols represent loci with either approved or evaluated treatments for CAD.

was positively regulated by PDGF-BB-PDGFRB stimulation in human coronary artery smooth muscle cells (26). *TCF21* dysregulation likely increases CAD risk by altering coronary artery smooth muscle cell responses to vascular injury during plaque remodeling (27, 28). Another vessel wall gene, *LMOD1*, an actin filament nucleator, was shown to be downregulated in vascular smooth muscle cells in response to PDGF treatment and serves as a potent marker of smooth muscle cell phenotypic modulation (29).

Extracellular Matrix Remodeling Pathways

The CARDIoGRAM consortium has highlighted numerous extracellular matrix and basement membrane genes involved in the pathogenesis of atherosclerosis, including *COL4A1/COL4A2*, *ITGB5*, and *FN1*. A *COL4A2* variant, rs4773144, was associated with both *COL4A1* and *COL4A2* expression, as well as smooth muscle cell survival, and plaque stability (30). The authors

suggest type IV collagen levels affect SMC proliferation, migration, extracellular matrix remodeling, apoptosis, and infiltration of immune cells through plaque remodeling. The CAD locus *MIA3* is involved in the endoplasmic reticulum export of large cargo such as pre-chylomicrons/VLDL (31) and collagens (including Col4a1 and Col4a2 in mice) (32). The CAD locus *SERPINH1* encodes heat-shock protein 47 (Hsp47) (33), a molecular chaperone involved in the collagen secretion pathway. *FN1* encodes fibronectin, a glycoprotein with established roles in cell adhesion, migration, growth, and differentiation. Though increased in atherosclerotic regions, the role of fibronectin in development of CAD remains unclear, with postulated roles in atherogenic lipoprotein retention, direct adverse effects on endothelial cell function, or roles in plaque stability (34). The *TNS1* gene encodes for the tensin-1 protein that attaches the plasma membrane to the extracellular matrix and positively regulates the small Rho GTPase, RhoA

(35). The *RHOA* gene itself was identified as a genome-wide significant locus in the latest CARDIoGRAMplusC4D meta-analysis (15) and is predicted to interact with several other CAD genes/pathways in smooth muscle cells and endothelial cells including TGFβ/SMAD3 and ECM proteins, such as collagens and fibronectin (36). RhoA also cooperates with Rac1 and cadherin to regulate barrier function in mural and endothelial cells (37). RhoA activation coincides with endothelial cell inflammation, permeability, and disturbed flow as a result of reduced PPAP2B (itself associated with CAD and ischemic stroke) (38). Lastly, the *ADAMTS7* gene, encoding a metalloproteinase, is proatherogenic based on mouse studies, with a direction of effect consistent with the human genetic association data (39). In the context of its association with CAD, it has been proposed that ADAMTS7 alters smooth muscle cell migration and extracellular matrix composition (40).

Inflammation and Immune Pathways

The role of inflammation in CAD pathogenesis is now well-established, yet the number of inflammatory genes mapping to CAD-associated loci is under-represented. One of the main CAD loci involved in inflammation is the interleukin 6 receptor (*IL6R*), which binds the pro-inflammatory cytokine IL-6 and its pathways have been causally linked to CAD using Mendelian randomization analyses (41). Another example is the CAD-associated *CXCL12* gene, which encodes an anti-inflammatory cytokine (also known as stromal derived factor 1; SDF-1) that binds the chemokine (C-X-C motif) receptor CXCR4, a G-protein coupled receptor. Given that *CXCL12* is induced immediately after vessel injury and specifically expressed in atherosclerotic lesions, this gene has potential to serve as a biomarker for early detection (42). The CAD-associated *SH2B3* gene encoding an adapter protein known as LNK is involved in hematopoiesis and suppression of cytokines and thrombopoietin signaling (43). In mice, loss of Sh2b3 was shown to promote both atherosclerosis and thrombosis only under the setting of hypercholesterolemia, suggesting an involvement in platelet/leukocyte activation during atherogenesis (44). It may also serve as an inflammatory link between vascular endothelial cells and immune cells and therapeutic target for hypertension and end-organ inflammation (45). Finally, the ligand *VEGFA* and the VEGF receptor (*FLT1*) loci both associate with CAD; inflammatory conditions in the plaque promote the release of angiogenic factors that result in neovascularization, plaque remodeling, and plaque instability (46).

CURRENT THERAPIES FOR CAD

Current therapies for CAD primarily focus on alleviating the symptoms of ischemic events as well as preventing thrombosis from ruptured plaque. Here we review the current treatments for CAD and also provide a genetically informed perspective on these drug targets (**Table 1**).

Statins

Statins represent the first line of treatment for elevated LDL-cholesterol levels associated with hyperlipidemia and CAD. By inhibiting HMG-coa (hydroxy-3-methylglutaryl-coenzyme A) reductase, statins decrease the production of cholesterol in the liver, thereby reducing its concentration in the circulation. Statins exhibit a pleiotropic effect by attenuating other risk factors for CAD (64). Genetic studies have identified variations in the *HMGCR* gene (rs12916) consistently associated with both blood lipids and LDL-cholesterol (52, 53), while an intergenic variant near *HMGCR* is also associated with CAD in the combined CARDIoGRAMplusC4D and UK Biobank analysis (20).

Anti-Platelet Therapies

As a prophylactic measure against thrombosis, antiplatelet drugs are utilized to reduce the risk of myocardial infarction. Two of the more popular antiplatelet drugs are acetylsalicylic acid (ASA) and clopidogrel. ASA is a COX inhibitor that prevents platelet activation by inhibiting the synthesis of thromboxane A2. On the other hand, clopidogrel is an ADP receptor (P2Y12) antagonist that prevents platelet aggregation and further amplification of the activation signal through the downregulation of glycoprotein IIb/IIIa receptor on its surface (65). While the gene targets of these drugs (*PTSG2* and *P2Y12*) do not harbor variants specifically associated with CAD, some of the effector signaling molecules in the pathway (*RHOA*, *ITGB5*, and *SH2B3*) indeed have CAD associations, as described above. This may represent an opportunity to understand some of the heterogeneity in responses to these commonly used agents by using a pathway approach.

ACE Inhibitors and Beta Blockers

Two classes of drugs, angiotensin converting enzymes (ACE) inhibitors and beta blockers both function in the maintenance of normal blood pressure. In the endothelium, ACE catalyzes the conversion of angiotensin I to angiotensin II where the latter is a potent vasoconstrictor. Additionally, ACE upregulation results in the degradation of bradykinin, a vasodilatory factor involved in the upregulation of nitric oxide and prostaglandins (66). Given the numerous CAD associations within the NO/cGMP pathway, the efficacy or toxicity profile of these drugs may be influenced by individual genetic variation. Beta blockers exert their cardioprotective effects by intervening in the adrenergic nervous system as competitive antagonists in both the myocardium and vasculature, depending on their selectivity for beta1 or beta2-adrenergic receptors. Clinically, reduced catecholamine stimulation results in decreased cardiac stress leading to decreased heart rate and blood pressure (67, 68). The third generation of beta blockers were shown to have more potent blood pressure lowering effects. Although it may be reasonable to speculate that NO mediated signaling is involved, it was recently demonstrated that nebivolol (compared to metoprolol) suppresses ET-1 mediated vasoconstriction to lower BP (69). This is important given that variation at the ET-1 gene *EDN1* (rs1629862) and the ET-1 receptor type A gene *EDNRA* (rs6841581) were recently identified as CAD loci (20).

Anti-Inflammatory Therapies

Therapies targeting inflammatory pathways have been extensively explored in cardiovascular disease. Two recent

TABLE 1 | List of current target genes for management of coronary artery disease and their genetic associations.

Target (gene name)	Genetic association with CAD/lipid trait/BP	Associated trait (lead SNP)	Drug(s)	Role	Phase	Administration
HMG-Coenzyme A reductase (*HMGCR*)	Yes	LDL-C (rs7703051) (47, 48) Total cholesterol (rs10038095) (49) Plasma LDL-C response to simvastatin (rs17244841) (50) Lipid traits (rs10045497) (51)	Statins (various)	Pharmacological inhibitors	Available with prescription	Oral
Cholesterol absorption in small intestine (gene unclear)	NA	NA	Ezetimibe (Zetia, Ezetrol)	Pharmacological inhibitor	Available with prescription	Oral
ATP citrate lyase (*ACLY*)	NA	NA	Bempedoic acid (Esperion)	Pharmacological inhibitor	Clinical trials (phase 3)	Oral
ApoB-100 (*APOB*)	Yes	LDL-C (rs1367117) (52–54) Total cholesterol (rs1367117) (52–54) Lipid metabolism phenotypes (rs1367117) (55) Triglycerides (rs1042034) (53) HDL-C (rs1042034) (53) Oxidized LDL (rs676210) (56) Lipid metabolism phenotypes (rs676210) (55) LDL-C (rs693) (47, 57–59) Total cholesterol (rs693) (57) Triglycerides (rs693) (48)	Mipomersen (Kynamro)	Antisense oligonucleotide (targets mRNA)	Available with prescription	Injection
PCSK9 (*PCSK9*)	Yes	CAD (rs11591147) (15, 20) LDL-C (rs11591147) (48, 49)	Repatha (evalocumab)	Monoclonal antibody	Available with prescription	Injection
			Praluent (alirocumab)	Monoclonal antibody	Available with prescription	Injection
			Bococizumab	Monoclonal antibody	Clinical trials	Injection
			Inclisiran	Long acting small interfering RNA (siRNA)	Clinical trials (phase 3)	Injection
Lipoprotein A (Lp(a)) (*LPA*)	Yes	CAD (rs10455872) (15) CAD (rs186696265) (20) LDL-C in response to statins (rs10455872) (60)	AKCEA-Apo(a)-LRx	Antisense oligonucleotide	Clinical trials (phase 2b)	Injection
ApoCIII (*APOC3*)	Yes	Triglyceride levels (rs76353203) (80) HDL-C (rs76353203) (80)	AKCEA-ApoCIII-LRx	Antisense oligonucleotide	Clinical trials (phase 2b)	Injection
ANGPTL3 (*ANGPTL3*)	Yes	Triglyceride levels (rs2131925) (52, 53) LDL-C (rs2131925) (52, 53) Total cholesterol (rs2131925) (52, 53)	AKCEA-ANGPTL3-LRx	Antisense oligonucleotide	Clinical trials (phase 2)	Injection
Interleukin 1 beta (*IL1B*)	NA	NA	Canakinumab	Monoclonal antibody	Clinical trials (phase 3)	Injection
Cyclooxygenase-2 (COX-2) (*PTGS2*)	NA	NA	Acetylsalicylic acid (ASA, Aspirin)	Pharmacological inhibitor (general anti-inflammatory effects)	Commercially available	Oral
P2Y$_{12}$ subunit of ADP receptor (*P2RY12*)	NA	NA	Clopidogrel (Plavix)	Pharmacological inhibitor	Available with prescription	Oral

(Continued)

TABLE 1 | Continued

Target (gene name)	Genetic association with CAD/lipid trait/BP	Associated trait (lead SNP)	Drug(s)	Role	Phase	Administration
Angiotensin converting enzyme (ACE) (*ACE*)	Yes	Diastolic blood pressure (rs4308) (61, 62)	ACE inhibitors	Pharmacological inhibitors	Available with prescription	Oral
Beta adrenergic receptor(s) (*ADRB1*, *ADRB2*, *ADRB3*)	NA	NA	Beta blockers	Pharmacological inhibitors	Available with prescription	Oral
Rho kinase (*ROCK1, ROCK2*)	NA	CAD, sudden cardiac arrest (rs6716724) (63)	Fasudil	Pharmacological inhibitor	Approved in China and Japan	Oral

studies investigating the role of clonal expansion of hematopoietic cells as a potential driver for age-related onset of atherosclerosis have provided evidence that IL1β secretion from TET2 deficient macrophages plays a role in the acceleration of disease (70, 71). TET2 is an epigenetic modifier that negatively regulates the expression of IL1β. Thus, loss of function of TET2 results in the upregulation of IL1β and IL-6 secretion from lesional macrophages (70). This elevated level of proinflammatory cytokines was positively correlated with increased plaque size in the aorta as well as severity of coronary artery calcification in mice and human patients, respectively(70, 71). Studies such as these underscore the potential of targeting the IL1β pathway in slowing down atherosclerosis progression.

The CANTOS (NCT01327846) clinical trial provided critical evidence that targeting IL1β alone with the monoclonal antibody canakinumab can reduce major cardiovascular events along with proinflammatory cytokines (IL-6) and high sensitivity C reactive protein in patients with atherosclerosis. Although the intermediate dose (150 mg) met the primary endpoint of reducing nonfatal myocardial infarction, nonfatal stroke, or cardiovascular death, a significant risk of fatal infection relative to placebo was observed (72). In addition to the high pricing and safety concerns, the marginal clinical benefits demand more development in this area. Given that IL-6 is a causal risk factor for CAD (73), anti-inflammatory therapies remain an attractive therapeutic approach for patients that do not respond to standard lipid lowering medication.

NEW CAD THERAPIES INFORMED FROM GENETIC STUDIES

PCSK9 Inhibitors and Antisense Oligonucleotides

One example of newly approved drug targets that have origins in genetic studies is the development of monoclonal antibodies against PCSK9. PCSK9 is a liver protease that targets LDL receptors for lysosomal degradation. The therapeutic potential of targeting PCSK9 was validated through Mendelian randomization studies that correlated a deleterious mutation in this gene with decreased risk (74). Large clinical trials [e.g., FOURIER (NCT01764633), ODYSSEY (NCT01623115)]

demonstrated that inhibition of this enzyme reduced systemic LDL levels to a greater extent than maximum statin therapy, with the most recent ODYSSEY trial (NCT01663402) reporting a reduction in both cardiovascular events and all-cause mortality for the first time. In addition to monoclonal antibodies, antisense oligonucleotides have also been developed against PCSK9, which should be evaluated for clinical outcomes in the near future.

Lipoprotein A and APOC3 Antisense Oligonucleotides

A high level of circulating lipoprotein A [Lp(a)] is considered a risk factor for cardiovascular disease. Two SNPs, rs3798220 and rs10455872, located within the lipoprotein A (*LPA*) gene correlate with increased levels of Lp(a) and are associated with increased risk for CAD. As one of the first therapies targeting lipoprotein A, AKCEA-APO(a)-L$_{Rx}$ is an antisense oligonucleotide that binds LPA mRNA leading to its degradation. Phase 2 clinical trial data has suggested that this approach is well tolerated and significantly reduced Lp(a) plasma concentrations (65, 66). Similarly, an antisense therapy was developed targeting *APOC3*, a gene involved in regulating plasma triglyceride levels. The antisense oligonucleotide therapy, volanesorsen was shown to reduce cellular levels of APOC3 and led to an overall reduction of triglyceride levels in phase 3 clinical trials (70, 71).

RhoA-ROCK Inhibition

The RhoA-Rock signaling pathway offers another avenue for CAD therapeutic targets. Aberrant activation of this signaling cascade has been implicated in vasoconstriction and endothelial dysfunction. Given the recent CAD association (rs7623687) at *RHOA*, further investigation is warranted to determine how to specifically target this gene. One opportunity is to target the downstream effectors, Rho-associated protein kinases (ROCK1, ROCK2), which control actin cytoskeleton arrangement, cell migration and contractility (75). In particular, a Rock2 inhibitor, Fasudil, has already been tested in clinical trials as a possible therapeutic for CAD as a vasodilatory agent through the upregulation of nitric oxide. It is also noteworthy to mention that Fasudil has been approved as a treatment option for cerebral vasospasm in Japan and China (75).

PERSPECTIVES AND FUTURE DIRECTIONS

Genetic Risk Scores

Besides the 9p21 locus, most loci uncovered from GWAS of CAD have small effects with odds ratios between 1.05 and 1.30. Nonetheless, GWAS results can be utilized to generate genetic risk scores for individuals based on the number of risk alleles they harbor. Therefore, in addition to traditional drug treatments such as statins, individuals that fall within the high CAD risk range based on their genetic risk score can be selected for more aggressive therapies and/or novel CAD treatments as mentioned above. With more data from sources such as the UK Biobank, the Million Veterans Project, and the NIH-funded All of Us project on the horizon, genetic risk scores will have more clinically-relevant predictive utility (76).

Feasible vs. Difficult Drug Targets

Since GWAS has highlighted the role of vessel wall genes and signaling pathways in the pathogenesis of CAD, it will be critical to apply this knowledge toward vessel wall therapeutic development. Strategies include non-specific targeting of the vessel wall (through upstream or downstream effector molecules), specifically targeting plaque vasculature, or specific cellular phenotypes (e.g., activated resident macrophages or phenotypically modulated smooth muscle cells).

Target New Cell Types (e.g., Endothelial Cells, Smooth Muscle Cells, Macrophages)

A CAD protective variant upstream of *ADAMTS7* confers greater protection against CAD for never-smokers compared to those that have smoked 100 or more cigarettes in their lifetime (77). This example highlights the importance of taking into account environmental factors in managing treatments. Other potential targets include the receptors for endothelin-1 on smooth muscle cells. Many of these potential vessel wall target proteins affect smooth muscle cell proliferation and migration, originally believed to drive atherogenesis. The current view suggests smooth muscle cell proliferation and migration could be reparative and promote plaque stability (78). Once the roles and timing of smooth muscle cell proliferation and migration are clarified, the TGF beta and PDGF pathways may be attractive targets due to their role in the regulation of smooth muscle cell genes.

Machine Learning/Systems Approaches

While GWAS has uncovered invaluable insights into potential therapies and validated existing ones, these associations require extensive follow-up to pinpoint causal variants, genes, pathways. More advanced algorithms such as machine learning can be leveraged to prioritize targets with diverse data inputs such as electronic health records, clinical notes, and -omics. These approaches can help to systematically decrease noise, reduce features, and identify gene sets of interest in addition to common GWAS methods of odds ratios, *p*-value statistics, and chi-square comparisons. Unsupervised learning algorithms have the capability to provide researchers and clinicians with an unbiased network of candidate genes that account for the greatest *variance* in CAD related phenotypes. A specific example is the use of machine learning for drug repurposing based on finding patterns from multi-dimensional datasets. Specific tools have been developed to provide an out-of-the-box approach for understanding diverse text, biological, and medical record data for non-data scientists. One such tool, RepurposeDB, combines drug and disease information to create a reference database for drug repositioning research (79). With the rapidly growing costs of drug discovery/development, such data-informed approaches can offer significant progress for the field.

CONCLUDING REMARKS

In summary, in this brief review we bring attention to the genetic loci discovered over the past decade which play critical roles in the vessel wall. Many of these genes are organized into distinct functional pathways, which will help redefine some of the pathogenic mechanisms and prioritize those pathways for future drug development or repurposing strategies.

AUTHOR CONTRIBUTIONS

AT and CM conceived of the manuscript. AT, DW, CD, and CM wrote the manuscript.

FUNDING

REFERENCES

1. Zdravkovic S, Wienke A, Pedersen NL, Marenberg ME, Yashin AI, De Faire U. Heritability of death from coronary heart disease: a 36-year follow-up of 20 966 Swedish twins. *J Intern Med.* (2002) 252:247–54. doi: 10.1046/j.1365-2796.2002.01029.x

2. Lloyd-Jones DM, Nam B-H, D'Agostino RB Sr, Levy D, Murabito JM, Wang TJ, et al. Parental cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults. *JAMA* (2004) 291:2204. doi: 10.1001/jama.291.18.2204

3. Watts GF, Lewis B, Sullivan DR. Familial hypercholesterolemia: a missed opportunity in preventive medicine. *Nat Clin Pract Cardiovasc Med.* (2007) 4:404–5. doi: 10.1038/ncpcardio0941

4. Innerarity TL, Mahley RW, Weisgraber KH, Bersot TP, Krauss RM, Vega GL, et al. Familial defective apolipoprotein B-100: a mutation of apolipoprotein B that causes hypercholesterolemia. *J Lipid Res.* (1990) 31:1337–49.

5. Abifadel M, Varret M, Rabès J-P, Allard D, Ouguerram K, Devillers M, et al. Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nat Genet.* (2003) 34:154–6. doi: 10.1038/ng1161

6. 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

7. 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

8. Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, Mayer B, et al. Genomewide association analysis of coronary artery disease. *N Engl J Med.* (2007) 357:443–53. doi: 10.1056/NEJMoa0 72366

9. Schunkert H, König IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet.* (2011) 43:333–8. doi: 10.1038/ng.784

10. Coronary Artery Disease (C4D) Genetics Consortium. A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. *Nat Genet.* (2011) 43:339–44. doi: 10.1038/ng.782

11. CARDIoGRAMplusC4D Consortium, 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

12. Nikpay M, Goel A, Won H-H, 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

13. Webb TR, Erdmann J, Stirrups KE, Stitziel NO, Masca NGD, Jansen H, et al. Systematic evaluation of pleiotropy identifies 6 further loci associated with coronary artery disease. *J Am Coll Cardiol.* (2017) 69:823–36. doi: 10.1016/j.jacc.2016.11.056

14. Howson JMM, Zhao W, Barnes DR, Ho W-K, Young R, Paul DS, et al. Fifteen new risk loci for coronary artery disease highlight arterial-wall-specific mechanisms. *Nat Genet.* (2017) 49:1113–9. doi: 10.1038/ng.3874

15. Nelson CP, Goel A, Butterworth AS, Kanoni S, Webb TR, Marouli E, et al. Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nat Genet.* (2017) 49:1385–91. doi: 10.1038/ng.3913

16. Ghosh S, Vivar J, Nelson CP, Willenborg C, Segrè AV, Mäkinen V-P, et al. Systems genetics analysis of genome-wide association study reveals novel associations between key biological processes and coronary artery disease. *Arterioscler Thromb Vasc Biol.* (2015) 35:1712–22. doi: 10.1161/ATVBAHA.115.305513

17. Zhao Y, Chen J, Freudenberg JM, Meng Q, Rajpal DK, Yang X. Network-based identification and prioritization of key regulators of coronary artery disease loci. *Arterioscler Thromb Vasc Biol.* (2016) 36:928–41. doi: 10.1161/ATVBAHA.115.306725

18. International Consortium for Blood Pressure Genome-Wide Association Studies, Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* (2011) 478:103–9. doi: 10.1038/nature10405

19. Kessler T, Wobst J, Wolf B, Eckhold J, Vilne B, Hollstein R, et al. Functional characterization of the GUCY1A3 coronary artery disease risk locus. *Circulation* (2017) 136:476–89. doi: 10.1161/CIRCULATIONAHA.116.024152

20. 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.117.3 12086

21. Turner AW, Martinuk A, Silva A, Lau P, Nikpay M, Eriksson P, et al. Functional analysis of a novel genome-wide association study signal in SMAD3 that confers protection from coronary artery disease. *Arterioscler Thromb Vasc Biol.* (2016) 36:972–83. doi: 10.1161/ATVBAHA.116.307294

22. Miller CL, Pjanic M, Wang T, Nguyen T, Cohain A, Lee JD, et al. Integrative functional genomics identifies regulatory mechanisms at coronary artery disease loci. *Nat Commun.* (2016) 7:12092. doi: 10.1038/ncomms 12092

23. Xu W, Angelis K, Danielpour D, Haddad MM, Bischof O, Campisi J, et al. Ski acts as a co-repressor with Smad2 and Smad3 to regulate the response to type beta transforming growth factor. *Proc Natl Acad Sci USA.* (2000) 97:5924–9. doi: 10.1073/pnas.090097797

24. Worthington JJ, Klementowicz JE, Travis MA. TGFβ: a sleeping giant awoken by integrins. *Trends Biochem Sci.* (2011) 36:47–54. doi: 10.1016/j.tibs.2010.08.002

25. Almontashiri NAM, Antoine D, Zhou X, Vilmundarson RO, Zhang SX, Hao KN, et al. 9p21.3 Coronary artery disease risk variants disrupt TEAD transcription factor-dependent transforming growth factor β regulation of p16 expression in human aortic smooth muscle cells. *Circulation* (2015) 132:1969–78. doi: 10.1161/CIRCULATIONAHA.114.0 15023

26. Miller CL, Anderson DR, Kundu RK, Raiesdana A, Nürnberg ST, Diaz R, et al. Disease-related growth factor and embryonic signaling pathways modulate an enhancer of TCF21 expression at the 6q23.2 coronary heart disease locus. *PLoS Genet.* (2013) 9:e1003652. doi: 10.1371/journal.pgen.10 03652

27. Nürnberg ST, Cheng K, Raiesdana A, Kundu R, Miller CL, Kim JB, et al. Coronary artery disease associated transcription factor TCF21 regulates smooth muscle precursor cells that contribute to the fibrous cap. *PLoS Genet.* (2015) 11:e1005155. doi: 10.1371/journal.pgen.1005155

28. Sazonova O, Zhao Y, Nürnberg S, Miller C, Pjanic M, Castano VG, et al. Characterization of TCF21 downstream target regions identifies a transcriptional network linking multiple independent coronary artery disease loci. *PLoS Genet.* (2015) 11:e1005202. doi: 10.1371/journal.pgen.1005202

29. Perisic Matic L, Rykaczewska U, Razuvaev A, Sabater-Lleal M, Lengquist M, Miller CL, et al. Phenotypic modulation of smooth muscle cells in atherosclerosis is associated with downregulation of LMOD1, SYNPO2, PDLIM7, PLN, and SYNM. *Arterioscler Thromb Vasc Biol.* (2016) 36:1947–61. doi: 10.1161/ATVBAHA.116.307893

30. Yang W, Ng FL, Chan K, Pu X, Poston RN, Ren M, et al. Coronary-heart-disease-associated genetic variant at the COL4A1/COL4A2 locus affects COL4A1/COL4A2 expression, vascular cell survival, atherosclerotic plaque stability and risk of myocardial infarction. *PLoS Genet.* (2016) 12:e1006127. doi: 10.1371/journal.pgen.1006127

31. Santos AJM, Nogueira C, Ortega-Bellido M, Malhotra V. TANGO1 and Mia2/cTAGE5 (TALI) cooperate to export bulky pre-chylomicrons/VLDLs from the endoplasmic reticulum. *J Cell Biol.* (2016) 213:343–54. doi: 10.1083/jcb.2016 03072

32. Wilson DG, Phamluong K, Li L, Sun M, Cao TC, Liu PS, et al. Global defects in collagen secretion in a Mia3/TANGO1 knockout mouse. *J Cell Biol.* (2011) 193:935–51. doi: 10.1083/jcb.201007162

33. Nagata K. Hsp47: a collagen-specific molecular chaperone. *Trends Biochem Sci.* (1996) 21:22–6. doi: 10.1016/S0968-0004(06)80023-X

34. Moore KJ, Fisher EA. The double-edged sword of fibronectin in atherosclerosis. *EMBO Mol Med.* (2012) 4:561–3. doi: 10.1002/emmm.201200238

35. Shih Y-P, Sun P, Wang A, Lo SH. Tensin1 positively regulates RhoA activity through its interaction with DLC1. *Biochim Biophys Acta* (2015) 1853:3258–65. doi: 10.1016/j.bbamcr.2015.09.028

36. Collins C, Osborne LD, Guilluy C, Chen Z, O'Brien ET, Reader JS, et al. Haemodynamic and extracellular matrix cues regulate the mechanical phenotype and stiffness of aortic endothelial cells. *Nat Commun.* (2014) 5:3984. doi: 10.1038/ncomms4984

37. Alimperti S, Mirabella T, Bajaj V, Polacheck W, Pirone DM, Duffield J, et al. Three-dimensional biomimetic vascular model reveals a RhoA, Rac1, and N-cadherin balance in mural cell-endothelial cell-regulated barrier function. *Proc Natl Acad Sci USA.* (2017) 114:8758–63. doi: 10.1073/pnas.1618333114

38. Wu C, Huang R-T, Kuo C-H, Kumar S, Kim CW, Lin Y-C, et al. Mechanosensitive PPAP2B regulates endothelial responses to atherorelevant hemodynamic forces. *Circ Res.* (2015) 117:e41–53. doi: 10.1161/CIRCRESAHA.117.306457

39. Bauer RC, Tohyama J, Cui J, Cheng L, Yang J, Zhang X, et al. Knockout of Adamts7, a novel coronary artery disease locus in humans, reduces atherosclerosis in mice. *Circulation* (2015) 131:1202–13. doi: 10.1161/CIRCULATIONAHA.114.012669

40. Pu X, Xiao Q, Kiechl S, Chan K, Ng FL, Gor S, et al. ADAMTS7 cleavage and vascular smooth muscle cell migration is affected by a coronary-artery-disease-associated variant. *Am J Hum Genet.* (2013) 92:366–74. doi: 10.1016/j.ajhg.2013.01.012

41. IL6R Genetics Consortium Emerging Risk Factors Collaboration, Sarwar N, Butterworth AS, Freitag DF, Gregson J, Willeit P, et al. Interleukin-6 receptor pathways in coronary heart disease: a collaborative meta-analysis

of 82 studies. *Lancet* (2012) 379:1205–13. doi: 10.1016/S0140-6736(11)61931-4

42. van der Vorst EPC, Döring Y, Weber C. MIF and CXCL12 in cardiovascular diseases: functional differences and similarities. *Front Immunol.* (2015) 6:373. doi: 10.3389/fimmu.2015.00373

43. Norddahl GL, Wahlestedt M, Gisler S, Sigvardsson M, Bryder D. Reduced repression of cytokine signaling ameliorates age-induced decline in hematopoietic stem cell function. *Aging Cell* (2012) 11:1128–31. doi: 10.1111/j.1474-9726.2012.00863.x

44. Wang W, Tang Y, Wang Y, Tascau L, Balcerek J, Tong W, et al. LNK/SH2B3 loss of function promotes atherosclerosis and thrombosis. *Circ Res.* (2016) 119:e91–103. doi: 10.1161/CIRCRESAHA.116.308955

45. Saleh MA, McMaster WG, Wu J, Norlander AE, Funt SA, Thabet SR, et al. Lymphocyte adaptor protein LNK deficiency exacerbates hypertension and end-organ inflammation. *J Clin Invest.* (2015) 125:1189–202. doi: 10.1172/JCI76327

46. Camaré C, Pucelle M, Nègre-Salvayre A, Salvayre R. Angiogenesis in the atherosclerotic plaque. *Redox Biol.* (2017) 12:18–34. doi: 10.1016/j.redox.2017.01.007

47. Burkhardt R, Kenny EE, Lowe JK, Birkeland A, Josowitz R, Noel M, et al. Common SNPs in HMGCR in micronesians and whites associated with LDL-cholesterol levels affect alternative splicing of exon13. *Arterioscler Thromb Vasc Biol.* (2008) 28:2078–84. doi: 10.1161/ATVBAHA.108.172288

48. Kathiresan S, Melander O, Guiducci C, Surti A, Burtt NP, Rieder MJ, et al. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat Genet.* (2008) 40:189–97. doi: 10.1038/ng.75

49. Surakka I, Horikoshi M, Mägi R, Sarin A-P, Mahajan A, Lagou V, et al. The impact of low-frequency and rare variants on lipid levels. *Nat Genet.* (2015) 47:589–97. doi: 10.1038/ng.3300

50. Medina MW, Gao F, Ruan W, Rotter JI, Krauss RM. Alternative splicing of 3-hydroxy-3-methylglutaryl coenzyme A reductase is associated with plasma low-density lipoprotein cholesterol response to simvastatin. *Circulation* (2008) 118:355–62. doi: 10.1161/CIRCULATIONAHA.108.773267

51. Zhou L, He M, Mo Z, Wu C, Yang H, Yu D, et al. A genome wide association study identifies common variants associated with lipid levels in the Chinese population. *PLoS ONE* (2013) 8:e82420. doi: 10.1371/journal.pone.0082420

52. 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

53. 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

54. Spracklen CN, Chen P, Kim YJ, Wang X, Cai H, Li S, et al. Association analyses of East Asian individuals and trans-ancestry analyses with European individuals reveal new loci associated with cholesterol and triglyceride levels. *Hum Mol Genet.* (2017) 26:1770–84. doi: 10.1093/hmg/ddx062

55. Chasman DI, Pare G, Mora S, Hopewell JC, Peloso G, Clarke R, et al. Forty-three loci associated with plasma lipoprotein size, concentration, and cholesterol content in genome-wide analysis. *PLoS Genet.* (2009) 5:e1000730. doi: 10.1371/journal.pgen.1000730

56. Mäkelä K-M, Seppälä I, Hernesniemi JA, Lyytikäinen L-P, Oksala N, Kleber ME, et al. Genome-wide association study pinpoints a new functional apolipoprotein B variant influencing oxidized low-density lipoprotein levels but not cardiovascular events: AtheroRemo Consortium. *Circ Cardiovasc Genet.* (2013) 6:73–81. doi: 10.1161/CIRCGENETICS.112.964965

57. Aulchenko YS, Ripatti S, Lindqvist I, Boomsma D, Heid IM, Pramstaller PP, et al. Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nat Genet.* (2009) 41:47–55. doi: 10.1038/ng.269

58. Sabatti C, Service SK, Hartikainen A-L, Pouta A, Ripatti S, Brodsky J, et al. Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nat Genet.* (2009) 41:35–46. doi: 10.1038/ng.271

59. Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes of BioMedical Research, Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PIW, 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

60. Postmus I, Trompet S, Deshmukh HA, Barnes MR, Li X, Warren HR, et al. Pharmacogenetic meta-analysis of genome-wide association studies of LDL cholesterol response to statins. *Nat Commun.* (2014) 5:5068. doi: 10.1038/ncomms6068

61. Warren HR, Evangelou E, Cabrera CP, Gao H, Ren M, Mifsud B, et al. Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. *Nat Genet.* (2017) 49:403–15. doi: 10.1038/ng.3768

62. Wain LV, Vaez A, Jansen R, Joehanes R, van der Most PJ, Erzurumluoglu AM, et al. Novel blood pressure locus and gene discovery using genome-wide association study and expression data sets from blood and the kidney. *Hypertension* (2017) 70:e4–19. doi: 10.1161/HYPERTENSIONAHA.117.09438

63. Aouizerat BE, Vittinghoff E, Musone SL, Pawlikowska L, Kwok P-Y, Olgin JE, et al. GWAS for discovery and replication of genetic loci associated with sudden cardiac arrest in patients with coronary artery disease. *BMC Cardiovasc Disord.* (2011) 11:29. doi: 10.1186/1471-2261-11-29

64. Finkel JB, Duffy D. 2013 ACC/AHA cholesterol treatment guideline: paradigm shifts in managing atherosclerotic cardiovascular disease risk. *Trends Cardiovasc Med.* (2015) 25:340–7. doi: 10.1016/j.tcm.2014.10.015

65. Knight CJ. Antiplatelet treatment in stable coronary artery disease. *Heart* (2003) 89:1273–8. doi: 10.1136/heart.89.10.1273

66. Ferrari R, Guardigli G, Ceconi C. Secondary prevention of CAD with ACE inhibitors: a struggle between life and death of the endothelium. *Cardiovasc Drugs Ther.* (2010) 24:331–9. doi: 10.1007/s10557-010-6244-x

67. Andreasen C, Andersson C. Current use of beta-blockers in patients with coronary artery disease. *Trends Cardiovasc Med.* (2018). doi: 10.1016/j.tcm.2017.12.014. [Epub ahead of print].

68. Barrese V, Taglialatela M. New advances in beta-blocker therapy in heart failure. *Front Physiol.* (2013) 4:323. doi: 10.3389/fphys.2013.00323

69. Diehl KJ, Stauffer BL, Dow CA, Bammert TD, Brunjes DL, Greiner JJ, et al. Chronic nebivolol treatment suppresses endothelin-1-mediated vasoconstrictor tone in adults with elevated blood pressure. *Hypertension* (2016) 67:1196–204. doi: 10.1161/HYPERTENSIONAHA.115.06979

70. Gaudet D, Alexander VJ, Baker BF, Brisson D, Tremblay K, Singleton W, et al. Antisense inhibition of apolipoprotein C-III in patients with hypertriglyceridemia. *N Engl J Med.* (2015) 373:438–47. doi: 10.1056/NEJMoa1400283

71. Gouni-Berthold I. The role of antisense oligonucleotide therapy against apolipoprotein-CIII in hypertriglyceridemia. *Atheroscler Suppl.* (2017) 30:19–27. doi: 10.1016/j.atherosclerosissup.2017.05.003

72. Ridker PM, Everett BM, Thuren T, MacFadyen JG, Chang WH, Ballantyne C, et al. Antiinflammatory therapy with canakinumab for atherosclerotic disease. *N Engl J Med.* (2017) 377:1119–31. doi: 10.1056/NEJMoa1707914

73. Interleukin-6 Receptor Mendelian Randomisation Analysis (IL6R MR) Consortium, Swerdlow DI, Holmes MV, Kuchenbaecker KB, Engmann JEL, Shah T, et al. The interleukin-6 receptor as a target for prevention of coronary heart disease: a mendelian randomisation analysis. *Lancet* (2012) 379:1214–24. doi: 10.1016/S0140-6736(12)60110-X

74. Kastelein JJP. Decade in review–dyslipidaemia: resurgence of targets and compounds to treat dyslipidaemia. *Nat Rev Cardiol.* (2014) 11:629–31. doi: 10.1038/nrcardio.2014.132

75. Lock FE, Ryan KR, Poulter NS, Parsons M, Hotchin NA. Differential regulation of adhesion complex turnover by ROCK1 and ROCK2. *PLoS ONE* (2012) 7:e31423. doi: 10.1371/journal.pone.0031423

76. Abraham G, Havulinna AS, Bhalala OG, Byars SG, De Livera AM, Yetukuri L, et al. Genomic prediction of coronary heart disease. *Eur Heart J.* (2016) 37:3267–78. doi: 10.1093/eurheartj/ehw450

77. Saleheen D, Zhao W, Young R, Nelson CP, Ho W, Ferguson JF, et al. Loss of cardioprotective effects at the ADAMTS7 Locus as a result of gene-smoking interactions. *Circulation* (2017) 135:2336–53. doi: 10.1161/CIRCULATIONAHA.116.0 22069

78. Bennett MR, Sinha S, Owens GK. Vascular smooth muscle cells in atherosclerosis. *Circ Res*. (2016) 118:692–702. doi: 10.1161/CIRCRESAHA.115.306361

79. Shameer K, Glicksberg BS, Hodos R, Johnson KW, Badgeley MA, Readhead B, et al. Systematic analyses of drugs and disease indications in RepurposeDB reveal pharmacological, biological and epidemiological factors influencing drug repositioning. *Brief Bioinformatics* (2017) 375:4. doi: 10.1093/bib/bbw136

80. Southam L, Gilly A, Süveges D, Farmaki A-E, Schwartzentruber J, Tachmazidou I, et al. Whole genome sequencing and imputation in isolated populations identify genetic associations with medically-relevant complex traits. *Nat Commun.* (2017) 8:15606. doi: 10.1038/ncomms 15606

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.

Integrating Genes Affecting Coronary Artery Disease in Functional Networks by Multi-OMICs Approach

Baiba Vilne[1,2] and Heribert Schunkert[1,2]*

[1] Deutsches Herzzentrum München, Klinik für Herz- und Kreislauferkrankungen, Technische Universität München, Munich, Germany, [2] Munich Heart Alliance, German Centre for Cardiovascular Research, Munich, Germany

Coronary artery disease (CAD) and myocardial infarction (MI) remain among the leading causes of mortality worldwide, urgently demanding a better understanding of disease etiology, and more efficient therapeutic strategies. Genetic predisposition as well as the environment and lifestyle are thought to contribute to disease risk. It is likely that non-linear and complex interactions occur between these multiple factors, involving simultaneous pathological changes in diverse cell types, tissues, and organs, at multiple molecular levels. Recent technological advances have exponentially expanded the breadth of available -omics data, from genome, epigenome, transcriptome, proteome, metabolome to even the microbiome. Integration of multiple layers of information across several -omics domains, i.e., the so-called multi-omics approach, currently holds the promise as a path toward precision medicine. Indeed, a more meaningful interpretation of genotype-phenotype relationships and the development of successful therapeutics tailored to individual patients are urgently needed. In this review, we will summarize recent findings and applications of integrative multi-omics in elucidating the etiology of CAD/MI; with a special focus on established disease susceptibility loci sequentially identified in genome-wide association studies (GWAS) over the last 10 years. Moreover, in addition to the autosomal genome, we will also consider the genetic variation in our "second genome"—the mitochondrial genome. Finally, we will summarize the current challenges in the field and point to future research directions required in order to successfully and effectively apply these approaches for precision medicine.

Keywords: cardiovascular disease, multi-omics, genomics, transcriptomics, metabolomics, gut microbiome

INTRODUCTION

In the current era of high-potency statin therapy it becomes increasingly clear that even individuals with normal LDL-cholesterol levels without any conventional risk factors may develop atherosclerosis (1). The most pertinent manifestation of atherosclerosis is coronary artery disease (CAD), a highly complex disease, influenced by both multiple genetic risk variants and lifetime exposure to an atherogenic environment (2). A better understanding of the etiology of CAD and directions toward hitherto therapeutically not addressed disease mechanisms are urgently demanded (3). During the last 10 years, the genetic risk has been thoroughly explored in numerous genome-wide association studies (GWAS), leading to identification of >300 chromosomal loci which all significantly affect the risk of CAD (4–15). More than 90% of these common disease

risk variants are located outside the protein-coding regions and have modest effect sizes (2, 16). Collectively they explain only ~25% of the overall disease heritability. This suggests that genetic variation may contribute to disease risk in a non-linear, interactive and complex way (17), leading to pathological changes in diverse cell types, tissues, and organs, at multiple molecular levels (18).

Recent technological advances have exponentially expanded the breadth of available -omics data (17). High-throughput monitoring of the abundance of various biological molecules and determination of their variation between different conditions on a global scale has become possible, promoting a paradigm shift in the way we approach biomedical problems (19). At the same time, it has been increasingly recognized that no single type of data can fully capture the intricacy of most complex molecular traits that manifest collectively as disease phenotypes (20–22). Rather, it is the integration of multiple layers of information across several -omics domains, i.e., the so-called multi-omics approach [also referred to as integromics or panomics (19)], that holds the promise for precision medicine (**Figure 1**) (19).

Of note, integrative analysis across multiple-omics layers can be conducted in two ways (**Figure 2**): pair-wise data integration and multi-dimensional i.e., network-based integration (22). Furthermore, pair-wise integrations can be divided into genetic and non-genetic correlations (22). In the first case, DNA variants (i.e., allelic distributions of single-nucleotide polymorphisms; SNPs) are tested for association with down-stream omics markers such as transcriptomic alterations, protein, metabolite or methylation levels or quantitative and qualitative measures of microbiome, via the so called quantitative trait loci (QTL) mapping. In the second scenario, one would explore correlations between down-stream omics data, e.g., correlation of CpG methylation levels to transcript expression or between metabolome and gut microbiome, however it may be difficult to infer causal relationships in such case (22). Considering the largely unexplored role of the established CAD risk loci from GWAS (23) and the central dogma that genetic variations control the transcriptome, which in turn affects e.g., the proteome (20), and metabolome (**Figure 2,** middle panel), our main focus will be pair-wise integrations linking genetic variation related to CAD risk to other down-stream omics layers such as epigenome, transcriptome, proteome or metabolome. Although multi-dimensional integrations have been widely used in the field of cancer research, their application in the context of CAD has so far been limited (22). Moreover, in addition to the autosomal genome, we will also consider the genetic variation in our "second genome"—the mitochondrial genome and its contribution to CAD.

INTEGRATING GENETIC VARIATION AND EPIGENOME

Epigenomic signatures reflect various DNA modifications and may affect gene regulatory mechanisms that do not involve changes in the DNA sequence *per se*. Thereby, epigenomics may become a critical mediator of environmental influences

and risk factors acting on the genome (20, 24). Three unique, but highly interrelated, epigenetic processes can be distinguished: DNA methylation, histone modifications (e.g., methylation, acetylation, phosphorylation, DP-ribosylation, and ubiquitination) and RNA-based mechanisms (e.g., microRNAs, long non-coding RNAs or lncRNAs, small interfering RNAs) (20, 24). Although, technically non-coding RNAs belong to the epigenome (20), we will discuss them in the next section, as the respective omics data are acquired via transcriptome profiling (RNA-seq).

DNA methylation and histone modifications are the best understood of the epigenetic mechanisms thus far and have been widely suggested to regulate gene expression and affect CAD risk factors including atherosclerosis, inflammation, hypertension and diabetes (25). DNA methylation consists of the covalent methylation of the C5 position of cytosine residues, when they are followed by guanine residues (CpG dinucleotides). It is partly heritable but it is also a dynamic process related to environmental stimuli and life style factors (26). Hedman et al. (27) analyzed epigenetic changes associated with lipid concentrations and identified a number of meQTLs, enriched in signals from GWAS on lipid levels and CAD. For example, genome-wide significant variants (rs563290 and its proxies), associated with LDL cholesterol and CAD at APOB, were meQTLs for a LDL cholesterol-related differentially methylated locus (**Table 1** and **Figure 3**).

Furthermore, the CDH13 (T-cadherin) locus may present an interesting example in the context of epigenetics and CAD. Putku et al. (39) reported several genetic variants in the promoter of CDH13 as meQTLs in hypertension patients (**Table 1**), several of them being also associated with high molecular weight adiponectin, a known ligand for CDH13, the binding of which results in increased proliferation and migration of endothelial cells (39). Moreover, recently Nelson et al. (13) identified a genetic variant in the intron of CDH13, which affects expression of this gene in vascular tissues, and is genome-wide significantly associated with CAD (28) (**Table 1**). Interestingly, the expression levels of CDH13 and lncRNAs from the same locus showed positive correlations, suggesting a functional link, as lncRNAs are known to display correlations with the expression of their neighboring protein-coding target genes (48).

An exciting field of future research will be studies conducting parallel profiling of genetic variation with histone modifications and Hi-C and ChIA-PET-based chromatin contact maps to uncover local and distal histone quantitative trait loci (hQTLs) (49) in CAD patients.

Overall, considering the critical role of epigenetic modifications as a critical mediator of environmental influences on the genome (20, 24), we urgently need more investigations studying DNA methylation and other epigenetic modifications genome-wide and in large enough cohorts, ideally also elucidating the differences between tissues and cells in healthy vs. CAD patients. Moreover, this should be supplemented with careful documentation of multiple environmental and lifestyle factors over time, i.e., the envirome, as well as comprehensive clinical information to draw a link between the environment and CAD.

FIGURE 1 | Multi-omics approach for precision medicine. Multi-omics (i.e., genome, epigenome, transcriptome, proteome, metabolome, microbiome, and envirome) data are collected from patients and integrated to create their individual molecular signatures (i.e., complex biomarkers), which are then used to select an appropriate drug for a particular patient, thus improving the treatment efficiency and reducing the possible side effects.

FIGURE 2 | Multi-omics (i.e., autosomal and mitochondrial genome, epigenome, transcriptome, proteome, metabolome, microbiome, and envirome) data integration can be conducted in two ways: pair-wise integrations, which can be further divided into non-genetic (left panel) and genetic correlations (middle panel). In the first case, one would examine the correlation patterns between the down-stream omics layers (e.g., metabolome and gut microbiome), whereas the second is achieved via the so called quantitative trait loci (QTL) mapping, linking genetic variation to methylation levels (meQTLs) or histone modifications (hQTLs), transcriptome (expression QTLs; eQTLs), protein (pQTLs), metabolite (mQTLs) or measures of microbiome (mbQTLs). Alternatively, multi-dimensional i.e., network-based integration approaches (right panel) exist, however their application in the context of CAD has so far been limited (22).

INTEGRATING GENETIC VARIATION AND TRANSCRIPTOME

Transcriptomics reflect genome-wide measures of RNA levels, both protein-coding RNA as well as the non-coding RNAs (i.e., microRNAs, lncRNAs, and small interfering RNAs) under specific conditions or in a specific cell. Moreover, the transcript levels are examined both qualitatively (i.e., which transcripts are present, identification of novel transcripts, splice sites, and RNA editing sites) and quantitatively (quantification of transcript abundance) (21).

Protein-Coding RNAs

Parallel assessments of genetic variation and transcriptome profiles across disease-relevant tissues, i.e., via mapping expression quantitative trait loci (eQTLs) to identify

susceptibility genes (mainly protein-coding), has been the most commonly applied approach (28, 29, 50–52). Björkegren et al. have performed a number of integrative network analysis, linking CAD risk variants and transcriptome data in seven disease-relevant vascular and metabolic tissues, collected from up to 600 CAD patients during coronary artery bypass surgery (28, 29, 53, 54). From these investigations, visceral abdominal fat has emerged as an important gene-regulatory site for blood lipids. Several risk SNPs for HDL-, LDL-, and total cholesterol levels, as well as for CAD demonstrated significant eQTL effects in visceral abdominal fat (28, 29).

Huan et al. (30) also used integrative analysis to investigate the molecular mechanisms of blood pressure regulation and identified a blood pressure associated SNP (rs3184504) in SH2B3, also associated with the expression (eQTL) of several

TABLE 1 | Genetic variation related to CAD/MI risk that has been associated with other down-stream omics layers such as transcriptome (mRNA, microRNAs and lncRNAs), epigenome, proteome or metabolome.

Data type	Tissue	Phenotypic Trait	SNP	Omics-Marker	Refernces
Transcriptome: mRNAs	Visceral abdominal fat	HDL cholesterol level	rs4148008	ABCAB/ABCA5	Franzén et al. (28)
			rs11869286	STARD3	
		Total cholesterol level	rs751557	EVI5	
			rs174546	TMEM258	
		LDL cholesterol level	rs12046679	PCSK9	
		CAD	rs892006	G3BP1	Foroughi Asl et al. (29)
			rs6908994	PSORSIC3	
			rs9930148	FLYWCH1	
	Internal mammary artery, atherosclerotic aortic root		rs7500448	CDH13	Nelson et al. (13)
	Blood	Blood pressure	Rs3184504	SH2B3,ALDH2,NAA25 (cis) and IL8,TAGAP (trans)	Huan et al. (30)
Transcriptome: microRNAs	Circulating leukocytes, human coronary artery smooth muscle cells (HCASMC)	CAD	rs12190287	miR-224: TCF21	Miller et al. (31) Bastami et al. (32)
		The effect of diet on plasma lipid levels	rs13702	miR-410:LPL	Richardson et al. (33)
		CAD	rs989727(rs7808424)	miR-202-5p:ASZ1	Bastami et al. (34)
			rs41269915(rs2229238)	miR-485-5p:UBE2Q1	
			rs15563	hsa-miR-130a-5p:UBE2Z	Brænne et al. (16)
			rs3088442	hsa-miR-130a-5p:SLC22A3	
			rs2266788	hsa-miR-4722-5p:APOA5	
			rs72932707	hsa-miR-4722-5p:ICA1L	
		HDL,LDL, and total cholesterol,triglycerides	rs2370747(rs7115089	miR-100-5p,miR-125b-5p	Huan et al. (35)
		CAD	rs11042699 (rs6578985)	miR-483-3p-IGF2	
		Platelet count	rs4905998 rs(7149242)	miR-127-3P, miR-134, miR-370, miR-376a-3p, miR-382-5p, miR-431-5p, miR-433, miR-329, miR-409-3p, miR-494, miR-411-3p, miR-654-5p, miR-668, miR-543, miR-323a-3p, miR-337-3p	
		3p/5p ratio	rs13064131	miR-28:LPP	Civelek et al. (36)
Transcriptome:lnc RNAs	Internal mammary artery, atherosclerotic aortic root	CAD	rs1333045	FPKM1_group_33469_transcript_1	Ballantyne et al. (37)
		MI	rs1333049	FPKM1_group_33469_transcript_2	
		T2D	rs2383208	FPKM1_group_33469_transcript_6	
		Early MI	rs10757274	ANRIL	McPherson et al. (38)
Epigenome:DNA methylation		Hypertension	rs113460564, rs12443878, rs12444338, rs62040565, rs8060301	CDH13	Putku et al. (39)
		Diastolic blood pressure,serum high-density lipoprotein,high molecular weight adiponectin	rs8060301	Cg09415485(CDH13)	

(Continued)

TABLE 1 | Continued

Data type	Tissue	Phenotypic Trait	SNP	Omics-Marker	Refernces
		High molecular weight adiponectin	rs2239857, rs77068073	CDH13	
		Smoking(no assosciation)	rs75509302	cg23576855(AHRR)	
		LDL cholesterol level	rs563290 (rs515135, rs562338)	Cg05337441(APOB)	Hedman et al. (27)
Proteome	Blood plasma	CAD	rs12740374	Granulin(CELSR2/SORT1)	Chen et al. (40)
			rs867186	Protein C (PROCR)	Howson et al. (14)
			rs1050362	apolipoprotein L1 (DHX38)	
Metabolome	Blood plasma	CAD	rs715	CPS1,urea cycle metabolites,plasma glycine	Hartiala et al. (41)
	Blood plasma		rs10450989(USP3), rs2228513(HER-C1)rs930491,rs11827377(STIM1), rs3853422(SEL1L), rs1869075(F-BXO25), rs9591507, rs17573278, rs894840, rs9285184(SUGT1)	Circulating short-chain di-carboxylacylcarnitine(SC-DA)	Kraus et al. (42)
Multi-OMICS	Low HDL and inflammatory pathways		rs241437	TAP2	Laurila et al. (43)
			rs9272143	HLA-DRB1,HLA-DQA1	
Mitochondrial Genome	Blood	Hypertension	m.8701A>G	MT-ATP6	Zhu et al. (44)
		CAD	Haplogroup T		Kofler et al. (45)
			m.16189T>C		Mueller et al. (46)
			m.15927G>A		Jia et al. (47)

genes, including SH2B3, in the genetically inferred causal blood pressure gene sets (**Table 1** and **Figure 4**). Some of these genes were also perturbed in Sh2b3$^{-/-}$ mice, demonstrating blood pressure-related phenotype (30). Rs3184504 has been previously also associated with CAD risk (9).

Much less investigated are non-coding RNA transcripts, such as micro-RNAs (miRNAs) and long non-coding RNAs (lncRNAs). Recent evidence suggests that at least some of these may play a role in CAD (55–58). Although, technically non-coding RNAs belong to the epigenome (20), we will discuss them in this section, as the respective omics data are acquired via transcriptome profiling (RNA-seq).

Micro RNAs

MiRNAs are involved in the transcriptional control of all main cell types participating in atherosclerosis progression, including endothelial cells, vascular smooth muscle cells, and macrophages (32, 59). Several studies have investigated the differential expression patterns of miRNAs in plasma/serum, microparticles, whole blood, platelets, blood mononuclear intimal, and endothelial progenitor cells in CAD vs. non-CAD patients, as summarized by Malik et al. (60). In majority of cases, up-regulation of different miRNA in CAD patients was observed (60). Moreover, growing body of evidence suggests that genetic variations in the miRNA targetome may lead to major

deleterious outcomes (61, 62). For example, Miller et al. (31) have shown that an established CAD risk variant (rs12190287) resides in the 3′ untranslated region of a transcription factor TCF21 and alters the seed binding sequence for miR-224. Moreover, allelic imbalance studies in circulating leukocytes and human coronary artery smooth muscle cells have demonstrated a significant imbalance of the TCF21 transcript levels, which correlated with genotype at rs12190287, consistent with this variant contributing to allele-specific expression differences (31). Richardson et al. (33) have reported that a variant (rs13702) in the 3′-UTR of lipoprotein lipase (LPL) disrupts the binding of miR-410 and modulates the effect of diet on plasma lipid levels (33). Recently, Bastami et al. (34) performed a more systematic computational screening, by mapping the established CAD risk variants to the miRNA targetome, identifying several links between SNPs and miRNAs (**Table 1**; https://www.ebi.ac.uk/gwas/). In a recent study from our group (16), we also mapped CAD risk variants from the CARDIoGRAMplusC4D GWAS meta-analyses (9), to 3′ UTR regions of genes to assess their overlaps with predicted target miRNA binding sites. Interestingly, the 3′ UTR region of MRAS was predicted to be targeted by 29 miRNAs and 23 miRNAs were predicted to bind more than one candidate CAD gene (**Table 1**). Thus far, there have been relatively few studies investigating genome-wide miRNA eQTLs (miR-eQTLs). Huan et al. (35) identified a

FIGURE 3 | Hedman et al. (27) identified SNP (rs515135) in an intron of APOB to be associated with LDL-C. Its proxy was also associated with CAD. Interestingly, this SNP represents a cis-meQTL. Black arrows indicate association findings. Red arrows indicate the presumed functional cascade leading to CAD.

FIGURE 4 | Huan et al. (30) uncovered a blood pressure associated SNP (rs3184504) in SH2B3, which also associates with the expression (eQTL) of several genes, including SH2B3 itself, in the genetically inferred causal blood pressure gene sets. Rs3184504 has been previously also associated with CAD risk. (9) Black arrows indicate association findings. Red arrows indicate the presumed functional cascade leading to CAD.

genetic variant (rs2370747) associated with miR-100-5p and miR-125b-5p expression, a proxy SNP of which was also associated with lipid traits (HDL-, LDL-, and total cholesterol as well as triglycerides). Moreover, it was found that both miRNAs were also differentially expressed in relation to HDL cholesterol (35).

FIGURE 5 | Civelek et al. (36) demonstrated a significant association of the SNP rs13064131 with the 3p/5p ratio of miR-28, encoded from the LPP gene. The miRNA processing and strand selection was adapted from (65).
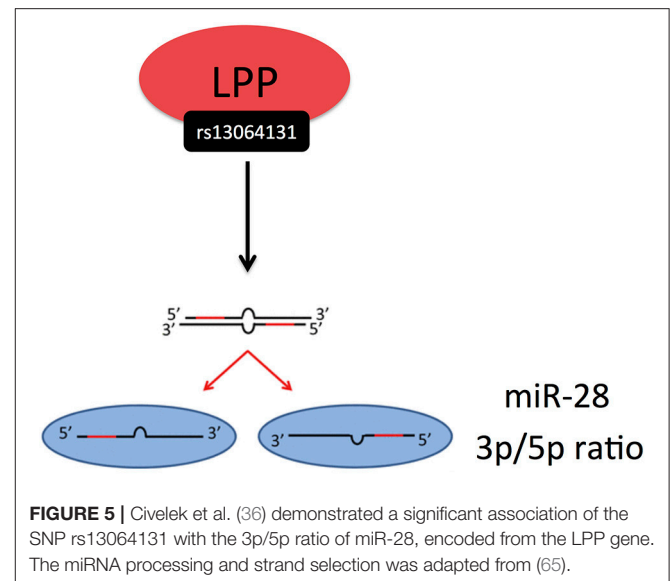
Civelek et al. (36) examined the genetic regulation of human adipose miRNA expression and its consequences for metabolic traits. Interestingly, this study showed, how genetic variation might influence the processing of miRNAs, i.e., the ratio of miRNA expression from the 3p and 5p arms. It is known that a miRNA precursor can give rise to two mature miRNAs from the 3p and 5p arm, one of which usually having higher expression than the other. The 3p/5p ratios of several miRNAs have been shown to be significantly different among various healthy tissues (63) and altered in pathological conditions compared with healthy controls (64). Civelek et al. demonstrated a significant association of the SNP rs13064131 with the 3p/5p ratio of miR-28, encoded from the LPP gene (**Figure 5**) (36). However, the SNP was not associated with the expression levels of the LPP transcript itself or with the abundance of miR-28-3p or miR-28-5p, suggesting that its effect on the 3p/5p ratio may be independent of transcription, possibly via degradation or stabilization mechanisms.

Long Non-Coding RNAs

The recent discovery of an extensive catalog of lncRNAs—i.e., long RNA transcripts that do not code for proteins—has opened a new perspective on the importance of the RNA-based mechanisms in gene regulation (24). LncRNAs are emerging as important regulators of various cellular processes, with many possible implications in cardiovascular disease pathophysiology (57, 58). In fact, the most prominent CAD risk locus at Chr9p21 (66, 67) harbors the lncRNA—ANRIL (Antisense Non-coding RNA in the INK4 Locus, CDKN2B antisense RNA). From these, rs10757274 is the strongest genetic predictor of early MI and is not associated with established CAD risk factors such as lipoproteins or hypertension, making ANRIL a key candidate (38). Interestingly, ANRIL is found both as a linear lncRNA (linANRIL), the transcript levels of which are known to positively correlate with disease severity (68), and is also capable of forming RNA circles (circANRIL)

FIGURE 6 | Recently, Holdt et al. (69) demonstrated that circANRIL regulates the maturation of precursor ribosomal RNA (pre-rRNA), by this impairing ribosome biogenesis and inducing nucleolar stress and apoptosis in vascular smooth muscle cells and macrophages. Moreover, carriers of the CAD-protective haplotype at 9p21 showed significantly increased expression of circANRIL.

(69). Recently, Holdt et al. (69) demonstrated that circANRIL regulates the maturation of precursor ribosomal RNA (pre-rRNA), by this impairing ribosome biogenesis and inducing nucleolar stress and apoptosis in vascular smooth muscle cells and macrophages (**Figure 6**). Carriers of the CAD-protective haplotype at 9p21 showed significantly increased expression of circANRIL (69).

Currently, there have not been many large-scale studies on lncRNAs in the context of CAD, though. Ballantyne et al. (37) recently conducted a genome-wide interrogation of long intergenic non-coding RNAs (lincRNAs) that associate with cardiometabolic traits in GWAS, including CAD and also identified a number of CAD/MI and type 2 diabetes associated SNPs at Chr9p21 that overlapped lincRNA transcripts (**Table 1**) (37). In STARNET (28), 5.4% of the identified cis-expression quantitative trait loci (eQTLs) were related to the expression of lncRNAs, however these have not been further explored, so far. Overall, more studies focusing on non-coding RNAs in different CAD relevant tissues in large enough cohorts will be required to yield insights into the possible functional roles of this portion of transcriptome and its genetic determinants, in healthy and disease states. Moreover, considering that lncRNAs are generally found to be more lowly-expressed, sufficient depth

of coverage for RNA-seq experiments will need to be guaranteed (21).

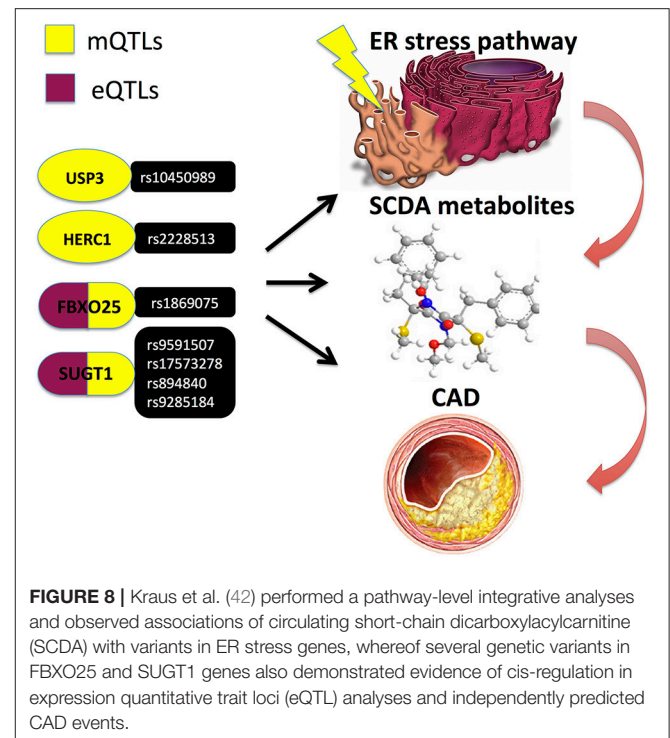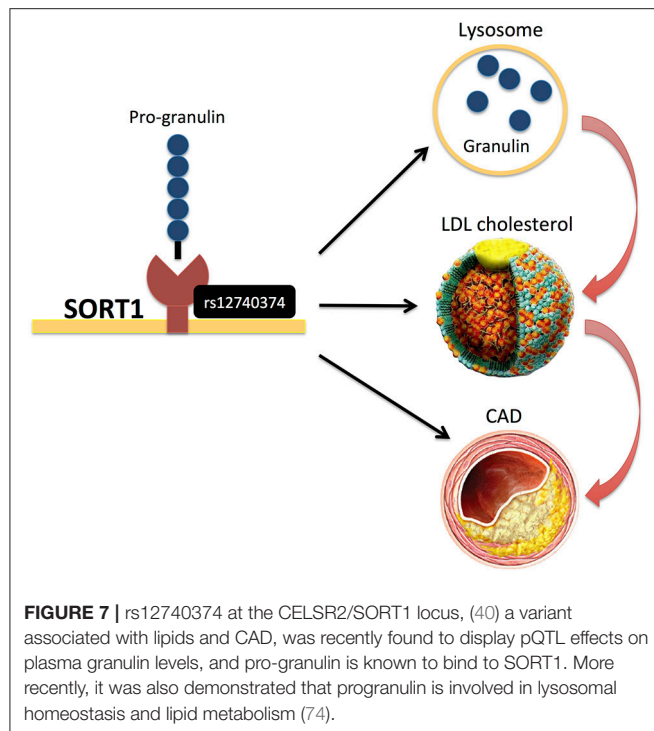## INTEGRATING GENETIC VARIATION AND PROTEOME

Proteomics uses high-throughput approaches (mainly MS-based) to quantify protein abundance, post-translational modifications and interactions (e.g., using phage display and yeast two-hybrid assays) in a tissue, cell or fluid compartment, such as plasma or urine (21). Considering that the transcriptome is not linearly proportional to proteome, that proteins are the biomolecules that execute cellular functions, and that many human diseases ultimately result from alterations in the proteome (70), such studies are urgently needed to facilitate the explorations of CAD etiology. However, proteome studies are still rare in relation to CAD, mostly due to the complex methodology involved. There have been some investigations in the past few years, aiming at characterizing the proteomes of several CAD-related tissues and cell types, including human arterial smooth muscle cells (71), platelets (72), as well as body fluids such as urine (73).

Only few studies (14, 40) have analyzed genetic variants that modify protein levels, i.e., the so-called protein quantitative trait loci (pQTLs) (**Table 1**). Chen et al. (40) assayed a pre-selected set of plasma proteins, identifying several pQTLs that overlapped with CAD risk SNPs and also explained a substantial proportion of inter-individual variation in protein abundance. For example, rs12740374 at the CELSR2/SORT1 locus, a variant associated with lipids and CAD, explained 15% of inter-individual variation in plasma granulin levels (**Figure 7**). Interestingly, progranulin binds to SORT1 and *Sort1* knockout mice show markedly elevated levels of progranulin (40). Recently, it was also demonstrated that progranulin is involved in lysosomal homeostasis and lipid metabolism (74).

As the proteomics technologies improve over time (21), more genome-wide investigations of CAD-related alterations in proteome and also phosphorpoteome in increasing numbers of disease relevant tissues are expected to be conducted in the near future. However, as proteins are more sensitive to their environment (21), caution will have to be taken during sample preparation steps to obtain accurate and reproducible results.

INTEGRATING GENETIC VARIATION AND METABOLOME

An important additional functional layer in mutli-omics data integration is the metabolome, as it represents an integrated state of all genetic, epigenetic and environmental factors, thus providing a link between genotype and phenotype (75). Metabolomics is an omics field that systematically identifies and quantifies multiple small molecule (typically <1,500 Daltons) types, such as amino acids, fatty acids, carbohydrates and biochemical intermediates, i.e., metabolites (21). A plethora of metabolites in blood and urine have been associated with CAD and subsequent cardiovascular events (76–79) and have been

FIGURE 7 | rs12740374 at the CELSR2/SORT1 locus, (40) a variant associated with lipids and CAD, was recently found to display pQTL effects on plasma granulin levels, and pro-granulin is known to bind to SORT1. More recently, it was also demonstrated that progranulin is involved in lysosomal homeostasis and lipid metabolism (74).

FIGURE 8 | Kraus et al. (42) performed a pathway-level integrative analyses and observed associations of circulating short-chain dicarboxylacylcarnitine (SCDA) with variants in ER stress genes, whereof several genetic variants in FBXO25 and SUGT1 genes also demonstrated evidence of cis-regulation in expression quantitative trait loci (eQTL) analyses and independently predicted CAD events.

demonstrated as promising biomarkers discriminating CAD vs. non-CAD subjects (78), as well as between thrombotic MI and stable CAD cases (80). Kraus et al. (42) recently identified several genetic loci demonstrating associations with blood plasma metabolites (i.e., metabolomic quantitative trait loci; mQTLs), the strongest findings being for the circulating short-chain dicarboxylacylcarnitine (SCDA) metabolite levels with variants in genes that regulate components of endoplasmic reticulum (ER) stress (**Table 1** and **Figure 8**) (42).

Besides blood and urine, metabolomic profiles of vascular and metabolomic tissues such as subcutaneous fat will need to be generated, ideally in conjunction with other omics layer data. Especially, gut microbiome would be of utmost interest, considering the close link between the two (81).

However, of note, metabolic profiles are even more prone to variability affected by sample preparation and storage conditions, as well as by several other factors including patient heterogeneity (21). Hence, the required sample size has to be carefully considered, to inspire confidence in the generated results.

INTEGRATING GENETIC VARIATION AND MICROBIOME

Microbiomics investigates all the microorganisms of a given community, including bacteria, viruses, and fungi, collectively known as the microbiota (and their genes constituting the microbiome) (21). The human microbiome is enormously complex and there are substantial variations in microbiota composition between individuals resulting from seed during birth and development, diet and other environmental factors,

drugs and age (21). Thousands of different bacterial species make up the human microbiomes, from which there is a small number of abundant species and a large number of rare or low abundance species, the differential functions of which remain poorly understood (82). Currently, several large scale initiatives are emerging including the American Gut Project http://americangut.org/ and the British Gut Project http://britishgut.org/, which are expected to produce a rich collection of anonymised human gut samples and lifestyle information for medical researchers.

Gut microbiome has emerged as another rich source of information and as a possible new player contributing to the CAD/MI pathogenesis (82–84). It has long been known that bacteria activate inflammatory pathways, and recent data demonstrate that the gut microbiome may also affect lipid metabolism and influences the development of obesity and atherosclerosis (84), suggesting that gut microbiota could be used as a diagnostic marker for CAD (85). The most investigated is the association between gut microbiota and fasting plasma trimethylamine N-oxide (TMAO) levels, a gut microbiota-dependent metabolite, previously also associated with CAD and stroke (81, 86). Org et al. (81) demonstrated that certain blood plasma metabolites strongly correlated with gut microbial community structure and that some of these correlations may be specific for the pre-diabetic state. LeChatelier et al. (84) used qunatitative gut microbiome information to distinguish between individuals with "high bacterial richness" and "low bacterial richness," were the latter were characterized by increased adiposity, insulin resistance and dyslipidemia in addition to a more pronounced inflammatory phenotype. Le Chatelier Fu et al.

(84) and Fu et al. (87) reported that gut microbiota richness and diversity were negatively correlated with triglycerides and positively correlated with HDL levels, however this effect was independent of age, sex and host genetics. So far, genome-wide mapping of the so-called microbiome quantitative trait loci (mbQTLs) (88) in the context of CAD has not been performed and is definitely next in line, ideally in conjunction with comprehensive profiling of metabolome in several tissues and body fluids in large enough cohorts.

INTEGRATING GENETIC VARIATION AND MULTIPLE OMICS DATASETS

An integrative analysis of genetic variation and transcriptome with additional high-throughput measurements may greatly improve the predictive power of disease networks. Zhu et al. (89) However, the number of studies conducting multi-omics integrations in the context of CAD is limited so far. Miller et al. (90) integrated genetic variation with investigations of chromatin state, enhancer activity and TF binding in human coronary artery smooth muscle cells and demonstrated, for example, that one of the lead candidate variants, rs17293632, located within an intergenic region of the SMAD3 gene, overlaps an open chromatin region. Moreover, it was observed that the major risk C allele was more associated with open chromatin and resided in a canonical AP-1 motif, which was effectively destroyed by the minor protective T allele. Preferential AP-1 binding to the risk C allele was experimentally validated using allele-specific ChIP analyses. Miller et al. (90) and Kraus et al. (42) performed a pathway-level integrative analyses, linking genetics, epigenetics, transcriptomics, and metabolomics profiles and implicating the ubiquitin proteasome system in cardiovascular disease pathogenesis. This study observed associations of circulating short-chain dicarboxylacylcarnitine (SCDA) with variants in ER stress genes, whereof several genetic variants (**Table 1** and **Figure 8**) in FBXO25 and SUGT1 genes also demonstrated evidence of cis-regulation in expression quantitative trait loci (eQTL) analyses and independently predicted CAD events (42). Moreover, two other genes from the same ER stress pathway—BRSK2 and HOOK2—were identified as differentially methylated, when comparing individuals with high and low SCDA levels. Subsequently, experimental validation using culture of human kidney cells in the presence of levels of fatty acids found in individuals with cardiometabolic disease, demonstrated induced accumulation of SCDA metabolites in parallel with increases in the ER stress marker BiP (42).

Shu et al. (20) investigated shared genetic regulatory networks for CAD and type 2 diabetes (T2D) and their key intervening drivers in multiple populations of diverse ethnicities by performing an integrative analysis of five multi-ethnic GWAS for CAD and T2D, eQTLs, ENCODE, as well as tissue-specific gene network models (both co-expression and graphical models) from disease-relevant tissues. This study identified pathways regulating the metabolism of lipids, glucose and branched-chain amino acids, as well as pathways governing oxidation, extracellular mat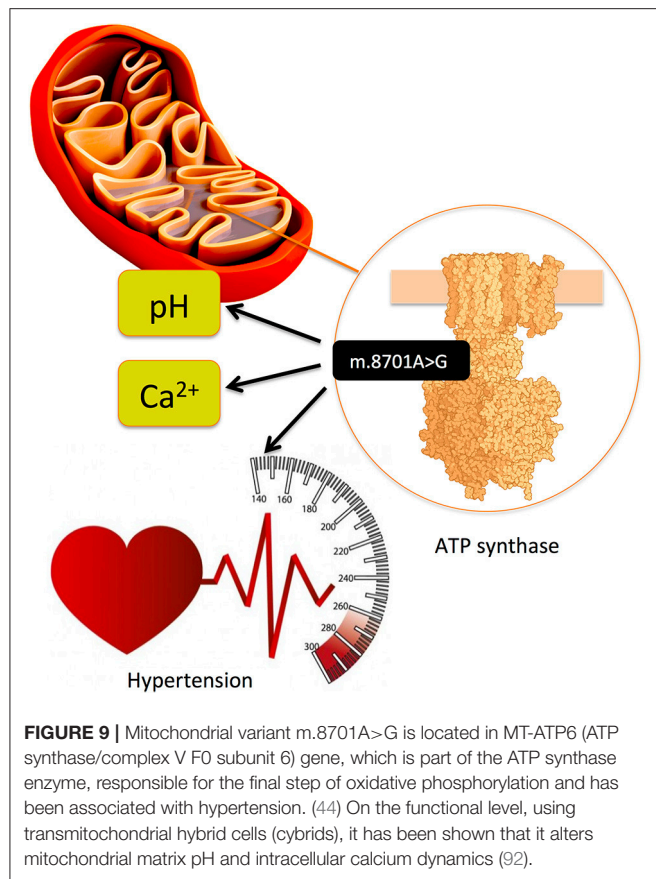rix and immune response as shared pathogenic processes for both diseases and identified 15 key drivers including HMGCR, CAV1, IGF1, and PCOLCE, whose network neighbors collectively accounted for ∼35% of known GWAS hits for CAD and 22% for T2D (20). Laurila et al. (43) applied a combined approach using both QTLs and canonical pathway analysis to link genomics and transcriptome analysis from the subcutaneous adipose tissue and plasma HDL lipidomics profiling, highlighting change in HDL particle quality toward putatively more inflammatory and less atheroprotective phenotype in subjects with low HDL, due to their reduced antioxidative capacity. Within the HLA region, this study found two significant, dose-dependent cis-eQTL associations with low HDL and inflammatory pathways: rs241437 in the intron of TAP2 and rs9272143 between HLA-DRB1 and HLA-DQA1, the latter also being associated with down-regulation of antioxidative pathways in HDL particles (43).

The application of multi-omics integrations in the field of CAD has so far been limited (22). Obviously, one of the main reasons for this is the current lack of appropriate data in large enough cohorts. However, considering the great promise such studies hold for precision medicine, it is expected that parallel measurements on multiple omics layers will be rapidly collected during the next couple of years, allowing also a comprehensive comparison, validation and improvement of the existing computational integration methods.

MITOCHONDRIAL GENETIC VARIATION AND DOWNSTREAM OMICS DATASETS

Dysfunction of mitochondria has been increasingly associated with obesity-related cardiometabolic diseases and CAD (91). Thus, genetic variation in the mitochondrial DNA (mtDNA), which codes for the 37 OXPHOS genes as well as further >1000 nuclear-coded genes imported into mitochondria constituting essential components for their proper functioning, needs exploration for a better understanding of CAD genetics. The mitochondrial haplogroup T (45) and mtDNA variants m.16189T>C (46) and m.15927G>A (47) have been associated with CAD in different ethnic groups. Another mitochondrial variant, m.8701A>G, has been associated with hypertension (44). This variant is located in MT-ATP6 (ATP synthase/complex V F0 subunit 6) gene, which is part of the ATP synthase enzyme, responsible for the final step of oxidative phosphorylation, and, on the functional level, using transmitochondrial hybrid cells (cybrids), it has been shown that it alters mitochondrial matrix pH and intracellular calcium dynamics (**Figure 9**) (92).

Similarly, other mitochondria-related omics data investigations could be of interest in the context of CAD, as Baccarelli et al. (93) reported that ATP synthesis genes including protein-encoding cytochrome c oxidase genes (MT-CO1, MT-CO2, and MT-CO3) and MT-TL1 were hypermethylated in platelets of CAD cases as compared to healthy controls (93). Using eQTLs in seven CAD relevant vascular and metabolic tissues (53) in conjunction with established CAD risk loci from GWAS (9) and time-resolved transcriptome data in the aortic arch in mice with reversible hypercholesterolemia (94, 95) we

FIGURE 9 | Mitochondrial variant m.8701A>G is located in MT-ATP6 (ATP synthase/complex V F0 subunit 6) gene, which is part of the ATP synthase enzyme, responsible for the final step of oxidative phosphorylation and has been associated with hypertension. (44) On the functional level, using transmitochondrial hybrid cells (cybrids), it has been shown that it alters mitochondrial matrix pH and intracellular calcium dynamics (92).

recently demonstrated a massive down-regulation of nuclear-encoded mitochondrial genes (96), specifically at the time of rapid atherosclerotic lesion expansion and foam cell formation, which was largely reversible by genetically lowering plasma cholesterol. Both mitochondrial signature genes were supported as causal for CAD in humans, as eQTLs representing their genes significantly overlapped with disease risk SNPs. In line with this, the STARNET (28) study recently examined mitochondrial (i.e., mtDNA-derived) gene expression and a markedly lower expression of mitochondrial genes in the atherosclerotic aortic arterial wall as compared to non-atherosclerotic arterial wall.

Furthermore, genetic variation of mitochondrial metabolome has remained largely unexplored. Hartiala et al. (41) searched for genetic factors associated with plasma betaine levels and determined their effect on CAD risk. This resulted in the identification of two significantly associated loci on chromosomes 2q34 and 5q14.1. The lead variant on 2q24—rs715—localized to carbamoyl-phosphate synthase 1 (CPS1), which encodes a mitochondrial enzyme that catalyzes the first committed reaction and rate-limiting step in the urea cycle. Rs715 was also significantly associated with decreased levels of urea cycle metabolites and increased plasma glycine levels. Finally, rs715 yielded a strikingly significant and protective association with decreased risk of CAD in women (41).

Finally, in recent years, it has become increasingly evident that the gut microbiome produces metabolites that influence

mitochondrial function and biogenesis (97), hence the ancestral gut microbiome-mitochondrion connection and its relation to CAD might need to be explored in the near future, as well.

Resent progress in next-generation sequencing (NGS) techniques has set a scene for a second "gold rush" in mitochondrial genomics and mtDNAs are presently the most sequenced type of eukaryotic chromosome (98). At the same time, multi-omics investigations in mitochondria, mapping the genomes, transcriptomes, proteomes, and metabolomes in parallel, apart from yeast (99) have not been conducted yet. Hence, although, mitochondrial dysfunction has been associated with many human diseases, the respective proteins and pathways are not well-characterized (99), presenting an exciting future field of investigation, especially considering the fact that mitochondria play a key role in plasticity and adaptation to environmental change, including adaptation to physiological stress (100).

CONCLUSIONS AND FUTURE DIRECTIONS

Given that CAD like other common complex disorders develops over time and involves both genetics and environment, full mechanistic insight will require coordinated sets of several-omics data at multiple time points, collected from many disease relevant tissues and body fluids in large enough cohorts (20, 21). Environmental risk factors can interact with the genome and perturb the epigenome to further modulate the transcriptome and proteome (20). Therefore, comprehensive monitoring and careful documentation of multiple environmental and lifestyle factors over time, i.e., the envirome, will be indispensable to yield significant insights into the complex etiology of CAD. Moreover, imaging and electronic health record data also will need to be considered. As more-omics and other data are generated, novel methods for efficient data integration, modeling, visualization and interpretation will be urgently needed to efficiently cope with this multi-dimensional data (101), and translate it into actionable precision medicine tools. Although, there has been major progresses in the development of multidimensional data integration algorithms and tools, the field is still in its infancy and the flexibility, effectiveness and robustness of data integration to extract biological insights is still restricted, especially when clinical outcomes (e.g., stable CAD vs. MI) need to be modeled (22, 101). In addition we still face a number of technical challenges related to patient sampling and profiling. For example, as already recognized by Hasin et al. and others (20, 21) human studies are often affected by various confounding factors, which are difficult or even impossible to control for (e.g., diet and medications). Clearly, also the available sample size will play an important role for the multi-omics approach to produce meaningful insights into CAD (21) and allow the generation of reliable prediction models for more efficient design of therapeutics, tailored to individual needs. According to Hasin et al. an underpowered study may not only miss true signals, but is also more likely to produce false positive results (21). Furthermore, already before and during data collection, careful

attention has to be paid to data analysis requirements, e.g., sufficient depth of coverage for RNA-seq experiments (21).

AUTHOR CONTRIBUTIONS

BV and HS drafted and edited the manuscript.

FUNDING

REFERENCES

1. Fernández-Friera L, Fuster V, López-Melgar B, Oliva B, García-Ruiz JM, Mendiguren J, et al. Normal LDL-cholesterol levels are associated with subclinical atherosclerosis in the absence of risk factors. *J Am Coll Cardiol.* (2017) 70:2979–91. doi: 10.1016/j.jacc.2017.10.024

2. Kessler T, Vilne B, Schunkert H. The impact of genome-wide association studies on the pathophysiology and therapy of cardiovascular disease. *EMBO Mol Med.* (2016) 8:688–701. doi: 10.15252/emmm.201506174

3. Ference BA, Ginsberg HN, Graham I, Ray KK, Packard CJ, Bruckert E, et al. Low-density lipoproteins cause atherosclerotic cardiovascular disease. 1. Evidence from genetic, epidemiologic, and clinical studies. A consensus statement from the European Atherosclerosis Society Consensus Panel. *Eur Heart J.* (2017) 38:2459–72. doi: 10.1093/eurheartj/ehx144

4. 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. *Nat Genet.* (2009) 41:280–2. doi: 10.1038/ng.307

5. Erdmann J, Willenborg C, Nahrstaedt J, Preuss M, König IR, Baumert J, et al. Genome-wide association study identifies a new locus for coronary artery disease on chromosome 10p11.23. *Eur Heart J.* (2011) 32:158–68. doi: 10.1093/eurheartj/ehq405

6. Schunkert H, König IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet.* (2011) 43:333–8. doi: 10.1038/ng.784

7. Charchar FJ, Bloomer LD, Barnes TA, Cowley MJ, Nelson CP, Wang Y, et al. Inheritance of coronary artery disease in men: an analysis of the role of the Y chromosome. *Lancet* (2012) 379:915–22. doi: 10.1016/S0140-6736(11)61453-0

8. Erdmann J, Stark K, Esslinger UB, Rumpf PM, Koesling D, de Wit C, et al. Dysfunctional nitric oxide signalling increases risk of myocardial infarction. *Nature* (2013) 504:432–6. doi: 10.1038/nature12722

9. CARDIoGRAMplusC4D Consortium, 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

10. 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

11. Myocardial Infarction Genetics and CARDIoGRAM Exome Consortia Investigators, Stitziel NO, Stirrups KE, Masca NG, Erdmann J, Ferrario PG, et al. Coding Variation in ANGPTL4, LPL, and SVEP1 and the risk of coronary disease. *N Eng J Med.* (2016) 374:1134–44. doi: 10.1056/NEJMoa1507652

12. Webb TR, Erdmann J, Stirrups KE, Stitziel NO, Masca NG, Jansen H, et al. Systematic evaluation of peiotropy identifies 6 further loci associated with coronary artery disease. *J Am Coll Cardiol.* (2017) 69:823–36. doi: 10.1016/j.jacc.2016.11.056

13. Nelson CP, Goel A, Butterworth AS, Kanoni S, Webb TR, Marouli E, et al. Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nat Genet.* (2017) 49:1385–91. doi: 10.1038/ng.3913

14. Howson J.MM, Zhao W, Barnes DR, Ho WK, Young R, Paul DS, et al. Fifteen new risk loci for coronary artery disease highlight arterial-wall-specific mechanisms. *Nat Genet.* (2017) 49:113–9. doi: 10.1038/ng.3874

15. van der Harst P, Verweij N. The identification of 64 novel genetic loci provides an expanded view on the genetic architecture of coronary artery disease. *Circ Res.* (2017) 122:433–43. doi: 10.1161/CIRCRESAHA.117.312086

16. Brænne I, Civelek M, Vilne B, Di Narzo A, Johnson AD, Zhao Y, et al. Prediction of causal candidate genes in coronary artery disease loci. *Arterioscler Thromb Vasc Biol.* (2015) 35:2207–17. doi: 10.1161/ATVBAHA.115.306108

17. Ritchie MD, Holzinger ER, Li R, Pendergrass SA, Kim D. Methods of integrating data to uncover genotype-phenotype interactions. *Nat Rev Genet.* (2015) 16:85–97. doi: 10.1038/nrg3868

18. Hartiala J, Schwartzman WS, Gabbay J, Ghazalpour A, Bennett BJ, Allayee H. The genetic architecture of coronary artery disease: current knowledge and future opportunities. *Curr Atheroscler Rep.* (2017) 19:6. doi: 10.1007/s11883-017-0641-6

19. Manzoni C, Kia DA, Vandrovcova J, Hardy J, Wood NW, Lewis PA, et al. Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences. *Brief Bioinformat.* (2016) 19:286–302. doi: 10.1093/bib/bbw114

20. Arneson D, Yang X, Shu L. Bioinformatics principles for deciphering cardiovascular diseases. *Encyclop Cardiovasc Res Med.* (2018) 1:273–92. doi: 10.1016/B978-0-12-801238-3.99576-X

21. Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. Genome Biology (2017) 18:83. doi: 10.1186/s13059-017-1215-1

22. Arneson D, Shu L, Tsai B, Barrere-Cain R, Sun C, Yang X. Multidimensional integrative genomics approaches to dissecting cardiovascular disease. *Front Cardiovasc Med.* (2017) 4:8. doi: 10.3389/fcvm.2017.00008

23. Erdmann J, Kessler T, Munoz Venegas L, Schunkert H. A decade of genome-wide association studies for coronary artery disease: the challenges ahead. *Cardiovasc Res.* (2018) 114:1241–57. doi: 10.1093/cvr/cvy084

24. Turgeon PJ, Sukumar AN, Marsden PA. Epigenetics of cardiovascular disease - a new "Beat" in coronary artery disease. *Med Epigenet.* (2014) 2:37–52. doi: 10.1159/000360766

25. Muka T, Koromani F, Portilla E, O'Connor A, Bramer WM, Troup J, et al. The role of epigenetic modifications in cardiovascular disease: a systematic review. *Int J Cardiol.* (2016) 212:174–83. doi: 10.1016/j.ijcard.2016.03.062

26. Fernández-Sanlés A, Sayols-Baixeras S, Subirana I, Degano IR, Elosua R. Association between DNA methylation and coronary heart disease or other atherosclerotic events: a systematic review. *Atherosclerosis* (2017) 263:325–33. doi: 10.1016/j.atherosclerosis.2017.05.022

27. Hedman ÅK, Mendelson MM, Marioni RE, Gustafsson S, Joehanes R, Irvin M.R. Epigenetic patterns in blood associated with lipid traits predict incident coronary heart disease events and are enriched for results

from genome-wide association studies. *Circulation* (2017) 10:e001487. doi: 10.1161/CIRCGENETICS.116.001487

28. Franzén O, Ermel R, Cohain A, Akers NK, Di Narzo A, Talukdar HA. et al. Cardiometabolic risk loci share downstream cis- and trans-gene regulation across tissues and diseases. *Science* (2016) 353:827–30. doi: 10.1126/science.aad6970

29. Foroughi Asl H, Talukdar HA, Kindt AS, Jain RK, Ermel R, Ruusalepp A, et al. Expression quantitative trait loci acting across multiple tissues are enriched in inherited risk for coronary artery disease. *Circ Cardiovasc Genet.* (2015) 8:305–15. doi: 10.1161/CIRCGENETICS.114.000640

30. Huan T, Meng Q, Saleh MA, Norlander AE, Joehanes R, Zhu J, et al. Integrative network analysis reveals molecular mechanisms of blood pressure regulation. *Mol Syst Biol.* (2015) 11:799. doi: 10.15252/msb.20145399

31. Miller CL, Haas U, Diaz R, Leeper NJ, Kundu RK, Patolla B, et al. Coronary heart disease-associated variation in TCF21 disrupts a Mir-224 binding site and miRNA-mediated regulation. *PLoS Genet.* (2014) 10:e1004263. doi: 10.1371/journal.pgen.1004263

32. Bastami M, Ghaderian SM, Omrani MD, Mirfakhraie R, Vakili H, Parsa SA, et al. MiRNA-related polymorphisms in MiR-146a and TCF21 are associated with increased susceptibility to coronary artery disease in an Iranian population. *Genet Test Mol Biomark.* (2016) 20:241–8. doi: 10.1089/gtmb.2015.0253

33. Richardson K, Nettleton JA, Rotllan N, Tanaka T, Smith CE, Lai CQ, et al. Gain-of-function lipoprotein lipase variant rs13702 modulates lipid traits through disruption of a microRNA-410 seed site. *Am J Hum Genet.* (2013) 92:5–14. doi: 10.1016/j.ajhg.2012.10.020

34. Bastami M, Nariman-Saleh-Fam Z, Saadatian Z, Nariman-Saleh-Fam L, Omrani MD, Ghaderian SMH, et al. The miRNA targetome of coronary artery disease is perturbed by functional polymorphisms identified and prioritized by in-depth bioinformatics analyses exploiting genome-wide association studies. *Gene* (2016) 594:74–81. doi: 10.1016/j.gene.2016.08.054

35. Huan T, Rong J, Liu C, Zhang X, Tanriverdi K, Joehanes R, et al. Genome-wide identification of microRNA expression quantitative trait loci. *Nat Commun.* (2015) 6:6601. doi: 10.1038/ncomms7601

36. Civelek M, Hagopian R, Pan C, Che N, Yang WP, Kayne PS, et al. Genetic regulation of human adipose microrna expression and its consequences for metabolic traits. *Hum Mol Genet.* (2013) 22:3023–37. doi: 10.1093/hmg/ddt159

37. Ballantyne RL, Zhang X, Nu-ez 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

38. 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

39. Putku M, Kals M, Inno R, Kasela S, Org R, KoŽich V. et al. CDH13 promoter SNPs with pleiotropic effect on cardiometabolic parameters represent methylation QTLs. *Hum Genet.* (2015) 134:291–303. doi: 10.1007/s00439-014-1521-6

40. Chen G, Yao C, Hwang S. J, Liu C, Song C, Huan T, et al. Abstract 18806: integrated proteomic analysis of cardiovascular disease reveals novel protein quantitative trait loci. *Circulation* (2016) 134:A18806.

41. Hartiala JA, Tang WH, Wang Z, Crow AL, Stewart AF, Roberts R, et al. Genome-wide association study and targeted metabolomics identifies sex-specific association of CPS1 with coronary artery disease. *Nat Commun.* (2016) 7:10558. doi: 10.1038/ncomms10558

42. Kraus WE, Muoio DM, Stevens R, Craig D, Bain JR, Grass E, et al. Metabolomic quantitative trait loci (mQTL) mapping implicates the ubiquitin proteasome system in cardiovascular disease pathogenesis. *PLoS Genet.* (2015) 11:e1005553. doi: 10.1371/journal.pgen.1005553

43. Laurila PP, Surakka I, Sarin AP, Yetukuri L, Hyötyläinen T, Söderlund S, et al. Genomic, transcriptomic, and lipidomic profiling highlights the role of inflammation in individuals with low high-density lipoprotein cholesterolsignificance. *Arteriosc Thromb Vasc Biol.* (2013) 33:847–57. doi: 10.1161/ATVBAHA.112.300733

44. Zhu Y, Gu X, Xu C. A mitochondrial DNA A8701G mutation associated with maternally inherited hypertension and dilated cardiomyopathy in a Chinese pedigree of a consanguineous marriage. *Chin Med J.* (2016) 129:259–66. doi: 10.4103/0366-6999.174491

45. Kofler B, Mueller EE, Eder W, Stanger O, Maier R, Weger M, et al. Mitochondrial DNA haplogroup T is associated with coronary artery disease and diabetic retinopathy: a case control study. *BMC Med Genet.* (2009) 10:35. doi: 10.1186/1471-2350-10-35

46. Mueller EE, Eder W, Ebner S, Schwaiger E, Santic D, Kreindl T, et al. The mitochondrial T16189C polymorphism is associated with coronary artery disease in Middle European populations. *PLoS ONE* (2011) 6:e16455. doi: 10.1371/journal.pone.0016455

47. Jia Z, Wang X, Qin Y, Xue L, Jiang P, Meng Y, et al. Coronary heart disease is associated with a mutation in mitochondrial tRNA. *Hum Mol Genet.* (2013) 22:4064–73. doi: 10.1093/hmg/ddt256

48. Kumar V, Westra HJ, Karjalainen J, Zhernakova DV, Esko T, Hrdlickova B. et al. Human disease-associated genetic variation impacts large intergenic non-coding RNA expression. *PLoS Genet.* (2013) 9:e1003201. doi: 10.1371/journal.pgen.1003201

49. Grubert F, Zaugg JB, Kasowski M, Ursu O, Spacek DV, Martin AR, et al. Genetic control of chromatin states in humans involves local and distal chromosomal interactions. *Cell* (2015) 162:1051–65. doi: 10.1016/j.cell.2015.07.048

50. Garnier S, Truong V, Brocheton J, Zeller T, Rovital M, Wild PS, et al. Genome-wide haplotype analysis of cis expression quantitative trait loci in monocytes. *PLoS Genet.* (2013) 9:e1003240. doi: 10.1371/journal.pgen.1003240

51. Rotival M, Zeller T, Wild PS, Maouche S, Szymczak S, Schillert A, et al. Integrating genome-wide genetic variations and monocyte expression data reveals trans-regulated gene modules in humans. *PLoS Genet.* (2011) 7:e1002367. doi: 10.1371/journal.pgen.1002367

52. Lempiäinen H, Brænne I, Michoel T, Tragante V, Vilne B, Webb TR. Network analysis of coronary artery disease risk genes elucidates disease mechanisms and druggable targets. *Sci Rep.* (2018) 8:3434. doi: 10.1038/s41598-018-20721-6

53. Hägg S, Skogsberg J, Lundström J, Noori P, Nilsson R, Zhong H, et al. Multi-organ expression profiling uncovers a gene module in coronary artery disease involving transendothelial migration of leukocytes and lim domain binding 2: the Stockholm Atherosclerosis Gene Expression (STAGE) Study. *PLoS Genet* (2009) 5:e1000754. doi: 10.1371/journal.pgen.1000754

54. Talukdar HA, Foroughi Asl H, Jain RK, Ermel R, Ruusalepp A, Franzén O, et al. Cross-tissue regulatory gene networks in coronary artery disease. *Cell Syst.* (2016) 2:196–208. doi: 10.1016/j.cels.2016.02.002

55. Small EM, Frost RJ, Olson EN. MicroRNAs add a new dimension to cardiovascular disease. *Circulation* (2010) 121:1022–32. doi: 10.1161/CIRCULATIONAHA.109.889048

56. Economou EK, Oikonomou E, Siasos G, Papageorgiou N, Tsalamandris S, Mourouzis K, et al. The role of microRNAs in coronary artery disease: from pathophysiology to diagnosis and treatment. *Atherosclerosis* (2015) 241:624–33. doi: 10.1016/j.atherosclerosis.2015.06.037

57. Archer K, Broskova Z, Bayoumi AS, Teoh JP, Davila A, Tang Y, et al. Long non-coding RNAs as master regulators in cardiovascular diseases. *Int J Mol Sci.* (2015) 16:23651–67. doi: 10.3390/ijms161023651

58. El Azzouzi H, Doevendans PA, Sluijter JP. Long non-coding RNAs in heart failure: an obvious lnc. *Ann Trans Med.* (2016) 4:182. doi: 10.21037/atm.2016.05.06

59. Madrigal-Matute J, Rotllan N, Aranda JF, Fernández-Hernando C. MicroRNAs and atherosclerosis. *Curr Atherosc Rep.* (2013) 15:322. doi: 10.1007/s11883-013-0322-z

60. Malik R, Mushtaque RS, Siddiqui UA, Younus A, Aziz MA, Humayun C, et al. Association between coronary artery disease and microRNA: literature review and clinical perspective. *Cureus* (2017) 23:e1188. doi: 10.7759/cureus.1188

61. Bulik-Sullivan B, Selitsky S, Sethupathy P. Prioritization of genetic variants in the microRNA regulome as functional candidates in genome-wide association studies. *Hum Mutat.* (2013) 34:1049–56. doi: 10.1002/humu.22337

62. Ghaedi H, Bastami M, Jahani MM, Alipoor B, Tabasinezhad M, Ghaderi O, et al. A bioinformatics approach to the identification of variants associated with Type 1 and Type 2 diabetes mellitus that reside in functionally validated miRNAs binding sites. *Biochem Genet.* (2016) 54:211–21. doi: 10.1007/s10528-016-9713-5

63. Cloonan N, Wani S, Xu Q, Gu J, Lea K, Heater S, et al. MicroRNAs and their isomiRs function cooperatively to target common biological pathways. *Genome Biol.* (2011) 12:R126. doi: 10.1186/gb-2011-12-12-r126

64. Li SC, Liao YL, Ho MR, Tsai KW, Lai CH, Lin WC. Mi-RNA arm selection and isomiR distribution in gastric cancer. *BMC Genomics* (2012) 13:S13. doi: 10.1186/1471-2164-13-S1-S13

65. Meijer HA, Smith EM, Bushell M. Regulation of miRNA strand selection: follow the leader? *Biochem Soc Trans.* (2014) 42:1135–40. doi: 10.1042/BST20140142

66. 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

67. Schunkert H, Götz A, Braund P, McGinnis R, Tregouet DA, Mangino M, et al. Repeated replication and a prospective meta-analysis of the association between chromosome 9p21.3 and coronary artery disease. *Circulation* (2008) 117:1675–84. doi: 10.1161/CIRCULATIONAHA.107.730614

68. Holdt LM, Beutner F, Scholz M, Gielen S, Gäbel G, Bergert H, et al. ANRIL expression is associated with atherosclerosis risk at chromosome 9p21. *Arterioscl Thromb Vasc Biol.* (2010) 30:620–7. doi: 10.1161/ATVBAHA.109.196832

69. 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

70. Langley SR, Dwyer J, Drozdov I, Yin X, Mayr M. Proteomics: from single molecules to biological pathways. *Cardiovasc Res.* (2012) 97:612–22. doi: 10.1093/cvr/cvs346

71. Dupont A, Pinet F. The proteome and secretome of human arterial smooth muscle cell. *Cardiovasc Proteom.* (2007) 357:225–33. doi: 10.1385/1-59745-214-9:225

72. Burkhart JM, Vaudel M, Gambaryan S, Radau S, Walter U, Martens L, et al. The first comprehensive and quantitative analysis of human platelet protein composition allows the comparative analysis of structural and functional pathways. *Blood* (2012) 120:73–82. doi: 10.1182/blood-2012-04-416594

73. Neisius U, Koeck T, Mischak H, Rossi SH, Olson E, Carty DM, et al. Urine proteomics in the diagnosis of stable angina. *BMC Cardiovasc Disord.* (2016) 16:70. doi: 10.1186/s12872-016-0246-y

74. Evers BM, Rodriguez-Navas C, Tesla RJ, Prange-Kiel J, Wasser CR, Yoo KS, et al. Lipidomic and transcriptomic basis of lysosomal dysfunction in progranulin deficiency. *Cell Rep.* (2017) 20:2565–74. doi: 10.1016/j.celrep.2017.08.056

75. Krumsiek J, Bartel J, Theis FJ. Computational approaches for systems metabolomics. *Curr Opin Biotechnol.* (2016) 39:198–206. doi: 10.1016/j.copbio.2016.04.009

76. Teupser D, Baber R, Ceglarek U, Scholz M, Illig T, Gieger C, et al. Genetic regulation of serum phytosterol levels and risk of coronary artery diseaseclinical perspective. *Circulation* (2010) 3:331–9. doi: 10.1161/CIRCGENETICS.109.907873

77. Shah SH, Sun JL, Stevens RD, Bain JR, Muehlbauer MJ, Pieper KS, et al. Baseline metabolomic profiles predict cardiovascular events in patients at risk for coronary artery disease. *Am Heart J.* (2012) 163:844–50. doi: 10.1016/j.ahj.2012.02.005

78. Krishnan S, Huang J, Lee H, Guerrero A, Berglund L, Anuurad E, et al. Combined high-density lipoprotein proteomic and glycomic profiles in patients at risk for coronary artery disease. *J Proteome Res.* (2015) 14:5109–18. doi: 10.1021/acs.jproteome.5b00730

79. Feng Q, Liu Z, Zhong S, Li R, Xia H, Jie Z, et al. Integrated metabolomics and metagenomics analysis of plasma and urine identified microbial metabolites associated with coronary heart disease. *Sci Rep.* (2016) 6:22525. doi: 10.1038/srep22525

80. Trainor PJ, Hill BG, Carlisle SM, Rouchka EC, Rai SN, Bhatnagar A, et al. Systems characterization of differential plasma metabolome perturbations following thrombotic and non-thrombotic myocardial infarction. *J Proteom.* (2017) 160:38–46. doi: 10.1016/j.jprot.2017.03.014

81. Org E, Blum Y, Kasela S, Mehrabian M, Kuusisto J, Kangas AJ, et al. Relationships between gut microbiota, plasma metabolites, and metabolic syndrome traits in the METSIM cohort. *Genome Biol.* (2017) 18:70. doi: 10.1186/s13059-017-1194-2

82. Ranjan R, Rani A, Metwally A, McGee HS, Perkins DL. Analysis of the microbiome: advantages of whole genome shotgun versus 16S amplicon sequencing. *Biochem Biophys Res Commun.* (2016) 469:967–77. doi: 10.1016/j.bbrc.2015.12.083

83. Caesar R, Fåk F, Bäckhed F. Effects of gut microbiota on obesity and atherosclerosis via modulation of inflammation and lipid metabolism. *J Int Med.* (2010) 268:320–8. doi: 10.1111/j.1365-2796.2010.02270.x

84. Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, Falony G, et al. Richness of human gut microbiome correlates with metabolic markers. *Nature* (2013) 500:541–6. doi: 10.1038/nature12506

85. Emoto T, Yamashita T, Kobayashi T, Sasaki N, Hirota Y, Hayashi T, et al. Characterization of gut microbiota profiles in coronary artery disease patients using data mining analysis of terminal restriction fragment length polymorphism: gut microbiota could be a diagnostic marker of coronary artery disease. *Heart Vessels* (2016) 32:39–46. doi: 10.1007/s00380-016-0841-y

86. Senthong V, Li XS, Hudec T, Coughlin J, Wu Y, Levison B, et al. Plasma trimethylamine N-Oxide, a gut microbe–generated phosphatidylcholine metabolite, is associated with atherosclerotic burden. *J Am Coll Cardiol.* (2016) 67:2620–8. doi: 10.1016/j.jacc.2016.03.546

87. Fu J, Bonder MJ, Cenit MC, Tigchelaar EF, Maatman A, Dekens JA, et al. The gut microbiome contributes to a substantial proportion of the variation in blood lipids. *Circ Res.* (2015) 117:817–24. doi: 10.1161/CIRCRESAHA.115.306807

88. Blekhman R, Goodrich JK, Huang K, Sun Q, Bukowski R, Bell JT, et al. Host genetic variation impacts microbiome composition across human body sites. *Genome Biol.* (2015) 16:191. doi: 10.1186/s13059-015-0759-1

89. Zhu J, Sova P, Xu Q, Dombek KM, Xu EY, Vu H, et al. Stitching together multiple data dimensions reveals interacting metabolomic and transcriptomic networks that modulate cell regulation. *PLoS Biol.* (2012) 10:e1001301. doi: 10.1371/journal.pbio.1001301

90. Miller CL, Pjanic M, Wang T, Nguyen T, Cohain A, Lee JD, et al. Integrative functional genomics identifies regulatory mechanisms at coronary artery disease loci. *Nat Commun.* (2016) 7:12092. doi: 10.1038/ncomms12092

91. Ballinger SW. Mitochondrial dysfunction in cardiovascular disease. *Free Radic Biol Med.* (2005) 38:1278–95. doi: 10.1016/j.freeradbiomed.2005.02.014

92. Kazuno AA, Munakata K, Nagai T, Shimozono S, Tanaka M, Yoneda M, et al. Identification of mitochondrial DNA polymorphisms that alter mitochondrial matrix pH and intracellular calcium dynamics. *PLoS Genet.* (2006) 2:e128. doi: 10.1371/journal.pgen.0020128

93. Baccarelli AA, Byun HM. Platelet mitochondrial dna methylation: a potential new marker of cardiovascular disease. *Clin Epigenet.* (2015) 7:44. doi: 10.1186/s13148-015-0078-0

94. Skogsberg J, Lundström J, Kovacs A, Nilsson R, Noori P, Maleki S, et al. Transcriptional profiling uncovers a network of cholesterol-responsive atherosclerosis target genes. *PLoS Genet* (2008) 4:e1000036. doi: 10.1371/journal.pgen.1000036

95. Björkegren JL, Hägg S, Talukdar HA, Foroughi Asl H, Jain RK, Cedergren C, et al. Plasma cholesterol-induced lesion networks activated before regression of early, mature, and advanced atherosclerosis. *PLoS Genet* (2014) 10:e1004201. doi: 10.1371/journal.pgen.1004201

96. Vilne B, Skogsberg J, Foroughi Asl H, Talukdar HA, Kessler T, Björkegren JLM, et al. Network analysis reveals a causal role of mitochondrial gene activity in atherosclerotic lesion formation. *Atherosclerosis* (2017) 267:39–48. doi: 10.1016/j.atherosclerosis.2017.10.019

97. Franco-Obregón A, Gilbert JA. The microbiome-mitochondrion connection: common ancestries, common mechanisms, common goals. *mSystems* (2017) 2:e00018–17. doi: 10.1128/mSystems.00018-17

98. Smith DR. The past, present and future of mitochondrial genomics: have we sequenced enough mtDNAs? *Brief Func Genom.* (2016) 15:47–54. doi: 10.1093/bfgp/elv027

99. Stefely JA, Kwiecien NW, Freiberger EC, Richards AL, Jochem A, Rush MJP, et al. Mitochondrial protein functions elucidated by multi-omic mass spectrometry profiling. *Nat Biotechnol.* (2016) 34:1191–7. doi: 10.1038/nbt.3683

100. Picard M, McEwen BS, Epel ES, Sandi C. An energetic view of stress: focus on mitochondria. *Front Neuroendocrinol.* (2018) 49:72–85. doi: 10.1016/j.yfrne.2018.01.001

101. Huang S, Chaudhary K, Garmire LX. More is better: recent progress in multi-omics data integration methods. *Front Genet.* (2017) 8:84. doi: 10.3389/fgene.2017.00084

Serum Biomarkers of Endothelial Dysfunction in Fabry Associated Cardiomyopathy

Jefferson Loso[1], Natalie Lund[1], Maxim Avanesov[2], Nicole Muschol[3], Susanne Lezius[4], Kathrin Cordts[5,6], Edzard Schwedhelm[5,6] and Monica Patten[1,6]*

[1] Department of General and Interventional Cardiology, University Heart Center Hamburg, Hamburg, Germany, [2] Department of Diagnostic and Interventional Radiology, University Medical Center Hamburg- Eppendorf, Hamburg, Germany, [3] Department of Pediatrics, University Medical Center Hamburg- Eppendorf, Hamburg, Germany, [4] Department of Medical Biometry and Epidemiology, University Medical Center Hamburg- Eppendorf, Hamburg, Germany, [5] Department of Experimental Pharmacology and Toxicology, University Medical Center Hamburg- Eppendorf, Hamburg, Germany, [6] DZHK (German Center for Cardiovascular Research e.V.), Hamburg, Germany

Background: Fabry disease (FD) is characterized by early development of vasculopathy and endothelial dysfunction. However, it is unclear whether these findings also play a pivotal role in cardiac manifestation. As Fabry cardiomyopathy (FC) is the leading cause of death in FD, we aimed to gather a better insight in pathological mechanisms of the disease.

Methods: Serum samples were obtained from 17 healthy controls, 15 FD patients with and 7 without FC. FC was defined by LV wall thickening of >12 mm in cardiac magnetic resonance imaging and serum level of proBNP, high sensitive Troponin T (hsT), and globotriaosylsphingosine (lyso-GB3) were obtained. A multiplex ELISA-Assay for 23 different angiogenesis markers was performed in pooled samples. Markers showing significant differences among groups were further analyzed in single samples using specific Elisa antibody assays. L-homoarginine (hArg), L-arginine, asymmetric (ADMA), and symmetric Dimethylarginine (SDMA) were quantified by liquid chromatography—mass spectrometry.

Results: Angiostatin and matrix metalloproteinase 9 (MMP-9) were elevated in FD patients compared to controls independently of the presence of FC (angiostatin: 98 ± 25 vs. 75 ± 15 ng/mL; $p = 0.001$; MMP-9: 8.0 ± 3.4 vs. 5.0 ± 2.4 µg/mL; $p = 0.002$). SDMA concentrations were highest in patients with FC (0.90 ± 0.64 µmol/l) compared to patients without (0.57 ± 0.10 µmol/l; $p = 0.027$) and vs. controls (0.58 ± 0.12 µmol/l; $p = 0.006$) and was positively correlated with indexed LV-mass ($r = 0.61$; $p = 0.003$), hsT ($r = 0.56$, $p = 0.008$), and lyso-Gb3 ($r = 0.53$, $p = 0.013$). Accordingly, the ratio of L-homoarginine to SDMA (hArg/SDMA) was lowest in patients with FC (2.63 ± 1.78) compared to controls (4.16 ± 1.44; $p = 0.005$). For L-arginine, hArg and ADMA no significant differences among groups could be detected, although a trend toward higher ADMA and lower hArg levels could be observed in the FC group. Furthermore, a significant relationship between kidney and cardiac function could be revealed ($p = 0.045$).

Conclusion: Elevated MMP-9 and angiostatin levels suggest an increased extracellular matrix turnover in FD patients. Furthermore, endothelial dysfunction may also be involved in FC, as SDMA and hArg/SDMA are altered in these patients.

Keywords: matrix metalloproteinase 9, angiostatins, SDMA, homoarginine, Fabry disease, Fabry cardiomyopathy, vasculopathy, endothelial dysfunction

INTRODUCTION

Fabry disease (FD) is an X-linked recessive multi-systemic storage disorder caused by a decreased activity of the lysosomal enzyme alpha-Galactosidase A (GLA) (1, 2). Due to the deposition of globotriaosylceramide (Gb3) in vascular lysosomes neutral glycosphingolipids accumulate in plasma and various tissues throughout the body (3). Typical manifestations of FD are cardiac, neurological, renal, ocular, dermatological, and gastrointestinal (4–6) with cardiovascular disease as the leading cause of death in FD patients (7). Prognosis of cardiomyopathy is particularly poor (8), thus, appropriate diagnosis and treatment of Fabry associated cardiomyopathy (FC) is crucial. Enzyme Replacement Therapy (ERT) has proven to significantly reduce accumulation of Gb3, especially intracellular deposits in the coronary endothelium (9, 10) and to halt or even partially reverse FC. However, in advanced stages of FD with a severe cardiac phenotype the effectiveness of ERT is profoundly diminished and the disease can even progress (11–13). Therefore, a better understanding of the underlying mechanisms contributing to the development of FC is urgently needed to improve treatment and outcome of FD patients.

Clinical studies provide evidence of increased intima-media thickness (IMT) and impaired artery flow-mediated dilatation in FD (14, 15) indicating an early onset of atherosclerosis in these patients. Moreover, different studies suggest that myocardial fibrosis, detected by cardiac magnetic resonance imaging (cMRI), may contribute to left ventricular remodeling in FD (16, 17). Myocardial fibrosis and subsequent remodeling are caused by an altered extracellular matrix turnover, which is catalyzed by Matrix Metalloproteinases (MMP) including MMP-9 (18). This is in line with the detection of increased serum MMP-9 level in Fabry patients compared to controls (19). Matsunaga et al showed that the inhibition of NO synthase resulted in increased MMP-9 and MMP-2 activities suggesting a link between oxidative stress and extracellular matrix turnover. Several clinical and experimental studies demonstrated inflammatory activity and endothelial nitric oxide synthase (eNOS) alterations in vascular cells of FD patients (20–23). These findings support the hypothesis of early occurring vasculopathy and endothelial dysfunction in FD. Whether these findings play a pivotal role in cardiac manifestation has not yet been sufficiently investigated. Accordingly, the aim of this study was to gather an insight in underlying pathological mechanisms by determining serum markers of endothelial dysfunction, angiogenesis and cardiac function in FD patients with and without FC.

METHODS

Study Population

Serum samples from 15 FD patients with FC, 7 without FC, and 17 healthy controls were collected between September 2014 and December 2016. FD was confirmed by molecular genetic analysis revealing the following mutations: 4x p.N215S, 2x p.E341K, 2x c.1277_1278delAA, 2x c.718_719delAA, p.N320l, p.A143T, p.A230_I232del, p.Q327L, p.A389V, c.717delAA, p.I384N, p.P205T, p.S247P, p.Q327L, p.R227Q. FC was defined by LV wall thickening of >12 mm assessed in cMRI. Furthermore, 11 FD patients with FC and one patient without FC received ERT at the time of blood sampling. The study was in line with the principles outlined in the Declaration of Helsinki and approved by the local ethics committee. All participants gave their written informed consent for participation in the study.

Enzyme-Linked Immunosorbent Assays

Blood samples were centrifuged at $4000 \times G$ for 10 min at room temperature and obtained serum was aliquoted and stored at $-80°C$ until use. For the multiplex Enzyme-Linked Immunosorbent Assay (ELISA) a Human Angiogenesis Antibody Array Membrane (Abcam PLC, Cambridge, UK, ab169808) was used. Aliquots of serum samples were pooled into three groups: FD with FC ($n = 15$), FD without FC ($n = 7$), and controls ($n = 17$). The assay allowed a simultaneous and semi-quantitative analysis of 23 targets (Angiopoietin 1, Angiopoietin 2, Angiostatin, Endostatin, G-CSF, GM-CSF, I-309, IL-10, IL-1 alpha, IL-1 beta, IL-2, IL-4, I-TAC, MCP-3, MCP-4, MMP-1, MMP-9, PECAM-1, Tie-2, TNF alpha, suPAR, VEGFR2, VEGFR3) and was performed in duplicates. Visualization of membrane signals was performed by ChemiDoc™ MP Imaging System and the densitometry software Image Lab™.

Markers showing significant differences among groups in the multiplex Elisa-Assay were further analyzed in single samples. Specific ELISAs from Abcam were used for MMP-9, angiostatin, soluble urokinase-type plasminogen activator receptor (suPAR), and vascular endothelial growth factor (VEGF) quantification according to the manufacturers' instructions.

Liquid Chromatography—Tandem Mass Spectrometry (LC-MS/MS) Measurements

L-Arginine, L-homoarginine (hArg), asymmetric (ADMA), and symmetric Dimethylarginine (SDMA) were quantified as described previously (24, 25). In brief, 25 µL of EDTA plasma were diluted with 100 µL $^2H_7$-arginine, $^{13}C_7^{15}N_4$-hArg, and $^2H_7$-ADMA solved in methanol. Proteins were

precipitated and residues were derivatised to their butylester derivatives. Twenty microliter of reconstituted samples were injected into the 1200 L Triple Quadrupole MS/MS system (Agilent Technologies, Waldbronn, Germany) chromatography. Analytes were separated on a Polaris C18-Ether column (Agilent Technologies; 50 × 2.0 mm) using an elution gradient of the two mobile phases (A): 1 mL/L formic acid in water and (B) acetonitrile-methanol (50/50, vol/vol) containing 1 mL/L formic acid in water (0:00 min 95/5 (A/B) – 0:30 95/5 – 2:00 50/50 – 2:01 95/5 – 4:00 95/5). The flow rate was 0.3 mL/min. Peak area ratios were calculated with internal standards and external calibration curves prepared in dialysed EDTA plasma. Intra- and interassay coefficients of variation were below 15 % for all analyses.

Cardiac MRI

Cardiac MRI was performed with a 1.5 Tesla MRI scanner (Achieva, Philips Medical Systems, Philips, Best, The Netherlands). The examination contained a retrospectively gated cine-MRI in cardiac short and long axis orientations using a steady-state free precession (SSFP) sequence to quantify regional and global left ventricular (LV) function and the LV-myocardial mass. Ten minutes after bolus injection of 0.075 mmol/kg Gd-BOPTA (MultiHance®), end-diastolic late gadolinium enhanced (LGE) images were acquired by phase-sensitive inversion recovery (PSIR) sequences to quantify areas of myocardial fibrosis. LGE images were obtained in the LV short-axis as well as in two-, three-, and four-chamber views and quantified using cvi42® software (Circle Cardiovascular Imaging Inc., Calgary, Alberta, Canada). All quantified parameters included LV function, end-diastolic and end-systolic volumes, stroke volume, ejection fraction, left ventricular mass, regional wall thickness, and LGE. Left ventricular mass was indexed to the body surface area calculated by DuBois & DuBois formula.

Statistical Analysis

Obtained data were analyzed using SPSS, version 23. Not normally distributed variables were log transformed if necessary. Assessment of group differences was performed by analysis of variance (ANOVA) or analysis of covariance (ANCOVA) when additionally adjusted for age, sex and eGFR. *Post-hoc* group comparisons were analyzed if the global differences were significant. Correlations were investigated by Pearson's correlation tests. Logistic regression was performed to investigate the relationship between renal and cardiac function. Concentrations are presented as mean ± standard deviation and a *p*-value of <0.05 was set as statistically significant.

RESULTS

Baseline Characteristics

Cardiac MRI measurements revealed significantly higher indexed left-ventricular masses of FD patients with FC compared to controls and FD patients without FC as shown in **Table 1**. Furthermore, similar results could be shown for septal thickness and mean LGE size, which were significantly thicker/higher in FD patients with FC. No differences among groups could be detected for left ventricular ejection fraction and stroke volume.

Laboratory values revealed significantly higher levels of proBNP, hsT, and lyso-Gb3 in the cardiomyopathy group. Moreover, patients with FC showed elevated renal function parameters.

Quantification of Angiogenesis Markers by Enzyme-Linked Immunosorbent Assays

The multiplex ELISA indicated differences among the pooled groups for angiostatin, MMP-9 and suPAR. In **Figure 1** the MMP-9 and angiostatin levels from the ensuing ELISA assays are demonstrated. MMP-9 concentrations were 1.51 times higher in FD patients with FC and 2.07 times higher in FD patients without FC compared to controls. The mean MMP-9 level of all FD patients was 1.67 times higher compared to controls (**Table 2**). Accordingly, angiostatin ELISA provided 1.33 times higher angiostatin levels in FD patients with FC and 1.23 times higher concentrations in FD patients without FC compared to controls. The mean angiostatin level of all FD patients was 1.3 times higher compared to controls (**Table 2**). However, no significant differences between the two FD groups could be detected for angiostatin and MMP-9. Moreover, neither suPAR nor VEGF concentrations revealed a significant difference between FD patients and healthy controls.

Quantification of Endothelial Dysfunction Markers by LC-MS/MS Measurements

LC-MS/MS measurements revealed 1.4 times higher SDMA concentrations in FD patients with vs. FD patients without FC and equally 1.4 times higher concentrations compared to controls, whereas no statistically significant difference could be shown between FD patients without FC and controls (**Table 2**, **Figure 2**). Accordingly, the ratio of L-homoarginine to SDMA (hArg/SDMA) was 0.53 times lower in FD patients with FC compared to controls. For L-arginine, hArg, ADMA and the ratio of hArg/ADMA no significant differences among groups could be detected, although a trend toward higher ADMA concentrations, lower hArg levels and accordingly a lower ratio of hArg/ADMA could be observed in the FD group with existing FC (**Table 2**).

Correlations of Biomarkers to Anthropometric and Clinical Phenotypes

No statistically relevant age or sex dependencies for MMP-9, angiostatin, SDMA, or hArg/SDMA could be observed. As shown in **Figure 3A**, SDMA was positively correlated with LV-mass, lyso-Gb3, and hsT and negatively correlated with eGFR. Accordingly, as shown in **Figure 3B**, hArg/SDMA showed a negative correlation to LV-mass, lyso-Gb3, and a positive correlation with eGFR. Moreover, correlations of indexed LV-mass with ADMA, hArg/ADMA, suPAR, lyso-Gb3, cardiac, and renal parameters could be revealed (**Table 3**). For MMP-9 and angiostatin no significant correlations to any of these variables could be observed (data not shown).

Relationship of SDMA and HArg/SDMA to Renal Function

Analysis of covariance with adjustment for eGFR diminished the significant group differences. Therefore, SDMA and hArg/SDMA

TABLE 1 | Baseline characteristics, cMRI measurements, and laboratory values.

	Control (1)	Fabry no FC (2)	Fabry FC (3)	p (ANOVA)	p (1 vs. 2)	p (2 vs. 3)	p (1 vs. 3)
n	17	7	15				
Age (years)	30.4 ± 6.3	36.3 ± 10.1	47.5 ± 13.0				
Male, n (%)	15 (88)	1 (14)	10 (67)				
ERT, n (%)	0 (0)	1 (14)	11 (73)				
MSSI	–	5.6 ± 3.2	25.7 ± 7.5				
cMRI-MEASUREMENTS							
LVEF (%)	61.1 ± 3.7	70.0 ± 8.8	68.0 ± 14.0	0.224	–	–	–
SV (ml/m^2)	114.2 ± 20.1	91.4 ± 13.0	100.8 ± 20.0	0.063	–	–	–
LVM indexed to BSA (g/m^2)	73.1 ± 8.2	73.3 ± 29.9	107.1 ± 34.7	**0.011**	0.988	**0.015**	**0.009**
Septal thickness (mm)	8.9 ± 1.7	8.6 ± 1.1	15.5 ± 3.7	**<0.001**	0.817	**<0.001**	**<0.001**
LGE positive, n	0	1	10				
LGE size LV mean (5th SD) (%)	0.0	0.014 ± 0.038	4.014 ± 3.970	**<0.001**	0.990	**0.001**	**<0.001**
LABORATORY VALUES							
proBNP (pg/ml)	37.7 ± 35.2	67.6 ± 44.3	492.9 ± 634.0	**<0.001**	0.248	**0.012**	**<0.001**
hsT (pg/ml)	6.9 ± 8.8	3.7 ± 0.8	28.8 ± 25.2	**0.001**	0.672	**0.003**	**0.001**
eGFR (ml/min/1.73 m^2)	104.2 ± 15.8	95.7 ± 19.6	71.9 ± 21.5	**<0.001**	0.321	**0.009**	**<0.001**
Creatinine (mg/dl)	0.94 ± 0.14	0.81 ± 0.14	1.46 ± 1.50	**0.042**	0.347	**0.020**	0.062
ACR (mg/g)	–	30.2 ± 24.2	1534.6 ± 4314.0	–	–	**0.008**	–
Lyso-Gb3 (ng/ml)	–	3.7 ± 3.9	36.0 ± 32.9	–	–	**0.003**	–

Data are presented as mean ± standard deviation. Significant p-values are marked in bold. Results of the analysis of variance (ANOVA) are shown. If ANOVA was significant post-hoc group comparisons were performed. ACR, albumin-to-creatinine ratio; cMRI, cardiac magnetic resonance imaging; eGFR, estimated glomerular filtration rate using the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) formula; ERT, enzyme replacement therapy; Fabry no FC, Fabry patients without Fabry cardiomyopathy; hsT, high sensitive cardiac Troponin T; LGE, late gadolinium enhancement; LGE size LV mean (5th SD), 5th standard deviation of left ventricular mean late gadolinium enhancement size; LVEF, left ventricular ejection fraction; LVM indexed to BSA, left ventricular end-diastolic mass indexed to body surface area; lyso-Gb3, Globotriaosylsphingosine; MSSI, Mainz severity score index; proBNP, prohormone of brain natriuretic peptide; SV, stroke volume.
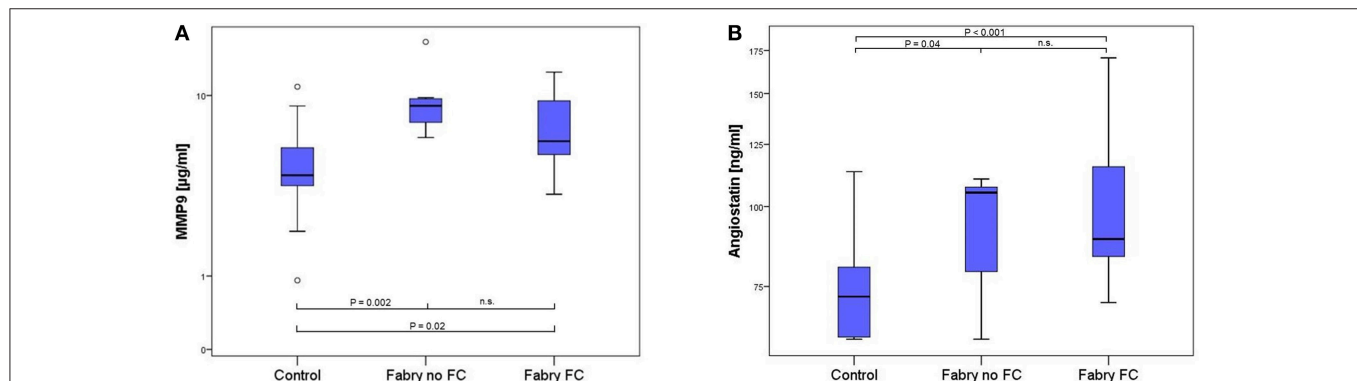
FIGURE 1 | Box-plots of MMP9 **(A)** and angiostatin **(B)** concentration in 17 healthy controls, 7 FD patients without and 15 FD patients with FC. Box plots represent median, 25th and 75th percentiles. Whiskers indicate minimum and maximum without outliers and extremes. On the Y-axis concentrations of the biomarkers are presented (log scale). Brackets indicate p-values from the analysis of variance (ANOVA) with *post-hoc* group comparisons. MMP9, matrix metalloproteinase 9; n.s., not significant.

are probably dependent on kidney function. Furthermore, an ensuing logistic regression of the eGFR and the two FD groups with and without FC was performed to evaluate the relationship between renal (eGFR) and cardiac function (existence of a FC) in FD patients. This test showed a significant relationship ($p = 0.045$). This finding is supported by the correlation of renal parameters with the LV-mass (**Table 3**).

DISCUSSION

The aim of this study was to investigate the role of vasculopathy and endothelial dysfunction in FD with special respect to Fabry associated cardiomyopathy. In blood samples of FD patients generally higher levels of MMP-9 and angiostatin could be detected independently of an existing FC, supporting the

TABLE 2 | Markers of endothelial dysfunction.

	Control (1)	Fabry no FC (2)	Fabry FC (3)	p (ANOVA)	p (1 vs. 2)	p (2 vs. 3)	p (1 vs. 3)
VEGF (pg/ml)	59.1 ± 51.7	70.1 ± 100.7	89.8 ± 58.6	0.127	–	–	–
suPAR (ng/ml)	1.6 ± 0.5	1.7 ± 0.7	3.2 ± 4.4	0.138	–	–	–
MMP-9 (μg/ml)	5.0 ± 2.4	9.6 ± 3.6	7.3 ± 3.1	**0.004**	**0.002**	0.163	**0.021**
- MMP-9 FD total	5.0 ± 2.4	8.0 ± 3.4*		**0.002****			
Angiostatin (ng/ml)	75.0 ± 15.2	92.9 ± 19.2	101.0 ± 27.3	**0.002**	**0.038**	0.468	**0.001**
- Angiostatin FD total	75.0 ± 15.2	98 ± 25*		**0.001****			
Arginine (μmol/l)	111.4 ± 30.1	114.4 ± 29.6	113.3 ± 31.0	0.944	–	–	–
L-homoarginine (μmol/l)	2.36 ± 0.82	2.24 ± 1.43	1.79 ± 0.79	0.133	–	–	–
ADMA (μmol/l)	0.63 ± 0.13	0.64 ± 0.09	0.77 ± 0.27	0.179	–	–	–
SDMA (μmol/l)	0.58 ± 0.12	0.57 ± 0.10	0.90 ± 0.64	**0.012**	0.959	**0.027**	**0.006**
hArg/ADMA	3.86 ± 1.54	3.48 ± 2.22	2.83 ± 2.14	0.061	–	–	–
hArg/SDMA	4.16 ± 1.44	4.17 ± 3.19	2.63 ± 1.78	**0.017**	0.629	0.078	**0.005**

*Data are presented as mean ± standard deviation. Significant p-values are marked in bold. Results of the analysis of variance (ANOVA) are shown. If ANOVA was significant post-hoc group comparisons were performed. ADMA, asymmetric dimethylarginine; FD total, both Fabry disease groups combined; hArg, L-homoarginine; MMP-9, matrix metalloproteinase 9; SDMA, symmetric dimethylarginine; suPAR, soluble urokinase-type plasminogen activator receptor; VEGF, vascular endothelial growth factor. *Mean ± standard deviation of both FD groups combined (2 + 3) **p-value of t-test: 2 + 3 vs. 1.*

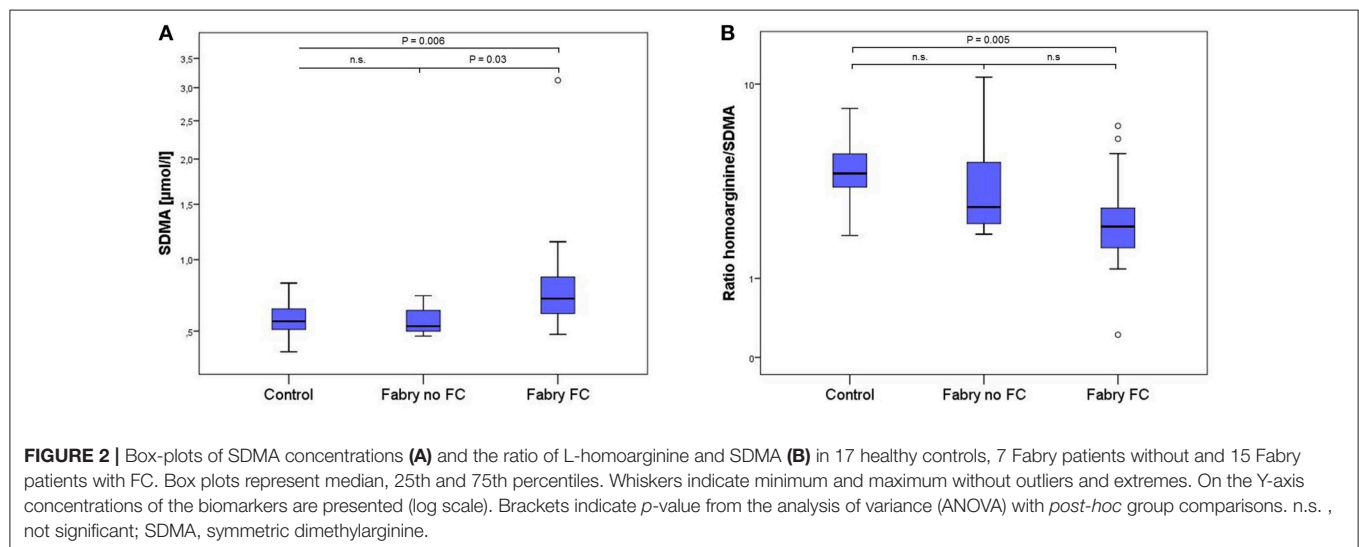**FIGURE 2 |** Box-plots of SDMA concentrations **(A)** and the ratio of L-homoarginine and SDMA **(B)** in 17 healthy controls, 7 Fabry patients without and 15 Fabry patients with FC. Box plots represent median, 25th and 75th percentiles. Whiskers indicate minimum and maximum without outliers and extremes. On the Y-axis concentrations of the biomarkers are presented (log scale). Brackets indicate p-value from the analysis of variance (ANOVA) with post-hoc group comparisons. n.s. , not significant; SDMA, symmetric dimethylarginine.

hypothesis of an altered extracellular matrix (ECM) turnover in FD.

In patients with FC higher concentrations of SDMA and a decreased ratio of hArg/SDMA could be revealed compared to healthy controls as well as to FD patients without overt cardiomyopathy. These parameters correlate with the ventricular mass as well as with cardiac and renal markers suggesting a potential causal relationship of kidney function and cardiac disease progress.

MMP-9, Angiostatin, and SuPAR

MMP-9 is part of a family of endogenous zinc-dependent endopeptidases. In the myocardium MMPs play an important role for structural integrity of the ECM (18). In patients with familial hypertrophic cardiomyopathy (HCM) an association of MMP-9 with gadolinium enhancement in cardiac MRI was

recently described and an important role of the MMP system in cardiac remodeling and fibrosis was proposed (26). In this context, Shah et al. identified significantly higher levels of MMP-9 in 29 FD patients compared to 21 healthy controls and hypothesized that MMP-9 plays an important role in the pathogenesis of FC and might be a valuable surrogate marker for the response to ERT (19). The findings from our study confirm the higher levels of MMP-9 in FD patients. However, a correlation of MMP-9 levels and Fabry associated cardiomyopathy cannot be confirmed.

MMP-9 cleaves matrix-bound plasminogen into angiostatin (27) which is a potent inhibitor of angiogenesis and has been shown to attenuate endothelial cell proliferation and migration (28). In FD the role of angiostatin has not yet been investigated. According to the elevated MMP-9 also Angiostatin concentrations increased in FD compared to controls and
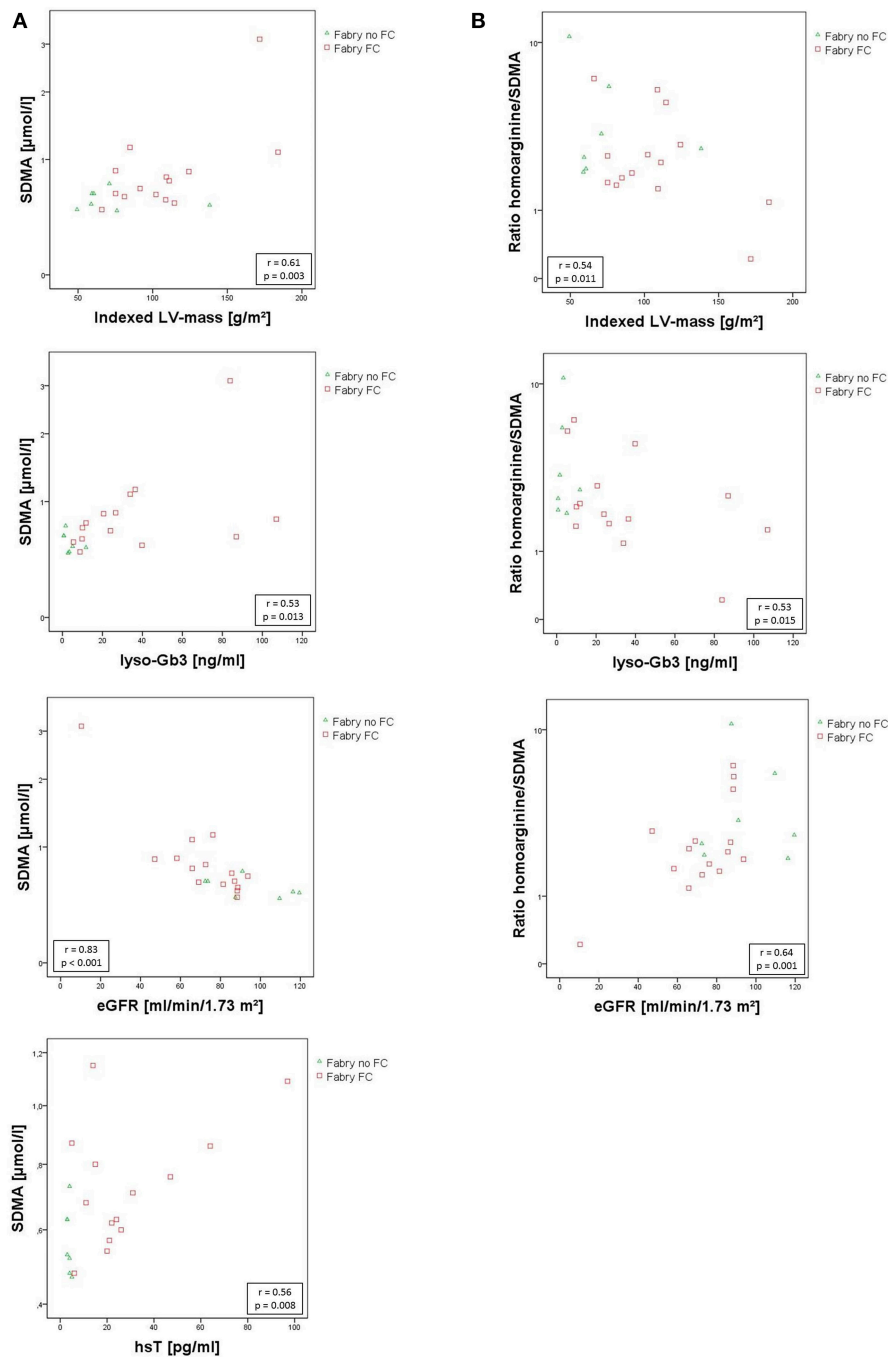
FIGURE 3 | **(A)** Scatterplots of SDMA correlating with indexed left ventricular mass, lyso-Gb3, eGfR, and hsT. **(B)** Scatterplots of hArg/SDMA correlating with indexed left ventricular mass, lyso-Gb3, and eGfR. SDMA concentrations and hArg/SDMA are presented on a log scale. Pearson correlation coefficient *r* and the corresponding *p*-values are given in boxes. eGFR, estimated glomerular filtration rate using the CKD-EPI. Formula; hsT, high sensitive cardiac Troponin T; indexed LV-mass, left ventricular mass indexed to body surface area; lyso-Gb3, Globotriaosylsphingosine; SDMA, symmetric dimethylarginine.

showed a trend toward higher concentrations in the FC group compared to FD patients without FC.

Interestingly, a study from Matsunaga et al. could demonstrate that reduced NO production leads to an increase of MMP-2 and MMP-9 activity and higher angiostatin concentrations concluding that NO production influences coronary angiogenesis

(29). Furthermore, an experimental study from Takahashi et al. revealed that angiostatin inhibits VEGF-stimulated NO production in human umbilical vein endothelial cells (30). In this context higher MMP-9 and angiostatin levels may contribute to an alteration of NO synthesis in FD. Moreover, an experimental animal study from Givvimani et al. found a switch to higher levels

TABLE 3 | Correlation to indexed left ventricular mass measured in cMRI.

	p-Value	*r*
SDMA (μmol/l)	0.003	0.61
ADMA (μmol/l)	0.031	0.47
hArg (μmol/l)	0.120	−0.35
hArg/SDMA	0.011	−0.54
hArg/ADMA	0.032	−0.47
suPAR (ng/ml)	0.002	0.64
hsT (pg/ml)	< 0.001	0.78
proBNP (pg/ml)	0.002	0.65
lyso-Gb3 (ng/ml)	0.026	0.50
eGFR (ml/min/1.73 m^2)	0.037	−0.46
Creatinine (mg/dl)	0.014	0.53
ACR (mg/g)	0.018	0.51

Results from the correlation tests of laboratory values and markers of endothelial dysfunction correlating with indexed left ventricular mass. ACR, albumin-to-creatinine ratio; ADMA, asymmetric dimethylarginine; eGFR, estimated glomerular filtration rate using the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) formula; hArg, L-homoarginine; hsT, high sensitive cardiac Troponin T; lyso-Gb3, Globotriaosylsphingosine; proBNP, prohormone of brain natriuretic peptide; r, Pearson's correlation coefficient; SDMA, symmetric dimethylarginine; suPAR, soluble urokinase-type plasminogen activator receptor.

of MMP-9 and anti-angiogenic markers such as angiostatin in the transition from compensatory hypertrophy to decompensated heart failure (31). Both, the effect of MMP-9 and angiostatin on NO synthesis and on a possible transition to decompensated heart failure in FC should be addressed in future studies.

Soluble uPAR is an important regulator of ECM proteolysis and is also involved in MMP activation by plasmin generation (32). For suPAR only a trend toward higher concentrations in FC could be determined, however, a positive correlation with the LV-mass could be found, as shown in **Table 3**. Soluble uPAR was shown to directly correlate with proteinuria (33). In this context a strong correlation to renal parameters and especially proteinuria ($p < 0.001$, $r = 0.74$) could be revealed.

Endothelial Dysfunction in Fabry Associated Cardiomyopathy

ADMA, SDMA, and hArg are non-proteinogenic amino acids structurally related to L-arginine. hArg has been shown to serve as an alternative substrate for NOS and to inhibit arginase. Thus, it is considered to increase NO formation (34, 35). In addition, low circulating concentrations of hArg have been proposed as a cardiovascular disease risk factor (36). ADMA on the other hand is an endogenous inhibitor of NOS (37), whereas its structural isomer SDMA does not directly interfere with NOS (38). However, SDMA inhibits the tubular L-arginine absorption in kidneys (39) and the y^+ transporter, which mediates the intracellular uptake of L-arginine (40). Therefore, SDMA has an indirect effect on NOS. Both dimethylarginines, ADMA and SDMA, are involved in endothelial dysfunction (41, 42), oxidative stress (43, 44), and atherosclerosis (45). A recent meta-analysis of prospective studies from Schlesinger et al. concluded that both markers are independently associated with cardiovascular disease and all-cause mortality (46).

In this context, the present study reveals a correlation of hArg, ADMA, and the ratio hArg/ADMA with the LV mass. Although group differences of ADMA and hArg showed a trend toward altered concentrations in FC patients, the effect was not significant. Moreover, the ratio of hArg/ADMA was superior compared to the single markers, but it equally missed significance in overall *F*-test of the analysis of variance with a *p*-value of 0.06.

However, in our cohort higher SDMA level in FC patients correlated with LV-mass, hsT, and lyso-Gb3 concentrations indicating the presence of endothelial dysfunction in these patients. Furthermore, this reveals a new possible mechanism of NO alteration in FD as this is the first study demonstrating higher SDMA level in FC patients' sera. Similar results are shown for the hArg/SDMA ratio. Although the alteration of the hArg/SDMA ratio was statistically not superior compared to the higher SDMA concentrations in FC, group differences of this ratio show that hArg might also contribute to a dysfunction of NO synthesis in FC. Due to the antagonistic effects of hArg and SDMA in arginine metabolism a further investigation of this ratio might be reasonable.

The discrepancy between significantly higher SDMA levels without ADMA group differences might be explained by their disparate ways of excretion. SDMA elimination is exclusively renal, whereas ADMA is also enzymatically excreted by the dimethylarginine dimethylaminohydrolases (DDAHs). DDAH-1 for example is highly expressed in the kidney and liver (47) and as FD patients do not typically show liver dysfunction enzymatic ADMA elimination might be sufficient, whereas SDMA accumulates due to the typical early occurring renal insufficiency in FD. In this regard, adjustment for eGFR showed a significant dependency of SDMA and hArg/SDMA to kidney function. Furthermore, an association of eGFR and FC could be revealed. It is well known that diastolic dysfunction caused by impaired LV relaxation may lead to congestive heart failure and consequently to renal insufficiency. In our cohort only 4 patients with FC presented with mild (grade 1) and 1 patient with moderate (grad 2) diastolic dysfunction. Thus, based on these findings one might speculate that renal insufficiency might rather contribute to FC due to an accumulation of SDMA and its negative effect on NO synthesis in the vascular system. However, a causal relationship between renal and cardiac function in Fabry disease has to be proven in further studies.

CONCLUSIONS

This study provides evidence for an altered ECM turnover with higher levels of MMP-9 and angiostatin in FD patients independent of an existing FC. Moreover, patients with FC showed higher SDMA and hArg/SDMA level, which correlated with LV mass, hsT, and lyso-Gb3 concentration but also with impaired renal function. Renal and cardiac function showed a relationship leading to the hypothesis that accumulation of SDMA due to renal insufficiency in FD might contribute to the development of endothelial dysfunction and subsequently lead to FC.

Limitations

A major limitation of this study is the small sample size of 22 investigated FD patients with only 7 patients without FC compared to more than twice as much patients with FC. Another limitation is the unequal distribution of male and female patients in the FD groups and the fact that this study is a single center study. Moreover, whether the relationship between kidney and cardiac function is causal cannot be proven by this study design.

Perspectives

To validate these findings multicenter studies including more patients to investigate the effects of MMP-9 and angiostatin on endothelial dysfunction in FD and to clarify the pathological impact of SDMA accumulation in Fabry associated cardiomyopathy are required.

ETHICS STATEMENT

This study was carried out in accordance with the principles outlined in the Declaration of Helsinki. The protocol was approved by the local ethics committee of the Ärztekammer Hamburg, Germany (approval no: PV4056).

AUTHOR CONTRIBUTIONS

JL: writing of the manuscript, experimental work. NL: experimental work, writing assistance. MA: MRI data acquisition, critical review of the data. NM: acquisition of clinical data, critical manuscript review. SL: statistical analysis. KC and ES: supervision and design of experiments, critical review of data analysis and manuscript. MP: experimental design, supervision of data analysis. writing of the manuscript.

FUNDING

REFERENCES

1. Zarate YA, Hopkin RJ. Fabry's disease. *Lancet* (2008) 372:1427–35. doi: 10.1016/S0140-6736(08)61589-5

2. Fabry H. Angiokeratoma corporis diffusum–Fabry disease: historical review from the original description to the introduction of enzyme replacement therapy. *Acta Paediatr Suppl.* (2002) 91:3–5. doi: 10.1111/j.1651-2227.2002.tb03102.x

3. Brady RO, Gal AE, Bradley RM, Martensson E, Warshaw AL, Laster L. Enzymatic defect in Fabry's disease. Ceramidetrihexosidase deficiency. *N Engl J Med.* (1967) 276:1163–7. doi: 10.1056/NEJM196705252762101

4. Favalli V, Disabella E, Molinaro M, Tagliani M, Scarabotto A, Serio A, et al. Genetic screening of anderson-fabry disease in probands referred from multispecialty clinics. *J Am Coll Cardiol.* (2016) 68:1037–50. doi: 10.1016/j.jacc.2016.05.090

5. Germain DP. Fabry disease. *Orphanet J Rare Dis.* (2010) 5:30. doi: 10.1186/1750-1172-5-30

6. Mehta A, Ricci R, Widmer U, Dehout F, Garcia de Lorenzo A, Kampmann C, et al. Fabry disease defined: baseline clinical manifestations of 366 patients in the Fabry Outcome Survey. *Eur J Clin Invest.* (2004) 34:236–42. doi: 10.1111/j.1365-2362.2004.01309.x

7. Mehta A, Clarke JT, Giugliani R, Elliott P, Linhart A, Beck M, et al. Natural course of Fabry disease: changing pattern of causes of death in FOS - Fabry Outcome Survey. *J Med Genet.* (2009) 46:548–52. doi: 10.1136/jmg.2008.065904

8. Felker GM, Thompson RE, Hare JM, Hruban RH, Clemetson DE, Howard DL, et al. Underlying causes and long-term survival in patients with initially unexplained cardiomyopathy. *N Engl J Med.* (2000) 342:1077–84. doi: 10.1056/NEJM200004133421502

9. Thurberg BL, Fallon JT, Mitchell R, Aretz T, Gordon RE, O'Callaghan MW. Cardiac microvascular pathology in Fabry disease: evaluation of endomyocardial biopsies before and after enzyme replacement therapy. *Circulation* (2009) 119:2561–7. doi: 10.1161/CIRCULATIONAHA.108.841494

10. Schiffmann R, Murray GJ, Treco D, Daniel P, Sellos-Moura M, Myers M, et al. Infusion of alpha-galactosidase A reduces tissue globotriaosylceramide storage in patients with Fabry disease. *Proc Natl Acad Sci USA.* (2000) 97:365–70. doi: 10.1073/pnas.97.1.365

11. Weidemann F, Niemann M, Stork S, Breunig F, Beer M, Sommer C, et al. Long-term outcome of enzyme-replacement therapy in advanced Fabry disease: evidence for disease progression towards serious complications. *J Intern Med.* (2013) 274:331–41. doi: 10.1111/joim.12077

12. Weidemann F, Niemann M, Breunig F, Herrmann S, Beer M, Stork S, et al. Long-term effects of enzyme replacement therapy on fabry cardiomyopathy: evidence for a better outcome with early treatment. *Circulation* (2009) 119:524–9. doi: 10.1161/CIRCULATIONAHA.108.794529

13. Pieroni M, Camporeale A, Della Bona R, Sabini A, Cosmi D, Magnolfi A, et al. Progression of Fabry cardiomyopathy despite enzyme replacement therapy. *Circulation* (2013) 128:1687–8. doi: 10.1161/CIRCULATIONAHA.113.002799

14. Rombach SM, van den Bogaard B, de Groot E, Groener JE, Poorthuis BJ, Linthorst GE, et al. Vascular aspects of Fabry disease in relation to clinical manifestations and elevations in plasma globotriaosylsphingosine. *Hypertension* (2012) 60:998–1005. doi: 10.1161/HYPERTENSIONAHA.112.195685

15. Kalliokoski RJ, Kalliokoski KK, Penttinen M, Kantola I, Leino A, Viikari JS, et al. Structural and functional changes in peripheral vasculature of Fabry patients. *J Inherit Metab Dis.* (2006) 29:660–6. doi: 10.1007/s10545-006-0340-x

16. Moon JC, Sachdev B, Elkington AG, McKenna WJ, Mehta A, Pennell DJ, et al. Gadolinium enhanced cardiovascular magnetic resonance in Anderson-Fabry disease. Evidence for a disease specific abnormality of the myocardial interstitium. *Eur Heart J.* (2003) 24:2151–5. doi: 10.1016/j.ehj.2003.09.017

17. Weidemann F, Breunig F, Beer M, Sandstede J, Stork S, Voelker W, et al. The variation of morphological and functional cardiac manifestation in Fabry disease: potential implications for the time course of the disease. *Eur Heart J.* (2005) 26:1221–7. doi: 10.1093/eurheartj/ehi143

18. Kelly D, Khan SQ, Thompson M, Cockerill G, Ng LL, Samani N, et al. Plasma tissue inhibitor of metalloproteinase-1 and matrix metalloproteinase-9: novel indicators of left ventricular remodelling and prognosis after acute myocardial infarction. *Eur Heart J.* (2008) 29:2116–24. doi: 10.1093/eurheartj/ehn315

19. Shah JS, Hughes DA, Tayebjee MH, MacFadyen RJ, Mehta AB, Elliott PM. Extracellular matrix turnover and disease severity in Anderson-Fabry disease. *J Inherit Metab Dis.* (2007) 30:88–95. doi: 10.1007/s10545-006-0360-6

20. Shen JS, Meng XL, Moore DF, Quirk JM, Shayman JA, Schiffmann R, et al. Globotriaosylceramide induces oxidative stress and up-regulates cell adhesion molecule expression in Fabry disease endothelial cells. *Mol Genet Metab.* (2008) 95:163–8. doi: 10.1016/j.ymgme.2008.06.016

21. Kang JJ, Shu L, Park JL, Shayman JA, Bodary PF. Endothelial nitric oxide synthase uncoupling and microvascular dysfunction in the mesentery of mice

deficient in alpha-galactosidase A. *Am J Physiol Gastrointest Liver Physiol.* (2014) 306:G140–6. doi: 10.1152/ajpgi.00185.2013

22. Kaneski CR, Moore DF, Ries M, Zirzow GC, Schiffmann R. Myeloperoxidase predicts risk of vasculopathic events in hemizygous males with Fabry disease. *Neurology* (2006) 67:2045–7. doi: 10.1212/01.wnl.0000247278.88077.09

23. Shu L, Vivekanandan-Giri A, Pennathur S, Smid BE, Aerts JM, Hollak CE, et al. Establishing 3-nitrotyrosine as a biomarker for the vasculopathy of Fabry disease. *Kidney Int.* (2014) 86:58–66. doi: 10.1038/ki.2013.520

24. Atzler D, Mieth M, Maas R, Boger RH, Schwedhelm E. Stable isotope dilution assay for liquid chromatography-tandem mass spectrometric determination of L-homoarginine in human plasma. *J Chromatogr B Analyt Technol Biomed Life Sci.* (2011) 879:2294–8. doi: 10.1016/j.jchromb.2011.06.016

25. Schwedhelm E, Maas R, Tan-Andresen J, Schulze F, Riederer U, Boger RH. High-throughput liquid chromatographic-tandem mass spectrometric determination of arginine and dimethylated arginine derivatives in human and mouse plasma. *J Chromatogr B Analyt Technol Biomed Life Sci.* (2007) 851:211–9. doi: 10.1016/j.jchromb.2006.11.052

26. Roldan V, Marin F, Gimeno JR, Ruiz-Espejo F, Gonzalez J, Feliu E, et al. Matrix metalloproteinases and tissue remodeling in hypertrophic cardiomyopathy. *Am Heart J.* (2008) 156:85–91. doi: 10.1016/j.ahj.2008.01.035

27. Cornelius LA, Nehring LC, Harding E, Bolanowski M, Welgus HG, Kobayashi DK, et al. Matrix metalloproteinases generate angiostatin: effects on neovascularization. *J Immunol.* (1998) 161:6845–52.

28. Moser TL, Stack MS, Asplin I, Enghild JJ, Hojrup P, Everitt L, et al. Angiostatin binds ATP synthase on the surface of human endothelial cells. *Proc Natl Acad Sci USA.* (1999) 96:2811–6. doi: 10.1073/pnas.96.6.2811

29. Matsunaga T, Weihrauch DW, Moniz MC, Tessmer J, Warltier DC, Chilian WM. Angiostatin inhibits coronary angiogenesis during impaired production of nitric oxide. *Circulation* (2002) 105:2185–91. doi: 10.1161/01.CIR.0000015856.84385.E9

30. Takahashi S, Shinya T, Sugiyama A. Angiostatin inhibition of vascular endothelial growth factor-stimulated nitric oxide production in endothelial cells. *J Pharmacol Sci.* (2010) 112:432–7. doi: 10.1254/jphs.10028FP

31. Givvimani S, Tyagi N, Sen U, Mishra PK, Qipshidze N, Munjal C, et al. MMP-2/TIMP-2/TIMP-4 versus MMP-9/TIMP-3 in transition from compensatory hypertrophy and angiogenesis to decompensatory heart failure. *Arch Physiol Biochem.* (2010) 116:63–72. doi: 10.3109/13813451003652997

32. Smith HW, Marshall CJ. Regulation of cell signalling by uPAR. *Nat Rev Mol Cell Biol.* (2010) 11:23–36. doi: 10.1038/nrm2821

33. Sinha A, Bajpai J, Saini S, Bhatia D, Gupta A, Puraswani M, et al. Serum-soluble urokinase receptor levels do not distinguish focal segmental glomerulosclerosis from other causes of nephrotic syndrome in children. *Kidney Int.* (2014) 85:649–58. doi: 10.1038/ki.2013.546

34. Moali C, Boucher JL, Sari MA, Stuehr DJ, Mansuy D. Substrate specificity of NO synthases: detailed comparison of L-arginine, homo-L-arginine, their N omega-hydroxy derivatives, and N omega-hydroxynor-L-arginine. *Biochemistry* (1998) 37:10453–60. doi: 10.1021/bi980742t

35. Hrabak A, Bajor T, Temesi A. Comparison of substrate and inhibitor specificity of arginase and nitric oxide (NO) synthase for arginine analogues and related compounds in murine and rat macrophages. *Biochem Biophys Res Commun.* (1994) 198:206–12. doi: 10.1006/bbrc.1994.1029

36. Atzler D, Gore MO, Ayers CR, Choe CU, Boger RH, de Lemos JA, et al. Homoarginine and cardiovascular outcome in the population-based Dallas Heart Study. *Arterioscler Thromb Vasc Biol.* (2014) 34:2501–7. doi: 10.1161/ATVBAHA.114.304398

37. Vallance P, Leone A, Calver A, Collier J, Moncada S. Accumulation of an endogenous inhibitor of nitric oxide synthesis in chronic renal failure. *Lancet* (1992) 339:572–5. doi: 10.1016/0140-6736(92)90865-Z

38. Bode-Boger SM, Scalera F, Kielstein JT, Martens-Lobenhoffer J, Breithardt G, Fobker M, et al. Symmetrical dimethylarginine: a new combined parameter for renal function and extent of coronary artery disease. *J Am Soc Nephrol.* (2006) 17:1128–34. doi: 10.1681/ASN.2005101119

39. Tojo A, Welch WJ, Bremer V, Kimoto M, Kimura K, Omata M, et al. Colocalization of demethylating enzymes and NOS and functional effects of methylarginines in rat kidney. *Kidney Int.* (1997) 52:1593–601. doi: 10.1038/ki.1997.490

40. Closs EI, Basha FZ, Habermeier A, Forstermann U. Interference of L-arginine analogues with L-arginine transport mediated by the y+ carrier hCAT-2B. *Nitric Oxide* (1997) 1:65–73. doi: 10.1006/niox.1996.0106

41. Memon L, Spasojevic-Kalimanovska V, Bogavac-Stanojevic N, Kotur-Stevuljevic J, Simic-Ogrizovic S, Giga V, et al. Assessment of endothelial dysfunction: the role of symmetrical dimethylarginine and proinflammatory markers in chronic kidney disease and renal transplant recipients. *Dis Markers* (2013) 35:173–80. doi: 10.1155/2013/306908

42. Boger RH. Association of asymmetric dimethylarginine and endothelial dysfunction. *Clin Chem Lab Med.* (2003) 41:1467–72. doi: 10.1515/CCLM.2003.225

43. Schepers E, Glorieux G, Dhondt A, Leybaert L, Vanholder R. Role of symmetric dimethylarginine in vascular damage by increasing ROS via store-operated calcium influx in monocytes. *Nephrol Dial Transplant.* (2009) 24:1429–35. doi: 10.1093/ndt/gfn670

44. Sydow K, Munzel T. ADMA and oxidative stress. *Atheroscler Suppl.* (2003) 4:41–51. doi: 10.1016/S1567-5688(03)00033-3

45. Miyazaki H, Matsuoka H, Cooke JP, Usui M, Ueda S, Okuda S, et al. Endogenous nitric oxide synthase inhibitor: a novel marker of atherosclerosis. *Circulation* (1999) 99:1141–6. doi: 10.1161/01.CIR.99.9.1141

46. Schlesinger S, Sonntag SR, Lieb W, Maas R. Asymmetric and symmetric dimethylarginine as risk markers for total mortality and cardiovascular outcomes: a systematic review and meta-analysis of prospective studies. *PLoS ONE* (2016) 11:e0165811. doi: 10.1371/journal.pone.0165811

47. Tran CT, Fox MF, Vallance P, Leiper JM. Chromosomal localization, gene structure, and expression pattern of DDAH1: comparison with DDAH2 and implications for evolutionary origins. *Genomics* (2000) 68:101–5. doi: 10.1006/geno.2000.6262

Exploring Coronary Artery Disease GWAs Targets With Functional Links to Immunometabolism

Maria F. Hughes [1,2,3,4*], Yvonne M. Lenighan [1,4], Catherine Godson [1,5] and Helen M. Roche [1,2,4]

[1] UCD Diabetes Complications Research Centre, Conway Institute of Biomolecular and Biomedical Research, University College Dublin, Dublin, Ireland, [2] Nutrigenomics Research Group, UCD Institute of Food and Health, School of Public Health Physiotherapy and Sports Science, University College Dublin, Dublin, Ireland, [3] Centre of Excellence for Public Health, Queen's University Belfast, Belfast, United Kingdom, [4] UCD Institute of Food and Health, School of Public Health Physiotherapy and Sports Science, University College Dublin, Dublin, Ireland, [5] School of Medicine, University College Dublin, Dublin, Ireland

Finding genetic variants that cause functional disruption or regulatory change among the many implicated GWAs variants remains a key challenge to translating the findings from GWAs to therapeutic treatments. Defining the causal mechanisms behind the variants require functional screening experiments that can be complex and costly. Prioritizing variants for functional characterization using techniques that capture important functional and regulatory elements can assist this. The genetic architecture of complex traits such as cardiovascular disease and type II diabetes comprise an enormously large number of variants of small effect contributing to heritability and spread throughout the genome. This makes it difficult to distinguish which variants or core genes are most relevant for prioritization and how they contribute to the regulatory networks that become dysregulated leading to disease. Despite these challenges, recent GWAs for CAD prioritized genes associated with lipid metabolism, coagulation and adhesion along with novel signals related to innate immunity, adipose tissue and, vascular function as important core drivers of risk. We focus on three examples of novel signals associated with CAD which affect risk through missense or UTR mutations indicating their potential for therapeutic modification. These variants play roles in adipose tissue function vascular function and innate immunity which form the cornerstones of immuno-metabolism. In addition we have explored the putative, but potentially important interactions between the environment, specifically food and nutrition, with respect to key processes.

Keywords: GWAS, immuno-metabolism, coronary artery disease, nutrition, omnigenic

INTRODUCTION

CVD is increasing and the distribution of risk factors is changing with increasing prevalence of obesity and type II diabetes (T2D), particularly among young adults (aged 18–45) in developed countries (1–3). The burden of CVD risk factors remains very high because of unhealthy contemporary lifestyles, with dysregulated balance between energy intake and physical activity. In addition, malnutrition, wherein excess energy is coupled by micronutrient deficiencies, amplifies genetic risk (4). The major cardiovascular consequences of obesity and T2D predominantly derive from dysregulated and inflamed adipose tissue, particularly perivascular or visceral fat surrounding

the organs (5). Visceral fat has limited expandability and becomes inflamed, with the resulting adipokine dysregulation adversely affecting vascular biology by promoting vasoconstriction, medial smooth muscle cell proliferation and endothelial dysfunction and is known as dysregulated immunometabolism (6). In the obese state, immune cells become activated and infiltrate metabolic tissues, chronic activation of inflammatory pathways in both vascular and immune components trigger stress kinase activation that impinge on the signaling of metabolic hormones such as insulin leading to impaired glucose and lipid homeostasis (7). Highly structured interactions between immune and metabolic responses are evolutionarily conserved and disruption of these interactions underlie many pathologies such as obesity and diabetes. Therapeutic solutions to tackle obesity, T2D and hypertension are drastically needed to reduce the overall burden of cardiovascular disease. However, many drugs or interventions have failed due to a lack of understanding of complex disease architecture (8, 9).

GWAs provided unique insights into the genetic architecture of complex diseases. Genetic architecture considers the overall composition of variants influencing a trait in terms of number, frequency and magnitude of effect and potential interactions, and can vary over traits (10). With increasing size and scope of GWAs, it has become clear that many complex traits are driven by enormously large numbers of variants of small effects. These variants are spread across the genome rather than in disease related pathways, include many without obvious connection to disease and/or related risk factors. These variants are potentially capturing most regulatory variants that are active in disease relevant tissues and the regulatory networks they form, may be so interconnected they affect the functions of core disease-related genes. This can be observed for variants that are heavily concentrated in regions that are transcribed or marked by active chromatin in disease-relevant tissues but with little enrichment for cell-type specific regulatory elements compared to broadly active regions. Boyle et al. (11) proposed that this pattern could be explained through an Omnigenic model of inheritance. This is an extension of RA Fischer's infinitesimal model of inheritance proposed nearly a century ago (12). The Omnigenic model considers that gene regulatory networks are so interconnected that all genes expressed in disease relevant cells are able to affect the functions of core disease-related genes. Most heritability is accounted for by effects of genes in peripheral pathways, outside of core pathways, which accounts for loci associated with multiple traits (pleiotrophy). Therefore, disease risk may be largely driven by genes with no direct relevance to disease and is propagated through regulatory networks to a much smaller number of core genes with direct effects.

In theory, the set of core genes must have a more pronounced effect on disease traits and proteins derived from these genes will drive pharmaceutical development and therapeutic strategies. However, how this works at the cellular regulatory network is incompletely understood. To understand the relevance of the variants for therapeutic development, it is crucial to understand their effect on protein level, activity or function. Even if a variant has a small effect on protein level and disease risk this protein

may still be a suitable target for disease prevention if considered in context with its disease architecture (10). For instance, the SNP associated with (HMG-CoA reductase) explains 0.26% of variance in LDL levels, manipulating this gene can reduce LDL levels by 30–40% and reduce CAD risk (13). Even if the variants affect protein level or function, there are numerous challenges to drug development.

AN OMNIGENIC ARCHITECTURE FOR CARDIOVASCULAR DISEASE?

GWAs has reproducibly associated over 160 variants with cardiovascular disease (14–20). By combining data from UK Biobank (34,541 cases and 261,984 non-cases) followed by replication in CardiogramplusC4D (88,192 cases and 162,544 controls), an additional 64 novel loci were recently prioritized (20). This identified a total of 163 loci associated with coronary artery disease (CAD) (21). Consistent with the omnigenic model for CAD genetic architecture, many novel candidate genes did not have an obvious connection to CAD and the genetic contribution was concentrated in regions transcribed or marked by active chromatin in relevant tissues (blood vessels and liver) but with little enrichment for cell-type specific regulatory elements. While they reconstituted a larger number of gene pathways/networks for CAD, increasing from 4 to 14%, overall the variants were spread throughout the genome and only 14% forming disease relevant pathways.

PRIORITIZING VARIANTS USING INFORMATION FROM FUNCTIONAL AND REGULATORY REGIONS

To prioritize core CAD-related genes, they fine mapped the regions characterizing the functional, cellular and regulatory contribution of the variants (22) and prioritized their significance using probabilistic models (23) to derive a set of genes with converging evidence of potential functional SNP-gene mechanisms for functional follow up studies. The fine mapping methods that they used are compared against related methods summarized in **Table 1**, reviewed in Schaid et al. (30). Integrating information from multiple omics approaches in this way provides a more comprehensive understanding of the flow of information from the disease driver to its functional consequence or interactions. Methods can now test the mediating mechanism of these genetic variants on complex traits (31). Their analysis prioritized 161 variants to candidate genes based on proximity, expression quantitative loci data, DEPICT analysis and long-range chromatin interactions of variants with gene promotors for signals of regulation using stringent conditions and identified 28 loci with convincing arguments for causal variation, 22 known and 6 novel or 19 potential core genes (with missense, intergenic, downstream or UTR mutations). Among known genes; APOE, PCSK9, ANGPTL4 and SORT1 are implicated as core genes in lipid metabolism (a key component of immunometabolism) and targeting the effects of these genes can reduce CAD risk (32–34). Of the 6
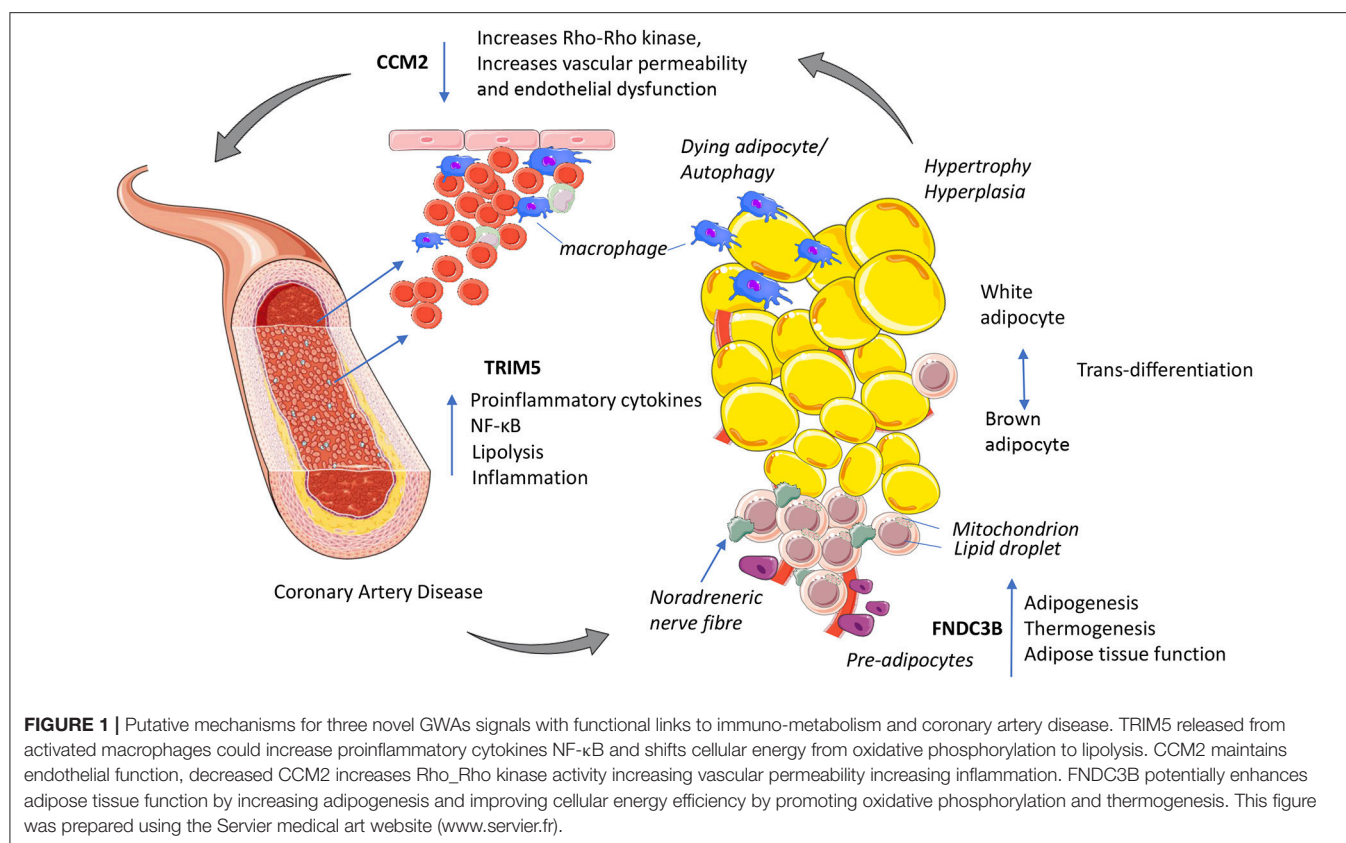
TABLE 1 | Summary of methods for fine mapping variants from GWAs.

Name	Combines	GWAs Data	Details	References
BIMBAM, GUESS	GWA and phenotype	Individual level data	Stepwise conditional analysis on SNPs with lowest p-value association until no additional SNPs reach preassigned threshold for strength of association with phenotypic trait	(24, 25)
FINEMAP	GWA and phenotype	Summary level data	Use GWAs summary statistics and SNP correlations to compute Bayes factors for strength of association with trait. Uses a shotgun stocastic search which allows more variants to be considered simultaneously	(26)
CAVIARBF	GWA and phenotype	Summary level data	CAVIAR differs from PAINTOR by modeling the uncertainty in the observed association statistics. CAVIARBF has been reported to be more accurate than PAINTOR in prioritizing variants when no annotation information is available	(27)
GARFIELD (GWAS analysis of regulatory and functional information enrichment with LD correction)	GWAs, functional annotation and phenotype	Summary or individual level data	Select SNPs from LD blocks to prioritize variants matched with regulatory/functional annotation (of 1,005 specific regions selected from ENCODE, GENCODE and Roadmap Epigenetics) incorporating genic annotation, chromatin sites, histone modifications, DNAse I hypersensitivity sites, transcription factor binding sites from cell lines from ENCODE with their strength of association with traits	(22)
PAINTOR (Probability Annotation INTegratOR); fastPAINTOR	GWAs, functional annotation and phenotype(s)	Summary or individual level data	Selects SNPs from LD blocks allowing for multiple causal variants and matched with functional/regulatory annotation data (ENCODE), PAINTOR up-weights variants in certain functional annotations (e.g., transcription start sites) while downweighting variants within annotations less relevant to the trait (e.g., intergenic) without making *ad-hoc* assumptions on which tissue-specific annotations are relevant to the trait of interest. fastPAINTOR updates previous method leveraging pleiotrophy across correlated traits with a new sampling scheme to improve efficiency, it integrates fine mapping across two (multiple) traits assuming same variants impact both traits though allowing potentially distinct effect sizes/opposite effects.	(23, 28)
SMR (summary data based Mendelian randomization) and HEIDI (heterogeneity in dependent instruments)	GWAS, eQTL, mQTLs	Summary or individual level data	Combines summary level multi-omics data to prioritize gene targets and their regulatory elements in 3 steps, using association tests, 1. map methylome QTL to genes (2 MB), map expression QTLs to trait, map trait to mQTL, if signals significant in all 3 steps infers target genes functionalyl relevant, can incorporate info from two independent studies.	(29)

novel signals, 3 are intergenic while 3 affect change through missense mutation or occur in a UTR3 region; these are TRIM5, FNDC3B, and CCM2 which are implicated in innate immunity, adipogenesis and vascular function, respectively, and all require functional follow up (**Figure 1**). Their study aimed to prioritize the CAD associations and elucidate regulatory connections that may influence the mechanism behind the associations, but according to the omnigenic model, broader regulatory connections between core genes must exist but are difficult to elucidate.

Several major challenges stand in the way to understanding how GWAs associations could become therapeutic targets. Most GWAs associations lie within non-coding regions making it difficult to predict their functions and identify targets/genes. Loci can be linked to multiple genes and the likely causal variant requires detailed investigation to elucidate the underlying mechanism. Functional follow up of important GWAs candidate loci now shows that multiple variants of small effect can synergistically drive dysfunction in regulatory networks, for example risk related to FTO (35), ANGPTL4 (17), GUCY1A3

(36), and SHROOM3 (37). To understand the mechanistic basis of increased adiposity associated with FTO, layers of OMICS data connecting epigenetic, gene co-expression and regulator expression followed by validation with genome editing elucidated the risk variant rs1421085 causes a loss of repression in AR1D5B which enhances expression of IRX3 and IRX5 increasing fat storage (35). Mining available OMIC data to gain insights into the complex regulatory circuitry behind these association signals has the potential to speed up functional follow-up by identifying novel links. We consider the three novel signals highlighted by van der Harst and Verweij for their strength of evidence and their importance to these pathways contributing to CAD risk or related traits such as adiposity and how these signals fit with other evidence supporting their contribution to disease risk. These may represent core genes but they may be signals that are context or cell specific to CAD. We also consider what the cell or tissue derived signals could offer therapeutically if they validated in independent studies. To this end, we explore a few examples wherein this paradigm may be relevant.

FIGURE 1 | Putative mechanisms for three novel GWAs signals with functional links to immuno-metabolism and coronary artery disease. TRIM5 released from activated macrophages could increase proinflammatory cytokines NF-κB and shifts cellular energy from oxidative phosphorylation to lipolysis. CCM2 maintains endothelial function, decreased CCM2 increases Rho_Rho kinase activity increasing vascular permeability increasing inflammation. FNDC3B potentially enhances adipose tissue function by increasing adipogenesis and improving cellular energy efficiency by promoting oxidative phosphorylation and thermogenesis. This figure was prepared using the Servier medical art website (www.servier.fr).

TRIM5, INNATE IMMUNE SIGNALING AND CAD RISK

The variant rs11601507 causes a missense mutation in TRIM5 and increases CAD risk ($p = 2.1 \times 10^{-12}$, OR 1.09 (95% C.I. 1.06, 1.11). Chromatin interactions between this variant and eQTLs in the promotors/enhancers of three other genes (TRIM6, OR52S1, OR52B6) suggest these genes enhance the expression of TRIM5. Chromatin interactions reveal relationships of chromatin organization in 3D space that may indicate biological function such as promotor-enhancer interactions. The evidence used to support rs11601507 is from a range of Hi-C experimental cell lines (20) (38). rs11601507 is a cis QTL for HBG2 (Hemoglobin) in whole blood (39) and shows significant tissue specific enrichment in veins and blood vessels [DEPICT analysis, (20)]. Ingenuity® pathway analysis (IPA®) prioritized TRIM5 and TRIM6 along with 14 other genes for association with CVD. IPA® considers upstream and downstream regulators of gene expression based on large scale causal networks (40).Interestingly, this same missense variant rs11601507 and a 5′UTR variant rs3824949 in TRIM5 has previously been associated with mean platelet volume ($p = 6 \times 10^{-19}$ and $p = 1 \times 10^{-24}$, respectively) (41) which is an example of pleiotrophy.

Given the enormous dimensionality of the phenome, it is unlikely that functional variants exist without pleiotrophic effects (42). Pleiotrophy can involve variants having effects on two or more traits via independent pathways (e.g., effects in different tissues) or effect of the variant in one trait causally related to variation in another trait. The risk allele of this variant has the same direction of effect for CAD and mean platelet volume. Using rs11601507 and other variants in a risk score, Astle et al. demonstrated a weak causal relationship between mean platelet volume and CAD risk using Mendelian randomization. Mean platelet volume is associated with increased hemolysis or free hemoglobin in the blood which is linked to increased inflammation. The TRIM5 association may be affecting both traits through inflammatory pathways.

TRIM5 promotes Interferon γ (IFNG) in macrophages, this forms part of the innate immune system (43). It has a capsid specific restriction factor that prevents infection from non-host adapted retroviruses. Interestingly, TRIM5 reciprocally enhances ubiquitination leading to co-operative action of IFNG and NF-κB pathways (44). There is a dynamic relationship between the innate immune system and metabolism, where re-configuration of energy metabolism between oxidative phosphorylation vs. glycolysis can define the immune-phenotype (45, 46). Fatty acids and other metabolites can influence and define immune cell functionality and cause metabolic reprogramming (45). It is hypothesized that this dynamic and reciprocal regulatory relationship between metabolism and inflammation plays a key role in metabolic disease including CAD (7). Macrophages play a key role as innate immune

cellular mediators of inflammation. Activated macrophages can recruit other monocytes/macrophages to a developing lesion and increase lipid uptake and instigate metabolic stress and reprogramming in adipose tissue. Macrophages can become "metabolically activated" in the presence of glucose, insulin and palmitate. Metabolically activated macrophages demonstrate similar effects to classically activated macrophages, where both types activate the TLR4 and NF-κB pathways to promote pro-inflammatory cytokine secretion. However, the metabolically active macrophages also activate PPARγ, therefore controlling inflammation by prompting lipid metabolism (47). TRIM5 promotes IFNG and through a mechanism of decreasing tryptophan metabolism (which viruses rely on), IFNG inhibits the central metabolic regulator mTOR and metabolically reprograms macrophages to switch from glycolysis to oxidative phosphorylation and upregulates inflammation. CVD is associated with changes in many immune cell types at multiple sites of critical metabolic function with a cumulative detrimental effect on cholesterol, lipid and glucose homeostasis (7, 22).

OLFACTORY SIGNALING INFLUENCES TRIM5 AND IS ALSO LINKED TO ADIPOSITY

The genetic mechanism associated with the TRIM5 variant suggests enhanced olfactory signaling enhances TRIM5 (innate immune signaling) to reduce lipolysis which enhances adiposity and increases risk of CAD. OR52B6, and OR52S1 are G-protein coupled olfactory signaling receptors (ORs) (48). These receptors interact with odorant molecules in the nose, to initiate a neuronal response that triggers the perception of a smell. OR52B6/OR52S1 have not been linked by GWAs signals as important regulatory variants but other variants related to olfaction have been been significantly linked to obesity development through GWAS (8). Olfactory signaling is highly complex and can play a bidirectional role in controlling energy homeostasis in response to sensory and hormonal signals from the central nervous system (CNS) (49). Essentially the ORs may alter the drive to eat a poor diet, leading to obesity, hence elaborating an environmental insult. Reduced olfactory signaling increases β-adrenergic receptors on white (WAT) and brown adipose tissue (BAT) increasing lipolysis and fatty acid oxidation reducing obesity in mice (49). Olfaction influences the loss of function mutation in ADCY3 (50) gene and its interaction with the major obesity gene MC4R which disrupts ciliary targeting in neuronal cells critical for body weight regulation (35, 51). Heterozygous or homozygous null mice for ADCY3 are unable to smell (35).

In summary, innate immunity is important in the pathogenesis of CVD, here the association between a variant linked to innate immunity is reinforced and mediated through a novel mechanism of olfaction. The immunosuppressant drug cyclosporine is an antagonist for TRIM5 suggesting a potential therapeutic intervention is available to explore for functional relevance (20). More generally, targeting systemic

inflammation through interleukin 1β (e.g., Canakinumab) has been shown to reduce CVD risk and by doing so has validated the inflammatory hypothesis of atherothrombosis (52). The variants TRIM5 and PROCR ($p = 6.8 \times 10^{-12}$) reaching GWAs significance are related to inflammation, which are relatively newly identified, show convergence between biological and genetic determinants of CVD and add to this inflammatory hypothesis (18, 20). An alternative therapeutic paradigm to anti-inflammatory modalities may be efforts to mimic the resolution of inflammation using specialized lipid mediators and their targets (53–55).

FNDC3B, ADIPOGENESIS AND CAD RISK

rs12897 is a common variant (MAF 0.41) showing a protective association with CAD; OR 0.96 (95% C.I. 0.95, 0.97) ($p = 1.9 \times 10^{-10}$), this SNP is an eQTL for the protein coding gene Fibronectin type III domain containing 3B (FNDC3B) (39) occurring in the 3′ UTR region of the mRNA likely affecting post-transcriptional regulation of gene expression. This SNP was the 3rd top gene prioritized by DEPICT ($p = 1 \times 10^{-21}$) in the overall analysis (20). IPA® prioritized a functional association between the protein of FNDC3B, TRIM5, TRIM6, VEGFA, and 12 other genes for association with CAD supporting a broader connectivity among these.

Adipogenesis is a key regulatory process, which determines adipose functionality, and its dysfunction is associated with metabolic-inflammation, hypoxia and related risks including insulin resistance (6) and deregulated cholesterol homeostasis and lipid metabolism (56) all of which lead to greater T2D and CVD risk (57). FNDC3B (alias FAD104) is a positive regulator of adipogenesis (58); specifically at the early stages of adipogenesis (59, 60). FNDC3B variant rs12897 was previously associated with large scale GWAs on height ($p = 3 \times 10^{-39}$) (61), waist-to-hip ratio (WHR) adjusted BMI ($p = 8 \times 10^{-10}$) and HIP adjusted BMI ($p = 3 \times 10^{-12}$) (62) and heart rate ($p = 1 \times 10^{-9}$) (63). Interestingly, intronic variants near FNDC3B strongly associated with intra-ocular pressure $p = 9 \times 10^{-48}$ (64) and $p = 5 \times 10^{-50}$ (65), however these variants are not in LD with rs12897. Although intraocular pressure may reflect changes in heart rate, this association may operate through a different, peripheral CAD pathway.

GWAs on specific adiposity traits and fat distributions (pericardial fat, visceral fat, WHRadjBMI, body fat percentage) have shown distinct genetic components (66, 67). WHR adjusted BMI and body fat percentage traits identified adipogenesis candidate genes to play key roles in adiposity. These genes included BMP2 ($p = 3.3 \times 10^{-14}$), CEBPA, PPARγ, HOXC-mir196, TBX15, and PEMT but these variants had no apparent regulatory links/eQTLs (62). While CEBPA and PPARγ are essential for white adipose tissue differentiation and are master regulators of adipogenesis, BMP2, like FNDC3B, is involved in early stage adipogenesis. FNDC3B and BMP2 are both involved in the early stage commitment of pre-adipocytes to proliferate and differentiate. FNDC3B (and BMP2) specifically

induce and/or regulate the differentiation of committed progenitor cells toward adipogenesis or osteogenesis (68). Adipocytes and in particular pre-adipocytes are now recognized as more than fat-storing organelles having the capability to secrete cytokines and adipokines thus contributing to inflammation (69).

ADIPOGENESIS AS A THERAPEUTIC MECHANISM TO REDUCE METABOLIC RISK

Defining the effectors which control the fate of adipocytes is of great interest to the therapeutic treatment of obesity. Obese individuals have a smaller proportion of brown adipose tissue (BAT) compared to white adipose tissue (WAT) which expands in response to lipid excess by hypertrophy, hyperplasia and inflammation and upon reaching a certain size become dysfunctional and necrotic, promoting macrophage infiltration. The conversion of WAT to the more functional energy dispersing BAT adipocytes would be a valuable approach to the treatment of obesity and its metabolic complications and is becoming the focus of anti-obesity research (9). Conversion of WAT to BAT can occur by two processes; adipogenesis (i.e., de-novo-adipocyte differentiation of precursor cells which FNDC3B may play a role) or more commonly trans-differentiation (i.e., WAT to beige/brite transition through molecular reprogramming, increasing mitochondrial oxidative phosphorylation/lipolysis requiring increased levels of uncoupling protein 1 (UCP-1) and enervated with b-adrenoreceptors (9). BAT derived from adipogenesis is more sensitive to stimuli from BMP7 (70) and BMP4 (71), irisin/FNDC5B (72), FGF21, and others. Irisin/FNDC5B is a myokine/cytokine that induces thermogenesis except in the obese state where it has a complex adaptive response to counterbalance decreased insulin sensitivity and other metabolic disorders associated with obesity (73) and is a key molecular target to induce browning of WAT (9). FNCD5B expression is highest at early stages of preadipocyte differentiation sharing 53% homology with FNDC3B but their relationship is unclear. While the eQTL affects the expression of FNDC3B, it is not known if this regulation is specific to a CVD relevant tissue or cell type. If regulation of FNDC3B is a key step in increasing adipogenesis, modulation of this process could enhance thermogenesis.

The strongest obesity variant associated to date, FTO, can act through a complex regulatory network also affecting pre-adipocyte differentiation highlighting the importance of this pathway (35). Interestingly, this regulation ensures it is restricted in a cell/tissue specific way to preadipocytes and mesenchymal adipocyte progenitors, not in brain or 120 other cell types (35). The causal variant associated with FTO disrupts AR1D5B binding in the risk haplotype leading to a loss of repression, this derepresses pre-adipocyte enhancer activity and increases IRX3 and IRX5 expression which represses mitochondrial thermogenesis and adipocyte browning making cells more likely to store fat.

OTHER VARIANTS OF GENES INFLUENCING TRANSDIFFERENTIATION OF WAT TO BAT ARE ALSO LINKED TO CAD RISK

In addition to FNDC3B, two other variants *PRDM16 and TWIST1* recently associated with CAD risk play key roles in adipogenic transdifferentiation of WAT to BAT which highlights this pathway as relevant to disease risk and therapeutic exploration. From a biological perspective PRDM16 is one of the most effective molecular targets to induce white-to-brown adipocyte trans-differentiation (9) and an intronic variant rs2493298 close to PRDM16 was recently identified to increase CAD risk (20). The variant rs2493298 $p = 1.9 \times 10^{-9}$, near PRDM16 occurs in an intronic region which physically interacts (chromatin, Hi-C experiments) with the promotors of three genes that act as enhancers (20) which have roles in metabolism (9). PRDM16 is essential for normal BAT function, it interacts with C/EBPβ and these are considered master regulators of BAT function (74), and functions in a feedback loop with PPARγ and SIRT1. No pharmacological targets for PRDM16 are advanced enough to explore in clinical trials (9). PPARγ coactivator 1 alpha (PGC1a) is another important control point of the BAT phenotype and it is repressed by TWIST1 which blocks target genes associated with PGC1a activity leading to browning of WAT (75). An eQTL rs21079595 intergenic to TWIST1 increases risk of CAD 1.3×10^{-24} and was prioritized as a core CAD related gene (20). Previously this variant had been linked to HDAC9 gene through proximity, but expression data from the Stockholm-Tartu Atherosclerosis Reverse Network Engineering Task Study (STARNET) in two different tissues prioritized this eQTL variant to TWIST1 (76). Manipulation of PGC1a can also achieve reductions in inflammatory disease risk and enhance adipogenesis through dietary fat modification (6).

In summary, several variants associated with adipogenesis have been associated with CAD risk. Dysregulation of adipocyte browning/thermogenesis, particularly in visceral fat surrounding thoracic and aortic arch, is important in the pathogenesis of CVD. FNDC3B is among several variants that impact adipogenesis, while the regulatory networks among these still need more complete understanding, manipulation of this pathway at the preadipocyte stage could impact CVD risk.

CCM2, ENDOTHELIAL FUNCTION AND CVD RISK

The variant rs2107732 causes a missense mutation in the CCM2 gene and is associated with reduced risk of CAD (OR 0.94 (95% C.I. 0.93–0.96), $p = 3.6 \times 10^{-8}$ (20). A variant in the promotor of MYOG1 forms a chromatin interaction with the CCM2 variant suggesting this regulates CCM2 (20). MYOG1 is a muscle specific transcription factor that induces myogenesis (muscle formation). Mutations in an orthologous mouse gene of CCM2 cause a cardiovascular phenotype in mice and mutations in MYOG1 caused abnormalities in inflammation/white blood cells of mice

(Mouse Genome Informatics database) (20). Inherited loss-of-function mutations in CCM2 (and also CCM1 and CCM3 genes) are implicated in abnormal vascular morphogenesis and can cause vascular lesions called cerebral cavernous malformations which develop in the human CNS (77). CCM2 is expressed in the brain and heart and CCM genes (including CCM2, CCM1, CCM3) are crucial regulators of heart and vessel formation and integrity by restricting vascular permeability and maintaining vascular homeostasis (78–80). These genes form complexes but also have complex independent roles (81). CCM2 restricts vascular permeability and maintain endothelial barrier function (tight and adherens junctions) by inhibiting Rho A-Rho kinase activity by enhancing Rho A proteasome degradation (79, 82). A lack of CCM2 increases Rho A Rho kinase activity which disrupts endothelial cell-cell contact causing permeability and stress fiber formation,, which is the initial phase in many cardiovascular diseases and characteristic of pathologically activated vascular endothelium. The response of CCM2 may be different in the inflammatory state and the MYOG1 transcription factor may influence this under certain conditions. Endothelial dysfunction reduces the ability of arteries to fully dilate, which stimulates vasodilators from the endothelium like nitric oxide (NO), decreased availability of NO or inactivation due to reactive oxygen species increases dysfunction. An intronic variant in NOS3 is also prioritized as a core causal variant of CAD alongside CCM2 both being important to blood vessel morphology and function (20) with several other rare and common variants in GUCY1A3 PDE5A and PEDE3A (16, 20, 36, 83, 84) highlighting the importance of the NO/cGMP signaling pathway to atherosclerosis and CAD risk.

Increased vascular permeability correlates with neo-angiogenesis (80). CCM proteins also control angiogenesis via Rho-kinase and other signaling pathways (78), CCM2 inhibits angiogenesis, loss of CCM2 causes dramatic angiogenic remodeling abnormalities (85). Adipose tissue is probably the most highly vascularized tissue in the body, as each adipocyte is encircled by capillaries, angiogenesis plays a key role in its function (86). Angiogenesis is driven by a complex interplay of angiogenic factors and inhibitors including vascular endothelial growth factor A (VEGFA). VEGFA is among the top 64 novel CAD loci increasing risk of CAD (OR 1.95 (95% C.I. 1.03–1.06), $p = 1.9 \times 10^{-12}$) (20) and waist to hip ratio adjusted BMI $p = 3 \times 10^{-27}$ (62). CCM proteins and particularly CCM3 can regulate VEGFA expression (80) [typically CCM2 and CCM3 function as a complex, (81)]. Lipid accumulation in adipocytes activates Rho Rho kinase signaling by breaking endothelial cell barriers/stress fiber formation triggering inflammatory changes (87). Vascular remodeling determines the flexibility and metabolic rate of adipose tissue and the communication between adipose and endothelial cells is crucial. Dysfunctional communication in obese individuals contributes to development and progression of T2D including impaired vasodilation, hypoxia and inflammation. CCM2 and VEGFA play roles at the interface of this cellular communication. A more complete understanding of the regulatory networks connecting CCM2 (inhibits angiogenesis) and VEGFA (stimulates angiogenesis, of which there are already targeted drugs), might synergistically increase the resulting therapeutic efficacy to combat obesity and CVD.

GWAS TARGETS WITH FUNCTIONAL LINKS TO IMMUNO-METABOLISM AND CORONARY ARTERY DISEASE

In summary, the three novel GWAs signals implicated in CAD risk play putative roles in immuno-metabolism (**Figure 1**). TRIM5 has potential to increase innate immunity, inflammation and CAD risk via macrophage infiltration of adipose tissue increasing metabolic stress. Activation of pro-inflammatory immune cells requires a shift to move from energy efficient oxidative phosphorylation to anaerobic glycolysis favoring glucose as substrate. This break or shift occurs when macrophages become polarized (M1) and is associated with nitric oxide production; an M1 effector molecule triggered by increasing oxidative stress. The mutation in CCM2 may reduce oxidative stress to maintain endothelial function, control angiogenesis and vascular remodeling of blood vessels including those surrounding adipose tissue to reduce CAD risk. Inhibiting glycolysis promotes the resolution of inflammation. FNDC3B enhances adipose tissue function by increasing adipogenesis and improving cellular energy efficiency by promoting oxidative phosphorylation and thermogenesis, with PRDM16 and TWIST1 playing similar roles in modifying CAD risk (**Figure 1**).

DIETARY INTERVENTIONS CONNECTING ADIPOGENESIS AND METABOLIC INFLAMMATION AS THERAPEUTIC MECHANISMS TO REDUCE METABOLIC RISK

Since the recognition that fatty acids can modulate an inflammatory response, e.g., via lipid induced re-programming of macrophage metabolism and inflammation (88) or the NLRP3 inflammasome (89, 90) they have been studied for their immunomodulatory effect on insulin resistance and dysregulated lipid metabolism pathways. Dietary manipulation and certain nutrients have the potential to modulate inflammatory responses. Obesity promotes adipose hypertrophy, with inflammation interacting with the adipogenic process. Pro-inflammatory cytokines IFNG, IL-1β and TNFα, inhibit adipogenesis by downregulating PPARγ and C/EBP (91–94) and several dietary components can modulate this effect e.g., reservatrol, flavonoids and polyphenols (95–97). Dietary fat modification to replace saturated fatty acids (SFA) with monounsaturated fatty acids (MUFA) and polyunsaturated fatty acids (PUFA) may provide a potential strategy to lessen inflammation that enhances adipogenesis to attenuate insulin resistance and dysregulated lipid metabolism (6) but the impact of dietary fat modification in humans has varied (98).

Efforts to explain this inter individual variability in response, has focused on the interaction between the genes, metabolites and diet. As diet is the exogenous source of many metabolites,

as well as affecting the generation of endogenous metabolites, interactions with the nutritional environment are plausible. However, many putative gene/variant-diet associations have failed to replicate in large studies (99) with various approaches to enhance power (100). Some intriguing examples of specific variant-metabolite interactions modulating disease risk exist from small studies (101, 102). The variant, rs5082 of APOA2 interacts with SFA intake to influence risk of obesity (101). This is modulated through an epigenetic effect on APOA2 regulatory region which promoted an APOA2 expression difference between APOA2 genotypes on a high SFA diet. This selectively dysregulated branched chain and tryptophan metabolic pathways with possible implications for food intake.

UNDERSTANDING THE REGULATORY NETWORKS UNDERLYING METABOLIC TRAITS

Attention is shifting to large scale studies integrating transcriptomic and metabolomic data to understand the interplay between genes and metabolites (103). To explore genes playing key roles in immunometabolism more specifically, Nath et al. integrated transcriptomics (focusing on immune networks) and metabolomics using 2,168 individuals from two general population cohorts (104). They identified significant expression quantitative loci in 8 immune gene networks highlighting the genetic foundations of these effects. For example, an eQTL in the ARHGEF3 gene (rs1354034 $p = 7 \times 10^{-28}$) had trans regulatory effects on several genes associated with platelet function and this module had diverse effects on 55 metabolites. Other important core immunometabolic associations related to neutrophil activation and viral response. A subset of the cohort measured repeatedly over 7 years, demonstrated the gene-metabolite effects were temporally stable (104). As long-term OMICs data will be collected on population cohorts over time, these signals are likely to become more reliable.

Identifying the genetic basis to these interactions can be useful therapeutically to modulate the variant itself for individualized treatment or modulate the pathway the variant functions in, which can have much wider implications for population treatment e.g., PCSK9 inhibitors for individual and population level treatment of hypercholesterolemia and CVD risk (34). With better understanding of metabolite-immune interactions, *in vivo* and interventional studies can be developed to modulate these interactions through existing lipid lowering medications, gut microbe effects or dietary changes. In this way, the immune system itself can be harnessed to reduce the burden of cardiovascular and metabolic disorders (55, 71). With distinct lifestyle strategies now known to differentially affect the way adipose tissue is stored and utilized in the body (105), it is important to understand where and how the drivers of these regulatory networks are acting, which might be under specific situations or locations.

CELLULAR SPECIFICITY OF REGULATORY NETWORKS

Progress has been made to determine the tissues and cell types underlying disease through the GTex consortium (106). GTex, Roadmap Epigenomics and Functional Annotation of Mammalian Genomes 5 (FANTOM5) provides reference sets for multi-tissue gene expression and epigenomics consistently evaluated on the same individuals with available tissues. Different layers of regulation can exist from post-transcriptional, post-translational, protein-protein interactions and intercellular signaling, mediated through chromatin interactions and expression quantitative trait loci. Assuming that most of the regulation occurs through genes (linked by eQTLs), regulation can occur at the tissue level, broad cell population level or in very specific cell types (106–110). For CVD, multiple cell types or highly specialized cell types may be involved (e.g., vascular, liver, adipose) where cellular networks could have variable expression across cell types (111). The effect of particular variants would then be an average of its effect size in each cell type weighted by cell type importance (11). Mapping GWAs signals to promotors/enhancers measured by cap analysis gene expression (CAGE) found regulation for specific diseases could be turned on/off in similar complex patterns across different cell types. For example, shared cell type specific regulatory networks distinguishing two sub-types of ulcerative colitis could be distinguished based on regulatory signals guided by GWAs signals enriched in either monocytes exposed to inflammatory signals or epithelial cells (108).

To identify cell type specific gene regulation, grouped cell types or deconvolution methods have been used, but the methods tend to be biased to specific cell types or difficult-to-identify less abundant cell types (112). It is possible to calculate the probability a GWAs variant and eQTL tag the same functional effect and infer the tissues where the effect for a trait is likely manifested (113). Single cell RNAseq (scRNAseq) can identify cell-type or context specific eQTLs, but the requirement for fresh tissue and costs limits large scale screening. Mapping monogenic kidney mutations or genome-wide variants associated with chronic kidney disease to gene expression from scRNAseq of 57,979 mouse kidney cells, Park et al. inferred that these variants were expressed in only one particular cell type (114). This suggests, most genetic diseases of the kidney can be traced to single cell types. Using intercellular variation from expression profiles from 25,000 peripheral blood mononuclear cells from 45 donors, scRNAseq identified cell type specific cis-eQTLs. Although gene regulatory networks were highly personal, their approach identified more genes under genetic control or specific cell type in which the effect is most prominent and found examples of SNPs influencing the co-expression of 2 genes (115). scRNAseq has the potential to group and examine the effects of cells along the cell cycle, along a differentiation path (e.g., adipocyte differentiation) or along a response to an environmental stimulus (e.g., inflammatory signaling) (116). With improved understanding of how these genes impact cell types and tissues, more specific

targeted interventions can be developed, for instance improved drugs, mobilizing specific fat deposits (105) or nutritional interventions (117).

CONCLUSIONS

We have highlighted three novel variants associated with CAD risk which have been prioritized and annotated based on systems genetics approaches including expression quantitative trait analysis and network analysis to infer their functional relevance. These core variants play roles in innate immunity, adipogenesis and endothelial function which drive coronary artery disease and principally in the role that obesity and T2D shape the pathogenesis of CAD through immuno-metabolism. Core variants representing these pathways provide a starting point to potential mechanism that may lead to therapeutic manipulation with further understanding of the regulatory networks connecting these is needed. Given that CAD is a multifactorial disease, it may be possible in the future to develop individual treatment strategies based on these variants or design relevant population level interventions based on the pathways these variants highlight for subsets of people in the population with subtypes of CAD risk related to obesity or T2D.

AUTHOR CONTRIBUTIONS

MH wrote the paper. YL, CG, and HR discussed and wrote sections, refined and edited the text. All agree with the submission.

FUNDING

REFERENCES

1. Leening MJG, Cook NR, Ridker PM. Should we reconsider the role of age in treatment allocation for primary prevention of cardiovascular disease? *Eur Heart J.* (2017) 38:1542–7. doi: 10.1093/eurheartj/ehw287

2. Tran J, Norton R, Conrad N, Rahimian F, Canoy D, Nazarzadeh M, et al. Patterns and temporal trends of comorbidity among adult patients with incident cardiovascular disease in the UK between 2000 and 2014: A population-based cohort study. *PLoS Med.* (2018) 15:e1002513. doi: 10.1371/journal.pmed.1002513

3. McMorrow AM, Connaughton RM, Magalhaes T. Personalized cardio-metabolic response to an eight week anti-inflammatory intervention in obese adolescents. *Mol Nut Food Res.* (2018) 62:e1701008. doi: 10.1002/mnfr.201701008

4. McArdle MA, Finucane OM, Connaughton RM, McMorrow AM, Roche HM. Mechanisms of obesity-induced inflammation and insulin resistance: insights into the emerging role of nutritional strategies. *Front Endocrinol (Lausanne).* (2013) 4:52. doi: 10.3389/fendo.2013.00052

5. Nazare JA, Smith JD, Borel AL, Haffner SM, Balkau B, Ross R, et al. Ethnic influences on the relations between abdominal sub- cutaneous and visceral adiposity, liver fat, and cardiometabolic risk profile: the International Study of Prediction of Intra- Abdominal Adiposity and Its Relationship With Cardiometabolic Risk/ Intra-Abdominal Adiposity. *Am J Clin Nutr.* (2012) 96:714–26. doi: 10.3945/ajcn.112.035758

6. Finucane OM, Lyons CL, Murphy AM, Reynolds CM, Klinger R, Healy NP, et al. Monounsaturated fatty acid-enriched high-fat diets impede adipose NLRP3 inflammasome-mediated IL-1β secretion and insulin resistance despite obesity. *Diabetes* (2015) 64:2116–28. doi: 10.2337/db14-1098

7. Hotamisligil GS. Foundations of immunometabolism and implications for metabolic health and disease. *Immunity* (2017) 47:406–20. doi: 10.1016/j.immuni.2017.08.009

8. Ghosh, S., Bouchard C. Convergence between biological, behavioural and genetic determinants of obesity. *Nat Rev Genet.* (2017) 18:731–748. doi: 10.1038/nrg.2017.72

9. Giordano A, Frontini A, Cinti S. Convertible visceral fat as a therapeutic target to curb obesity. *Nat Rev Drug Discov.* (2016) 15:405–24. doi: 10.1038/nrd.2016.31

10. Timpson NJ, Greenwood CMT, Soranzo N, Lawson DJ, Richards JB. Genetic architecture: the shape of the genetic contribution to human traits and disease. *Nat Rev Genet.* (2018) 19:110–24. doi: 10.1038/nrg.2017.101

11. 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

12. Fisher R. The correlation between relatives on the supposition of mendelian inheritance. *Trans R Soc Edinburgh* (1919) 52:399–433. doi: 10.1017/S0080456800012163

13. Illingworth DR, Stein EA, Mitchel YB, Dujovne CA, Frost PH, Knopp RH, et al. Comparative effects of lovastatin and niacin in primary hypercholesterolemia. A prospective trial. *Arch Intern Med.* (1994) 154:1586–95. doi: 10.1001/archinte.1994.00420140051007

14. Schunkert, H., König, I.R., Kathiresan S. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet.* (2011) 43:333–8. doi: 10.1038/ng.784.

15. CARDIoGRAMplusC4D Consortium, 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

16. Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, et al. A comprehensive 1000 Genomes–based genome-wide association meta-analysis of coronary artery disease. *Nat Genet.* (2015) 47:1121–30. doi: 10.1038/ng.3396

17. Verweij N, Eppinga RN, Hagemeijer Y, van der Harst P. Identification of 15 novel risk loci for coronary artery disease and genetic risk of recurrent events, atrial fibrillation and heart failure. *Sci Rep.* (2017) 7:2761. doi: 10.1038/s41598-017-03062-8

18. Howson JMM, Zhao W, Barnes DR, Ho WK, Young R, Paul DS, et al. Fifteen new risk loci for coronary artery disease highlight arterial-wall-specific mechanisms. *Nat Genet.* (2017) 49:1113–9. doi: 10.1038/ng.3874

19. Nelson CP, Goel A, Butterworth AS, Kanoni S, Webb TR, Marouli E, et al. Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nat Genet.* (2017) 49:1385–91. doi: 10.1038/ng.3913

20. 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

21. Erdmann J, Kessler T, Munoz Venegas L, Schunkert H. A decade of genome-wide association studies for coronary artery disease: the challenges ahead. *Cardiovasc Res.* (2018) 114:1241–57. doi: 10.1093/cvr/cvy084

22. Iotchkova V, Huang J, Morris JA, Jain D, Barbieri C, Walter K, et al. Discovery and refinement of genetic loci associated with cardiometabolic risk using dense imputation maps. *Nat Genet.* (2016) 48:1303–12. doi: 10.1038/ng.3668

23. Kichaev G, Yang WY, Lindstrom S, Hormozdiari F, Eskin E, Price AL, et al. Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *PLoS Genet.* (2014) 10:e1004722. doi: 10.1371/journal.pgen.1004722

24. Servin B, Stephens M. Imputation-based analysis of association studies: candidate regions and quantitative traits. *PLoS Genet.* (2007) 3:e114. doi: 10.1371/journal.pgen.0030114

25. Bottolo L, Chadeau-Hyam M, Hastie DI, Zeller T, Liquet B, Newcombe P, et al. GUESSing polygenic associations with multiple phenotypes using a gpu-based evolutionary stochastic search algorithm. *PLoS Genet.* (2013) 9:e1003657. doi: 10.1371/journal.pgen.1003657

26. Benner C, Spencer CCA, Havulinna AS, Salomaa V, Ripatti S, Pirinen M. FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* (2016) 32:1493–501. doi: 10.1093/bioinformatics/btw018

27. Chen W, Larrabee BR, Ovsyannikova IG, Kennedy RB, Haralambieva IH, Poland GA et al. Fine mapping causal variants with an approximate Bayesian method using marginal test statistics. *Genetics* (2015). 200:719–36. doi: 10.1534/genetics.115.176107

28. Kichaev G, Roytman M, Johnson R, Eskin E, Lindstrom S, Kraft P, et al. Improved methods for multi-trait fine mapping of pleiotropic risk loci. *Bioinformatics* (2017) 33:248–55. doi: 10.1093/bioinformatics/btw615

29. Wu C, Pan W. Integration of enhancer-promoter interactions with GWAS summary results identifies novel schizophrenia-associated genes and pathways. *Genetics* (2018) 209:699–709. doi: 10.1534/genetics.118.300805

30. Schaid DJ, Chen W, Larson NB. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat Rev Genet.* (2018) 19:1–14. doi: 10.1038/s41576-018-0016-z

31. Wu Y, Zeng J, Zhang F, Zhu Z, Qi T, Zheng Z, et al. Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. *Nat Commun.* (2018) 9:918. doi: 10.1038/s41467-018-03371-0

32. Myocardial Infarction Genetics and CARDIoGRAM Exome Consortia Investigators, Stitziel NO, Stirrups KE, Masca NG, Erdmann J, Ferrario PG, et al. Coding variation in ANGPTL4, LPL and SVEP1 and the risk of coronary disease. *N Eng J Med.* (2016) 373:1134–44. doi: 10.1056/NEJMoa1507652

33. Ridker PM, Revkin J, Amarenco P, Brunell R, Curto M, Civeira F, et al. Cardiovascular efficacy and safety of bococizumab in high-risk patients. *N Engl J Med.* (2017) 376:1527–39. doi: 10.1056/NEJMoa1701488

34. Ference BA, Robinson JG, Brook RD, Catapano AL, Chapman MJ, Neff DR, et al. Variation in *PCSK9* and *HMGCR* and risk of cardiovascular disease and diabetes. *N Engl J Med.* (2016) 375:2144–53. doi: 10.1056/NEJMoa1604304

35. Claussnitzer M, Dankel SN, Kim K-H, Quon G, Meuleman W, Haugen C, et al. *FTO* obesity variant circuitry and adipocyte browning in humans. *N Engl J Med.* (2015) 373:895–907. doi: 10.1056/NEJMoa1502214

36. Kessler T, Wobst J, Wolf B, Eckhold J, Vilne B, Hollstein R, et al. Functional characterization of the GUCY1A3 coronary artery disease risk locus. *Circulation* (2017) 136:476–89. doi: 10.1161/CIRCULATIONAHA.116.024152

37. Prokop JW, Yeo NC, Ottmann C, Chhetri SB, Florus KL, Ross EJ, et al. Characterization of coding/noncoding variants for SHROOM3 in patients with CKD. *J Am Soc Nephrol.* (2018) 29:1525–35. doi: 10.1681/ASN.2017080856

38. Nguyen KD, Qiu Y, Cui X, Goh YPS, Mwangi J, Mukundan L, et al. HHS Public Access. (2012) 480:104–8. Available online at: http://yunliweb.its.unc.edu/HUGIn

39. Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet.* (2013) 45:1238–43. doi: 10.1038/ng.2756

40. Krämer A, Green J, Pollard J Jr, Tugendreich S. Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics* (2014) 30:523–30. doi: 10.1093/bioinformatics/btt703

41. Astle WJ, Elding H, Jiang T, Allen D, Ruklisa D, Mann AL, et al. The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell* (2016) 167:1415–29.e19. doi: 10.1016/j.cell.2016.10.042

42. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet.* (2018) 50:693–8. doi: 10.1038/s41588-018-0099-7

43. Pertel T, Hausmann S, Morger D, Züger S, Guerra J, Lascano J, et al. TRIM5 is an innate immune sensor for the retrovirus capsid lattice. *Nature* (2011) 472:361–5. doi: 10.1038/nature09976

44. Chang T-H, Yoshimi R, Ozato K. Tripartite Motif (TRIM) 12c, a mouse homolog of TRIM5, is a ubiquitin ligase that stimulates type I IFN and NF-κB pathways along with TNFR-associated factor 6. *J Immunol.* (2015) 195:5367–79. doi: 10.4049/jimmunol.1402064

45. Ralston JC, Lyons CL, Kennedy EB, Kirwan AM, Roche HM. Fatty acids and NLRP3 inflammasome–mediated inflammation in metabolic tissues. *Annu Rev Nutr.* (2017) 37:77–102. doi: 10.1146/annurev-nutr-071816-064836

46. van Diepen JA, Thiem K, Stienstra R, Riksen NP, Tack CJ, Netea MG. Diabetes propels the risk for cardiovascular disease: sweet monocytes becoming aggressive? *Cell Mol Life Sci.* (2016) 73:4675–84. doi: 10.1007/s00018-016-2316-9

47. Kratz M, Coats BR, Hisert KB, Hagman D, Mutskov V, Peris E, et al. Metabolic dysfunction drives a mechanistically distinct proinflammatory phenotype in adipose tissue macrophages. *Cell Metab.* (2014) 20:614–25. doi: 10.1016/j.cmet.2014.08.010

48. Genecards. *Olfactory Receptor Family*. Available online at: https://www.genecards.org/Search/Keyword?queryString=olfactory%20 (accessed April 9, 2018).

49. Riera CE, Tsaousidou E, Halloran J, Follett P, Hahn O, Pereira MMA, et al. The sense of smell impacts metabolic health and obesity. *Cell Metab.* (2017) 26:198–211.e5. doi: 10.1016/j.cmet.2017.06.015

50. Grarup N, Moltke I, Andersen MK, Dalby M, Vitting-Seerup K, Kern T, et al. Loss-of-function variants in ADCY3 increase risk of obesity and type 2 diabetes. *Nat Genet.* (2018) 50:172–4. doi: 10.1038/s41588-017-0022-7

51. Siljee JE, Wang Y, Bernard AA, Ersoy BA, Zhang S, Marley A, et al. Subcellular localization of MC4R with ADCY3 at neuronal primary cilia underlies a common pathway for genetic predisposition to obesity. *Nat Genet.* (2018) 50:180–5. doi: 10.1038/s41588-017-0020-9

52. Ridker PM, Everett BM, Thuren T, MacFadyen JG, Chang WH, Ballantyne C, et al. Antiinflammatory therapy with canakinumab for atherosclerotic disease. *N Engl J Med.* (2017) 377:1119–31. doi: 10.1056/NEJMoa1707914

53. Crean D, Godson C. Specialised lipid mediators and their targets. *Semin Immunol.* (2015) 27:169–76. doi: 10.1016/j.smim.2015.05.002

54. Brennan EP, Mohan M, McClelland A, de Gaetano M, Tikellis C, Marai M, et al. Lipoxins protect against inflammation in diabetes-associated atherosclerosis. *Diabetes* (2018) 67:db171317. doi: 10.2337/db17-1317

55. Börgeson E, Johnson AMF, Lee YS, Till A, Syed GH, Ali-Shah ST, et al. Lipoxin A4 attenuates obesity-induced adipose inflammation and associated liver and kidney disease. *Cell Metab.* (2015) 22:125–37. doi: 10.1016/j.cmet.2015.05.003

56. O'Reilly M, Dillon E, Guo W, Finucane O, McMorrow A, Murphy A, et al. High-density lipoprotein proteomic composition, and not efflux capacity, reflects differential modulation of reverse cholesterol transport by saturated and monounsaturated fat diets. *Circulation* (2016) 133:1838–50. doi: 10.1161/CIRCULATIONAHA.115.020278

57. Connaughton RM, McMorrow AM, McGillicuddy FC, Lithander FE, Roche HM. Impact of anti-inflammatory nutrients on obesity-associated metabolic-inflammation from childhood through to adulthood. *Proc Nutr Soc.* (2016) 75:115–24. doi: 10.1017/S0029665116000070

58. Tominaga K, Kondo C, Johmura Y, Nishizuka M, Imagawa M. The novel gene fad104, containing a fibronectin type III domain, has a significant role in adipogenesis. *FEBS Lett.* (2004) 577:49–54. doi: 10.1016/j.febslet.2004.09.062

59. Kishimoto K, Kato A, Osada S, Nishizuka M, Imagawa M. Fad104, a positive regulator of adipogenesis, negatively regulates osteoblast differentiation. *Biochem Biophys Res Commun.* (2010) 397:187–91. doi: 10.1016/j.bbrc.2010.05.077

60. Peng Y, Li H, Li X, Yu S, Xiang H, Peng J, et al. MicroRNA-215 impairs adipocyte differentiation and co-represses FNDC3B and CTNNBIP1. *Int J Biochem Cell Biol*. (2016) 79:104–12. doi: 10.1016/j.biocel.2016.08.014

61. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet*. (2014) 46:1173–86. doi: 10.1038/ng.3097

62. Shungin D, Winkler T, Croteau-Chonka D, Ferreira T, Locke A, Mägi R, et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature* (2015) 518:187–96. doi: 10.1038/nature14132

63. den Hoed M, Eijgelsheim M, Esko T, Brundel BJ, Peal DS, Evans DM, et al. Identification of heart rate associated loci and their effects on cardiac conduction and rhythm disorders. *Nat Genet*. (2013) 45:621–31. doi: 10.1038/ng.2610

64. Choquet H, Thai KK, Yin J, Hoffmann TJ, Kvale MN, Banda Y, et al. A large multi-ethnic genome-wide association study identifies novel genetic loci for intraocular pressure. *Nat Commun*. (2017) 8:2108. doi: 10.1038/s41467-017-01913-6

65. Gao XR, Huang H, Nannini DR, Fan F, Kim H. Genome-wide association analyses identify new loci influencing intraocular pressure. *Hum Mol Genet*. (2018) 27:2205–13. doi: 10.1093/hmg/ddy111

66. Fox CS, Liu Y, White CC, Feitosa M, Smith AV, Heard-Costa N, et al. Genome-wide association for abdominal subcutaneous and visceral adipose reveals a novel locus for visceral fat in women. *PLoS Genet*. (2012) 8:e1002695. doi: 10.1371/journal.pgen.1002695

67. Lu Y, Day FR, Gustafsson S, Buchkovich ML, Na J, Bataille V, et al. New loci for body fat percentage reveal link between adiposity and cardiometabolic disease risk. *Nat Commun*. (2016) 7:10495. doi: 10.1038/ncomms10495

68. Kishimoto K, Nishizuka M, Katoh D, Kato A, Osada S, Imagawa M. FAD104, a regulatory factor of adipogenesis, acts as a novel regulator of calvarial bone formation. *J Biol Chem*. (2013) 288:31772–83. doi: 10.1074/jbc.M113.452961

69. Wang B, Wood IS, Trayhurn P. Hypoxia induces leptin gene expression and secretion in human preadipocytes: differential effects of hypoxia on adipokine expression by preadipocytes. *J Endocrinol*. (2008) 198:127–34. doi: 10.1677/JOE-08-0156

70. Schulz TJ, Huang TL, Tran TT, Zhang H, Townsend KL, Shadrach JL, et al. Identification of inducible brown adipocyte progenitors residing in skeletal muscle and white fat. *Proc Natl Acad Sci USA*. (2011) 108:143–8. doi: 10.1073/pnas.1010929108

71. Gustafson B, Hammarstedt A, Hedjazifar S, Hoffmann JM, Svensson PA, Grimsby J, et al. BMP4 and BMP antagonists regulate human white and beige adipogenesis. *Diabetes* (2015) 64:1670–81. doi: 10.2337/db14-1127

72. Boström P, Wu J, Jedrychowski MP, Korde A, Ye L, Lo JC, et al. A PGC1a dependent myokine that derives browning of white fat and thermogenesis. *Nature* (2012) 481:463–8. doi: 10.1038/nature10777

73. Pérez-Sotelo D, Roca-Rivada A, Baamonde I, Baltar J, Castro AI, Domínguez E, et al. Lack of adipocyte-Fndc5/irisin expression and secretion reduces thermogenesis and enhances adipogenesis. *Sci Rep*. (2017) 7:16289. doi: 10.1038/s41598-017-16602-z

74. van Marken Lichtenbelt WD. Cold-activated brown adipose tissue in healthy men. *Yearb Endocrinol*. (2010) 2010:115–7. doi: 10.1056/NEJMoa0808718

75. Pan D, Fujimoto M, Lopes A, Wang YX. Twist-1 is a PPARδ-inducible, negative-feedback regulator of PGC-1α in brown fat metabolism. *Cell* (2009) 137:73–86. doi: 10.1016/j.cell.2009.01.051

76. Franzen O, Ermel R, Cohain A, Akers N, Di Narzo A, Talukdar HA, et al. Cardiometabolic risk loci share downstream cis- and trans-gene regulation across tissue and diseases. The Stockholm-Tartu Atherosclerosis Reverse Network Engineering Task (STARNET) study. *Science* (2016) 353:827–31. doi: 10.1126/science.aad6970

77. Liquori CL, Berg MJ, Siegel AM, Huang E, Zawistowski JS, Stoffer T, et al. Mutations in a gene encoding a novel protein containing a phosphotyrosine-binding domain cause type 2 cerebral cavernous malformations. *Am J Hum Genet*. (2003) 73:1459–64. doi: 10.1086/380314

78. Fischer A, Zalvide J, Faurobert E, Albiges-Rizo C, Tournier-Lasserve E. Cerebral cavernous malformations: from CCM genes to endothelial cell homeostasis. *Trends Mol Med*. (2013) 19:302–8. doi: 10.1016/j.molmed.2013.02.004

79. Crose LES, Hilder TL, Sciaky N, Johnson GL. Cerebral cavernous malformation 2 protein promotes smad ubiquitin regulatory factor 1-mediated RhoA degradation inendothelial cells. *J Biol Chem*. (2009) 284:13301–5. doi: 10.1074/jbc.C900009200

80. DiStefano PV, Kuebel JM, Sarelius IH, Gladding AJ. KRIT1 protein depletion modifies endothelial cell behavior via increased vascular endothelial growth factor (VEGF) Signaling. *J Biol Chem*. (2014) 289:33054–65. doi: 10.1074/jbc.M114.582304

81. Draheim KM, Li X, Zhang R, Fisher OS, Villari G, Boggon TJ, et al. CCM2-CCM3 interaction stabilizes their protein expression and permits endothelial network formation. *J Cell Biol*. (2015) 208:987. doi: 10.1083/jcb.201407129

82. Whitehead KJ, Chan AC, Navankasattusas S, Koh W, London NR, Ling J, et al. The cerebral cavernous malformation signaling pathway promotes vascular integrity via Rho GTPases. *Nat Med*. (2009) 15:177–84. doi: 10.1038/nm.1911

83. Erdmann J, Stark K, Esslinger UB, Rumpf PM, Koesling D, DeWit C, et al. Dysfunctional nitric oxide signalling increases risk of myocardial infarction. *Nature* (2013) 504:432–6. doi: 10.1038/nature12722

84. Klarin D, Zhu QM, Emdin CA, Chaffin M, Horner S, McMillan BJ, et al. Genetic analysis in UK Biobank links insulin resistance and transendothelial migration pathways to coronary artery disease. *Nat Genet*. (2017) 49:1392–7. doi: 10.1038/ng.3914

85. Boulday G, Blecon A, Petit N, Chareyre F, Garcia LA, Niwa-Kawakita M, et al. Tissue-specific conditional CCM2 knockout mice establish the essential role of endothelial CCM2 in angiogenesis: implications for human cerebral cavernous malformations. *Dis Model Mech*. (2009) 2:168–77. doi: 10.1242/dmm.001263

86. Cao Y. (2010). Adipose tissue angiogenesis as a therapeutic target for obesity and metabolic diseases. *Nat Rev Drug Discov*. (2010) 9:107–15. doi: 10.1038/nrd3055

87. Hara Y, Wakino S, Tanabe Y, Saito M, Tokuyama H, Washida N, et al. Rho and Rho-kinase activity in adipocytes contributes to a vicious cycle in obesity that may involve mechanical stretch. *Sci Signal*. (2011) 4:ra3. doi: 10.1126/scisignal.2001227

88. Lancaster GI, Langley KG, Berglund NA, Kammoun HL, Reibe S, Estevez E, et al. Evidence that TLR4 is not a receptor for saturated fatty acids but mediates lipid-induced inflammation by reprogramming macrophage metabolism. *Cell Metab*. (2018) 27:1096–110.e5. doi: 10.1016/j.cmet.2018.03.014

89. Reynolds CM, Mcgillicuddy FC, Harford KA, Finucane OM, Mills KHG, Roche HM. Dietary saturated fatty acids prime the NLRP3 inflammasome via TLR4 in dendritic cells-implications for diet-induced insulin resistance. *Mol Nutr Food Res*. (2012) 56:1212–22. doi: 10.1002/mnfr.201200058

90. Wen H, Gris D, Lei Y, Jha S, Zhang L, Huang MTH, et al. Fatty acid-induced NLRP3-ASC inflammasome activation interferes with insulin signaling. *Nat Immunol*. (2011) 12:408–15. doi: 10.1038/ni.2022

91. McGillicuddy FC, Harford KA, Reynolds CM, Oliver E, Claessens M, Mills KHG, et al. Lack of interleukin-1 receptor I (IL-1RI) protects mice from high-fat diet-induced adipose tissue inflammation coincident with improved glucose homeostasis. *Diabetes* (2011) 60:1688–98. doi: 10.2337/db10-1278

92. Isakson P, Hammarstedt A, Gustafson B, Smith U. Role of Wnt, Tumor Necrosis Factor-α, and Inflammation. *Diabetes* (2009) 58:1550–7. doi: 10.2337/db08-1770

93. Vidal C, Bermeo S, Li W, Huang D, Kremer R, Duque G. Interferon gamma inhibits adipogenesis in vitro and prevents marrow fat infiltration in oophorectomized mice. *Stem Cells* (2012) 30:1042–8. doi: 10.1002/stem.1063

94. Zuniga LA, Shen W-J, Joyce-Shaikh B, Pyatnova EA, Richards AG, Thom C, et al. IL-17 regulates adipogenesis, glucose homeostasis, and obesity. *J Immunol*. (2010) 185:6947–59. doi: 10.4049/jimmunol.1001269

95. Rayalam S, Yang JY, Ambati S, Della-Fera MA, Baile CA. Resveratrol induces apoptosis and inhibits adipogenesis in 3T3-L1 adipocytes. *Pytother Res*. (2008) 22:1367–71. doi: 10.1002/ptr.2503

96. Kim GS, Park HJ, Woo JH, Kim MK, Koh PO, Min W, et al. *Citrus aurantium* flavonoids inhibit adipogenesis through the Akt signaling pathway in 3T3-L1 cells. *BMC Complement Altern Med*. (2012) 12:31. doi: 10.1186/1472-6882-12-31

97. Ejaz A, Wu D, Kwan P, Meydani M. Curcumin inhibits adipogenesis in 3T3-L1 adipocytes and angiogenesis and obesity in C57/BL mice. *J Nutr.* (2009) 139:919–25. doi: 10.3945/jn.108.100966

98. Lyons CL, Kennedy EB, Roche HM. Metabolic inflammation-differential modulation by dietary constituents. *Nutrients* (2016) 8:E247. doi: 10.3390/nu8050247

99. Li SX, Imamura F, Ye Z, Schulze MB, Zheng J, Ardanaz E, et al. Interaction between genes and macronutrient intake on the risk of developing type 2 diabetes; systematic review and findings from European Prospective Investigation into Cancer (EPIC)-InterAct. *Am J Clin Nutr* (2017) 5:263–75. doi: 10.3945/ajcn.116.150094

100. Shungin D, Deng WQ, Varga TV, Luan J, Mihailov E, Metspalu A, et al. Ranking and characterization of established BMI and lipid associated loci as candidates for gene-environment interactions. *PLoS Genet.* (2017) 13:e1006812. doi: 10.1371/journal.pgen.1006812

101. Lai C-Q, Smith CE, Parnell LD, Lee Y-C, Corella D, Hopkins P, et al. Epigenomics and metabolomics reveal the mechanism of the APOA2-saturated fat intake interaction affecting obesity. *Am J Clin Nutr.* (2018) 108:188–200. doi: 10.1093/ajcn/nqy081

102. Weng LC, Guan W, Steffen LM, Pankow JS, Pankratz N, Chen MH, et al. Pleiotropic effects of n-6 and n-3 fatty acid-related genetic variants on circulating hemostatic variables. *Thromb Res.* (2018) 168:53–9. doi: 10.1016/j.thromres.2018.05.032

103. Bartel J, Krumsiek J, Schramm K, Adamski J, Gieger C, Herder C, et al. The human blood metabolome-transcriptome interface. *PLoS Genet.* (2015) 11:e1005274. doi: 10.1371/journal.pgen.1005274

104. Nath AP, Ritchie SC, Byars SG, Fearnley LG, Havulinna AS, Joensuu A, et al. An interaction map of circulating metabolites, immune gene networks, and their genetic regulation. *Genome Biol.* (2017) 18:146. doi: 10.1186/s13059-017-1279-y

105. Gepner Y, Shelef I, Schwarzfuchs D, Zelicha H, Tene L, Meir AY, et al. Effect of distinct lifestyle interventions on mobilization of fat storage pools: CENTRAL magnetic resonance imaging randomized controlled trial. *Circulation* (2018) 137:1143–57. doi: 10.1161/CIRCULATIONAHA.117.030501

106. Ongen H, Brown AA, Delaneau O, Panousis NI, Nica AC, Dermitzakis ET. Estimating the causal tissues for complex traits and diseases. *Nat Genet.* (2017) 49:1676–83. doi: 10.1038/ng.3981

107. Shu L, Chan KHK, Zhang G, Huan T, Kurt Z, Zhao Y, et al. Shared genetic regulatory networks for cardiovascular disease and type 2 diabetes in multiple populations of diverse ethnicities in the United States. *PLoS Genet.* (2017) 13:1–25. doi: 10.1371/journal.pgen.1007040

108. Baillie JK, Bretherick A, Haley CS, Clohisey S, Gray A, Neyton LPA, et al. Shared activity patterns arising at genetic susceptibility loci reveal underlying genomic and cellular architecture of human disease. *PLoS Comput Biol.* (2018) 14:1–24. doi: 10.1371/journal.pcbi.1005934

109. Fagny M, Paulson JN, Kuijjer ML, Sonawane AR, Chen C-Y, Lopes-Ramos CM, et al. Exploring regulation in tissues with eQTL networks. *Proc Natl Acad Sci USA.* (2017) 114:E7841–50. doi: 10.1073/pnas.1707375114

110. Saha A, Kim Y, Gewirtz ADH, Jo B, Gao C, McDowell IC, et al. Co-expression networks reveal the tissue-specific regulation of transcription and splicing. *Genome Res.* (2017) 1843–58. doi: 10.1101/gr.216721.116

111. Sonawane AR, Platig J, Fagny M, Chen CY, Paulson JN, Lopes-Ramos CM, et al. Understanding tissue-specific gene regulation. *Cell Rep.* (2017) 21:1077–88. doi: 10.1016/j.celrep.2017.10.001

112. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* (2015) 15:453–7. doi: 10.1038/nmeth.3337

113. Brown AA, Viñuela A, Delaneau O, Spector TD, Small KS, Dermitzakis ET. Predicting causal variants affecting expression by using whole-genome sequencing and RNA-seq from multiple human tissues. *Nat Genet.* (2017) 49:1747–51. doi: 10.1038/ng.3979

114. Park J, Shrestha R, Qiu C, Kondo A, Huang S, Werth M, et al. Single-cell transcriptomics of the mouse kidney reveals potential cellular targets of kidney disease. *Science* (2018) 360:758–63. doi: 10.1126/science.aar2131

115. van der Wijst MG, Brugge H, de Vries DH, Deelen P, Swertz MA, Franke L. Single-cell RNA sequencing identifies celltype-specific cis-eQTLs and co-expression QTLs. *Nat Genet.* (2018) 50:493–7. doi: 10.1038/s41588-018-0089-9

116. Macaulay IC, Ponting CP, Voet T. Single-cell multiomics: multiple measurements from single cells. *Trends Genet.* (2017) 33:155–68. doi: 10.1016/j.tig.2016.12.003

117. Syed I, Lee J, Moraes-Vieira PM, Donaldson CJ, Sontheimer A, Aryal P, et al. Palmitic acid hydroxystearic acids activate GPR40, which is involved in their beneficial effects on glucose homeostasis. *Cell Metab.* (2018) 27:419–427.e4. doi: 10.1016/j.cmet.2018.01.001

Long Noncoding RNA *ANRIL*: Lnc-ing Genetic Variation at the Chromosome 9p21 Locus to Molecular Mechanisms of Atherosclerosis

Lesca M. Holdt and Daniel Teupser*

Institute of Laboratory Medicine, University Hospital, LMU Munich, Munich, Germany

Ever since the first genome-wide association studies (GWAS) on coronary artery disease (CAD), the Chr9p21 risk locus has emerged as a top signal in GWAS of atherosclerotic cardiovascular disease, including stroke and peripheral artery disease. The CAD risk SNPs on Chr9p21 lie within a stretch of 58 kilobases of non-protein-coding DNA, containing the gene body of the long noncoding RNA (lncRNA) *antisense non coding RNA in the INK4 locus* (*ANRIL*). How risk is affected by the Chr9p21 locus in molecular detail is a matter of ongoing research. Here we will review recent advances in the understanding that *ANRIL* serves as a key risk effector molecule of atherogenesis at the locus. One focus of this review is the shift in understanding that genetic variation at Chr9p21 not only affects the abundance of *ANRIL*, and in some cases expression of the adjacent *CDKN2A/B* tumor suppressors, but also impacts *ANRIL* splicing, such that 3′-5′-linked circular noncoding *ANRIL* RNA species are produced. We describe how the balance of linear and circular *ANRIL* RNA, determined by the Chr9p21 genotype, regulates molecular pathways and cellular functions involved in atherogenesis. We end with an outlook on how manipulating circular *ANRIL* abundance may be exploited for therapeutic purposes.

Keywords: lncRNA (long non-coding RNA), circRNA, GWAS (genome-wide association study), eQTL analysis, transcription, splicing, tumor suppressor proteins, cardiovascular diseases

INTRODUCTION

Since publication of the first genome-wide association studies (GWAS) of coronary artery disease (CAD) in 2007, Chr9p21 has emerged as the most significant risk locus associated with this frequent disease (1–4). The region contains a number of strongly interlinked SNPs within a stretch of 58 kilobases (kb) of non-protein-coding DNA. Later, the same haplotype block has been associated with other endpoints of atherosclerosis, such as stroke (5–11), peripheral artery disease (12–14), and also with different types of aneurysms (2, 8, 15, 16). Due to the availability of large study cohorts and the better resolution of genetic recombination in this region, it has now become clear that associations with other phenotypes at Chr9p21 fall in distinct haplotype blocks not overlapping with the CAD block (**Figure 1A**). Closely nearby, and proximal to the CAD locus, GWAS found associations with cancer, such as melanoma, glioma, basal cell carcinoma, and acute lymphoblastic leukemia [see (40) for review], and also with glaucoma, and diverse proliferative or

inflammatory diseases, such as endometriosis of the reproductive tract (41), periodontitis (42), and platelet reactivity (43). The region located distally to the CAD region contains a distinct haplotype block associated with type 2 diabetes (44, 45).

In the last 10 years, GWAS have been successfully used to increase the number of genetic loci implicated in CAD risk inheritance. The number of CAD risk loci in the genome rose from 56 by 2013 (24, 46–52) to 80 by 2015 (53–56), to 243 by 2017 (17). Concerning the Chr9p21 locus in these studies, the association rose steadily from $p = 5.40 \times 10^{-23}$ (rs4977575) (57), over $p = 4.68^{-101}$ (rs4977574) (17) to $p = 8.8 \times 10^{-223}$ (rs4977574) (58). In populations of European descent, the allele frequency is very high (0.48), leading to the situation that approximately one-fourth of people are homozygous for the CAD risk alleles. CAD risk SNPs on Chr9p21 have recurrently been shown to have one of the top-ranking effect sizes [allele-specific odds ratio (OR) for CAD > 1.3] (3, 24). Despite the extent of effects, the Chr9p21 risk is independent of classically known CAD risk determinants, such as dyslipidemia, diabetes mellitus, age, and sex.

The Chr9p21 region contains at least 5 genes, which are, in part, tightly clustered and overlapping. These include the 3.8 kb long *ANRIL* non-coding RNA, and the tumor suppressors *cyclin dependent kinase inhibitor CDKN2A/p16^INK4A*, *CDKN2A/p14^ARF*, *CDKN2B/p15^INK4B*, and *methylthioadenosine phosphorylase (MTAP)*. *ANRIL* overlaps in antisense the full length of the *p15* gene body, while sharing a bidirectional promoter with *CDKN2A*. Hence, it was also termed *CDKN2B antisense RNA (CDKN2B-AS1)*. Only recently, the picture got even more complex: Advances in high-throughput sequencing and adaptions in bioinformatics mapping of RNA reads to reference genomes have revealed that thousands of genes in our genome produce not only mature linear RNA but also $3'$-$5'$ covalently linked circular RNAs (circRNAs) (59). So far, two studies have shown that a number of circular *ANRIL* (*circANRIL*) isoforms exist, comprised of different exons, whereby a downstream exon is fused to an upstream exon by the enzymatic activity of the spliceosome in a reaction termed "backsplicing" [see (60, 61) for review]. Circularizing exons in *ANRIL* stemmed mostly from middle parts of the lncRNA (**Figure 1A**), which are in part also shared by the linear *ANRIL* isoforms. *CircANRIL* was found not only in many different cell lines, but also in many primary cell types, including vascular smooth muscle cells (VSMCs) and macrophages, as well as in heart and vascular tissue (22, 36).

A major focus in exploring how risk is effected by Chr9p21 has been on whether genetic variation affected expression of genes at the locus *in cis* (**Figure 1A**) or whether it elicited gene expression changes *in trans*. Top CAD-associated SNPs lie within the distal parts of long linear *ANRIL* isoforms (**Figure 1A**) and several studies have shown that they co-localize with sequences marked by chromatin modifications, RNA polymerase II transcription patterns and DNA motifs characteristic of *bona-fide* transcriptional enhancers (19, 35, 62–65). Using expression quantitative trait locus (eQTL) analyses in patient samples, several groups have by now investigated if the risk alleles at the locus were associated with the expression of specific

target genes *in cis* (cis-eQTLs). Whereas studies investigating *ANRIL* expression have mostly used quantitative PCRs (qPCRs) targeting different exons from the lncRNA, expression of *p14, p15, p16,* or *MTAP* has either been investigated using genome-wide expression arrays or isoform-specific qPCRs. Here, we focus on studies investigating eQTLs in atherosclerosis cohorts but do not cover studies related to other phenotypes, such as cancer, which are reviewed elsewhere (66).

CIS-eQTLs AT Chr9p21

ANRIL expression at Chr9p21 is complex and at least 20 linear isoforms as well as multiple circular isoforms have been reported [www.ensembl.org, (22, 36, 39)]. In principal, linear and circular isoforms can be distinguished by the fact that the latter derive from a backsplice event, where splicing of a downstream exon (e.g., exon 7) to an upstream exon (e.g., to exon 5) can be detected. Backsplicing of ex7-5 was the most common event observed in our own study in peripheral blood monocytes (36). Concordantly, Burd and colleagues have reported dominant backsplice isoforms spanning ex14-4 in peripheral blood T lymphocytes (22). In both studies, exon 1 and exons 17-20 were not contained in circularized *ANRIL* (**Table 1**). Thus, for classification reasons, results from studies targeting these exons will be referred to as proximal linear isoforms (containing the first *ANRIL* exons) and long linear isoforms (containing the distal exons 17-20) (**Table 1**). Since both linear and circular *ANRIL* may contain exons from the middle portion of the lncRNA (e.g., exons 4-16), a clear distinction as to whether linear or circular isoforms were investigated cannot be made in cases where these exons were targeted by qPCRs which were non-specific for backsplice junctions (**Table 1**).

As one of the first studies on Chr9p21, Jarinova et al. have shown that *ANRIL* expression was induced by the CAD risk SNP rs1333049 in peripheral blood monocytes (PBMCs). No significant effects on *CDKN2A* or on *CDKN2B* were recorded in that study (19). Over the years, comparable quantifications of these genes followed in whole blood, peripheral blood T lymphocytes, lymphoblastoid cells lines, aortic smooth muscle cells (SMCs) and in different tissue samples that are known to have a role in atherosclerosis. For example, vascular tissues such as carotid atherosclerotic plaque samples, samples from aorta, mammary artery, and from the heart ventricles have been analyzed, but also tissues like subcutaneous or omental fat have been used (**Table 1**). Of the 23 cis-eQTL studies conducted in the Chr9p21 CAD region to date, 16 investigated different isoforms of *ANRIL*, out of which 10 used assays targeting proximal *ANRIL* exons, 8 used assays targeting the middle region, 6 used assays targeting downstream linear *ANRIL* exons, and two investigated backsplices contained in *circANRIL* (**Table 1**). Complicating a clear-cut interpretation, in the different studies, different risk genotypes were used to indicate risk haplotypes. The expression of *CDKN2A* and of *CDKN2B* was investigated in 18 studies and *MTAP* in 10 studies (**Table 1**).

Overall, 80% of the studies investigating *ANRIL* expression found an association with the Chr9p21 genotype. Here, a trend

FIGURE 1 | ANRIL and the CAD risk locus at Chr9p21. **(A)** Genomic mapping of SNPs identified in a region ± 300 kb around the top-ranking sentinel SNP rs4977574 based on data of the most recent large CAD GWAS (17). Chromosome ideogram and zoom-in onto RefSeq transcripts for *ANRIL*, *CDKN2A*, *CDKN2B,* and *MTAP*

(Continued)

toward higher expression of the proximal and distal exons contained in linear *ANRIL* in patients carrying the CAD-risk allele was observed (7 of 10 and 3 of 6 studies). In contrast, circular *ANRIL* was downregulated in the two published studies in patients carrying the Chr9p21 risk haplotype. No clear tendency was observed when assays targeting the middle region of *ANRIL* were used (**Table 1**). This is likely explained by the fact that these assays target both, linear and circular, *ANRIL* isoforms, which seem to be inversely regulated. With respect to the tumor suppressor genes contained at the Chr9p21 locus, 78 and 67% of the studies failed to find an association of *CDKN2A* and *CDKN2B* with Chr9p21, respectively. When reporting an association, specifically *CDKN2B* was down-regulated in the majority of studies (94%), yet its expression was not always anticorrelating with *ANRIL* expression (19, 21, 22, 29). *MTAP* expression was not associated with the Chr9p21 genotype in any of the published studies. Overall, the picture emerges that circular *ANRIL* and *CDKN2B* tend to be down-regulated in patients carrying the risk allele, whereas linear *ANRIL* isoforms tend to be inversely regulated (**Figure 1B**). It is currently unclear, why expression of *p15* or of *p14* and *p16* were in many cases positively correlated with *ANRIL* (19, 21, 22, 27, 29, 32, 34, 65). Also, *MTAP*, which was not associated with Chr9p21 (**Table 1**), was in some conditions anticorrelating to *ANRIL,* but not in all cases or contexts (20, 34, 67). SNPs in *ANRIL* can hypothetically affect enhancers in both directions, either by disrupting transcription factor binding sites in open chromatin (68) or by increasing enhancer activity through yet unknown primary effects (24, 65).

In summary, many studies document *cis*-eQTLs for *ANRIL* or, separately, for *CDKN2B* (35). Throughout, from the existing data, it can be concluded that these effects are cell-type specific and combinatorial. Of note, many studies have investigated only very small cohorts and those, simultaneously testing both *ANRIL* and *CDKN2B* in larger cohorts (>1000 samples) identified much stronger effects of Chr9p21 on *ANRIL* than on *CDKN2B* (13, 33, 36). This observation might be explained by the haplotype block structure of the region, where effects of CAD lead SNPs are located within *ANRIL* but bleed through due to linkage disequilibrium, resulting in more subtle concomitant effects on *CDKN2B* expression. Another possibility is that the Chr9p21 genotype impacts transcription enhancers at the locus which contact and activate gene promoters affecting CAD. The consequences of such contacts would not be expected to be captured through traditional non-allelic RNA expression analysis. In fact, when allelic expression control through 3D-enhancer looping was specifically measured in a separate study in human coronary aortic SMCs (64), physical contacts of CAD

variant-containing enhancers in the locus and the promoters of *CDKN2A, CDKN2B,* and *ANRIL* were corroborated.

Taken together, these data suggest that genetic variation within the core 9p21 CAD region relates to differential expression not only of *ANRIL*, but in specific cells or conditions, also of the *CDKN2A/B* tumor suppressors encoded in the locus. While either of these factors could potentially increase cell proliferation, or lead to unscheduled senescence, or elicit out of context inflammatory signaling, as far as based on work with cells *in vitro*, no study in humans or in mouse models has been able to decisively implicate a downstream effector pathway *in vivo*.

TRANS-eQTLs AT Chr9p21 AND MOLECULAR FUNCTIONS OF *ANRIL* IN TRANSCRIPTIONAL REGULATION

As opposed to *cis* effects, two eQTL studies have so far detected modest and tissue-selective differential expression of dozens of genes associated with Chr9p21 genotype with genome-wide significance (19, 27). Affected genes were from a broad range of classes (*AVPR2, PEAK1, FBLN1, KALRN, DAZL, STAU2, HLA-DQA1, BTNL8, PLEKHA6, TDGF1*) in whole blood (19) and different, non-overlapping gene sets linked to tissue wounding, cell migration and inflammatory response, when analyzing heart tissue, plaques, aortas, and arteries (27).

Other, and in part, larger studies in vascular tissue (20), peripheral blood mononuclear cells (PBMC, n = 2280) (33) and in blood monocytes (n = 1490) (23) reported no significant expression association.

Though not directly comparable, another study showed that in macrophages cultured *in vitro* under stress-bearing IFNγ and LPS stimulation, the CAD risk genotype led to differential up- and downregulation of target genes outside the Chr9p21 locus and yet distinct from the previously mentioned studies (*IL1B, IL12B, CASP5, CCL8, MT1A, MT1E, MUCL1, TNIP3, VCAN, ENPP2, NDP, CD163*) (30). Also *ANRIL* knockdown in cultured cell lines (69–72) and overexpression of linear *ANRIL* affected the expression of non-overlapping gene sets in the genome *in trans* (33, 36).

How *ANRIL* exerts *trans*-regulation is not known, and despite a study that showed a physical interaction of *ANRIL* with promoters of target genes (33), this role is likely not a classical function as enhancer RNA [eRNA (73)], because it involved both up-and down-regulated genes, and was suggested to involve sequence homology (33). In the case of *ANRIL*, *trans*-regulation of target genes was ascribed to an ALU motif

TABLE 1 | Chr9p21 cis-eQTLs in patients.

References	CAD SNPs	Cells or tissues	Sample size	Assay	*ANRIL* 17 linear and dozens of circular isoforms				CDKN2B	CDKN2A		MTAP
					Proximal exons 1–5*	Middle exons 4–16		Distal exons 17–20	p15	p16^INK4A	p14^ARF	
					Linear*	Linear or circular	Circular**	Linear				
(18)	rs10757278	Peripheral blood T cells (healthy)	170	qPCR	n.d.	ex4-5	n.d.	n.d.				
(19)	rs10757272, rs4977574, rs2891168, rs1333048, rs1333049	Whole blood (healthy)	120	qPCR	ex1-5	ex13		n.d.				n.d.
		Whole blood (CAD)	42	Microarray, qPCR	n.d.	ex15-16 / ex13				n.d.		
(20)	rs2891168, rs2383207, rs1333049, rs13333045, rs10757278, rs10757274, rs10116277	Lymphoblastoid cell lines from HapMap	233	Microarray	ex1-3	ex6,7,14	n.d.	ex17-19				
		Vascular tissue	294		ex1-3	ex6,7,14	n.d.	ex17-19				
(13)	SNPs rs10757274, rs2383206, rs2383207, rs10757278; rs10738605	PBMCs (healthy, CAD)	1098	qPCR	ex1-5	ex4-5	n.d.	ex18-19				
		Whole blood (healthy, CAD)	769		ex1-5	ex4-5	n.d.	ex18-19				
		Vascular tissues	41		ex1-5	ex4-5	n.d.	ex18-19				
(21)	rs10757274, rs10757278; rs1333049; rs2383206; rs3217992; rs7044859, rs496892, rs564398 and others	Whole blood (healthy)	487	qPCR	ex1-2	n.d.	n.d.	n.d.	n.d.			n.d.
(22)	rs10757278	Peripheral blood T cells (healthy)	106	qPCR	ex1-2	n.d.	ex14-5 circular	ex18-19	n.d.	n.d.	n.d.	n.d.
(23)	rs1075727	PBMCs (healthy/subclinical CAD)	1490	Microarray	n.d.	n.d.	n.d.	n.d.	n.d.			n.d.
(24)	rs4977574	Omental adipose tissue (healthy, CAD)	2430	Microarray	n.d.	n.d.	n.d.	n.d.		n.d.	n.d.	n.d.
(25)	rs10757274, rs2383206, rs2383207	Vascular tissue (CAD)	57	qPCR	n.d.	n.d.	n.d.	n.d.				
(26)	rs4977574	aortic SMCs (healthy)	79	qPCR	n.d.	n.d.	n.d.	n.d.				
(27)	rs1333049	Heart (healthy/subclinical)	108	Microarray, qPCR	ex1-2	n.d.	n.d.	n.d.		n.d.	n.d.	
		Combined analysis of myocardium, vascular tissues (healthy, CAD)	406	Microarray	ex1-2	n.d.	n.d.	n.d.				

TABLE 1 | Continued

References	CAD SNPs	Cells or tissues	Sample size	Assay	ANRIL 17 linear and dozens of circular isoforms — Proximal exons 1-5* — Linear*	Middle exons 4-16 — Linear or circular	Circular**	Distal exons 17-20 — Linear	CDKN2B — p15	CDKN2A — $p16^{INK4A}$	$p14^{ARF}$	MTAP
(28)	rs10757274	whole blood (healthy, CAD)	205	qPCR	ex1-5 / ex4-5	n.d.	n.d.	ex17-18				n.d.
(29)	rs496892, rs564398, rs6475606, rs1063192, rs10811650, rs10738607 (prox. ANRIL); rs10965235, rs3731217 (dist. ANRIL); rs2383208 (CDKN2B)	Whole blood (healthy, CAD)	57	qPCR	ex1-2	n.d.	n.d.	ex17-18		n.d.	n.d.	n.d.
	rs564398, rs10965224 (short ANRIL); rs2811712, rs7855162 (dist. ANRIL); rs10965235 (CDKN2A)				ex1-2			ex17-18	n.d.		n.d.	n.d.
	rs1333049				ex1-2			ex17-18	n.d.	n.d.	n.d.	n.d.
(30)	rs7044859, rs496892, rs7865618, rs1333049	PBMCs differentiated to macrophages (healthy, CAD)	68	microarray, qPCR	n.d.	n.d.	n.d.	n.d.				
(31)	rs1333049	Primary VSMCs (arterial from umbilical cords, healthy)	69	qPCR, Western, histochemistry	n.d.	ex14-15	n.d.	n.d.		n.d.		n.d.
(32)	rs944799 (ex12-13 ANRIL)	PBMCs (healthy, CAD)	281	qPCR	n.d.	ex12-13	n.d.	n.d.		n.d.	n.d.	n.d.
	rs944799, rs598664, rs1002878 (ANRIL); rs598664 (CDKN2B)				n.d.	ex13-14	n.d.	n.d.		n.d.	n.d.	n.d.
	rs10965228 (ANRIL)	Platelet-rich blood plasma (without PBMCs)			n.d.	ex12-13 / ex13-14	n.d.	n.d.			n.d.	n.d.
(33)	rs10757274, rs2383206, rs2383207, rs10757278	PBMCs (healthy, CAD)	1933	microarray, qPCR	ex 1-5	ex7-13	n.d.	ex18-19				
		Whole blood (healthy, CAD)	960		ex 1-5	ex7b, ex7-13, ex1-5, ex10-13b	n.d.	ex18-19				
		Vascular tissues	193		ex 1-5	ex7b, ex7-13, ex1-5, ex10-13b	n.d.	ex18-19				
(34)	rs10757278, rs10757274, rs2383206, rs1333049, rs4977574	Whole blood; plasma/healthy, CAD)	200	qPCR (blood), ELISA (plasma)	ex1-5	n.d.	n.d.	n.d.	n.d.		n.d.	n.d.

(Continued)

TABLE 1 | Continued

References	CAD SNPs	Cells or tissues	Sample size	Assay	ANRIL 17 linear and dozens of circular isoforms				CDKN2B	CDKN2A		MTAP
					Proximal exons1-5*	Middle exons 4-16		Distal exons 17-20	p15	p16^INK4A^	p14^ARF^	
					Linear*	Linear or circular	Circular**	Linear				
(36)	rs10757274, rs2383206, rs2383207, rs10757278	PBMCs (healthy, CAD)	1933	Microarray, qPCR	n.d.	n.d.	ex7-5 circular	n.d.				
		Whole blood (healthy, CAD)	1933		n.d.	n.d.	ex7-5 circular	n.d.				
		Vascular tissues	218		n.d.	n.d.	ex7-5 circular	n.d.		n.d.	n.d.	n.d.
(37)	rs1537371, rs1333040	Vascular tissues, fat, muscle, blood (CAD)	600	RNAseq	n.d.	n.d.			#	n.d.	n.d.	n.d.
(18)	rs10965215, rs10738605	PBMCs (healthy, CAD)	66	qPCR	n.d.	ex7	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
(17)	rs1537371	aorta (CAD)	600	RNAseq	n.d.	n.d.	ex7-5 circular	n.d.		n.d.	n.d.	n.d.
		Total number of relevant studies			10	8	2	6	19	18		10
		% of studies showing upregulation			70	37	0	50	11	6		0
		% of studies showing downregulation			30	50	100	17	47	17		0
		% of studies showing unchanged expression			40	63	0	67	63	78		100

CAD-relevant SNPs encoded in Chr9p21 and defined as cis-eQTLs for regulation of ANRIL, p15, p16, p14, and/or MTAP through gene expression profiling with microarrays or targeted qPCR analysis.

Upregulation of target gene expression (green), downregulation (red), no effect on RNA abundance (gray), effect not determined (n.d.). Effects on different isoforms of ANRIL (linear or circular) is specified when exon-specific qPCR analysis was performed targeting proximal, middle or distal ANRIL exons (ex). In total 23 studies are reviewed here. References: (13, 17–39). Note that the percentages (%) of studies showing up- and down-regulation of ANRIL isoforms do not necessarily add up to 100%, because different ANRIL exons were quantified in different studies, not all classes of ANRIL transcripts were analyzed in each study, and because one study can report on both up- and downregulation of different isoforms belonging to the same class of ANRIL transcripts (proximal/middle/distal).

*(ANRIL linearity determined by PCR forward primer residing in exons 1, 2, or 3). ** (ANRIL circularity determined by PCR primers detecting backsplicing between exons 10-2, 5-intron3, 6-intron3, 6-4, 7-4, 14-4, 10-4, 12-4, 13-4, 14-4, 16-4, 6-5, 7-5, 8-5, 10-5, 14-5 16-5, 19-5, 7-6, 10-6, 14-6, 16-13*, 16-15) (22, 36, 39).

(cis-eQTL locating in enhancer element, but with unspecified direction of effect on expression).

in both *ANRIL* and the target gene promoters (33). Similarly, an independent study found that *ANRIL* did not only silence its targets, but unexpectedly also upregulated target genes: For example, proinflammatory interleukins *IL6/8* were found to be co-stimulated by *ANRIL* and YY1, a transcription-regulating factor that bound to the *ANRIL RNA,* especially in the context of TNFα/NFκB signaling (70). Therefore, opposite to what could have been expected from the reported physical interaction of *ANRIL* with proteins from the repressive Polycomb group complexes (74), *ANRIL* might be an activator, at least for some *trans*-regulated genes (33, 70) (see chapter 4 for details). Whether *circANRIL,* beyond regulating rRNA maturation, is involved in primary transcriptional control, alone or via impacting linear *ANRIL*'s function, is not known (36). Nevertheless, it is interesting to note that *circANRIL* isoforms linked to CAD are produced from exons located in the middle of the *ANRIL* gene (22, 36), and as such do not include the ALU motif, which is important for gene *trans*-regulation by linear *ANRIL* and is located more distally in the gene (33). Thus, variation in *ANRIL* RNA at the molecular level (linear vs. circular) might impose a fundamental alteration in *ANRIL* effector function, while not offering any explanation *per se* on how linear *ANRIL* regulates genes, as scaffold for promoter-activating complexes, or as decoy/inhibitor of repressive chromatin-modifying complexes. Conservatively speaking, it seems possible that Chr9p21 CAD risk genotypes affects genomic expression both *in cis* and *in trans*, and linear *ANRIL* RNA may be one, but not the sole, important effector molecule for how the Chr9p21 locus transduces such effects (**Figure 1B**).

CORRELATION OF CHR9P21 GENES WITH ATHEROSCLEROSIS SEVERITY IN HUMANS AND MOUSE MODELS

Another piece of evidence for a functional role of *ANRIL* in determining CAD risk stems from correlation analysis with disease features in patient cohorts. Aside of the genetic association, *ANRIL* levels were often increased in CAD patients, and not only in atherosclerotic plaque tissue, but also in circulating PBMCs or whole blood. Here, linear *ANRIL* levels were positively correlated with the severity of atherosclerosis (13, 29, 75) whereas *circANRIL* was anticorrelated (36) (**Figure 1B**). Thus, while the genotype of Chr9p21 determines the production of atherogenic (linear) over antiatherogenic *ANRIL* RNA species (circular), CAD and peripheral artery disease-dependent changes may additionally feed into *ANRIL* regulation. For *CDKN2B*, two studies reported a correlation of the expression with atherosclerosis severity (34, 76), where the direction of the correlation (downregulation in plaques) was consistent from what could be expected from the association results. But another study reported increased *p16*^{INK4A} levels to positively correlate with inflammation markers in plaques instead of anticorrelation (25). Together, results from association as well as correlation analyses have established *ANRIL* lncRNA as prime candidate at the Chr9p21 locus.

MOLECULAR FUNCTION OF *ANRIL* AND *CDKN2A/B* IN ATHEROGENESIS

ANRIL belongs to the group of long non-coding RNAs and as such has been suggested to act as a molecular scaffold of chromatin-modifying complexes that control gene expression through modifying histone tails. Specifically*, ANRIL* was found to physically interact with the CBX7 protein inside the PRC1 Polycomb complex, one of the major gene repression complexes in cells (74). Knockdown of members of this Polycomb group complex led to increased expression of the *CDKN2A* and *CDKN2B* tumor suppressors in the Chr9p21 locus. Also, ongoing RNA polymerase II transcription was important for the association of the Polycomb proteins with the locus, indicative of the importance of RNA for recruitment. It was concluded that *ANRIL*'s function may be, at least in part, to repress the *CDKN2A* and *CDKN2B* tumor suppressors. As a consequence increased *ANRIL* levels are thought to promote overproliferation and to be incompatible with senescence onset, a major function of *CDKN2A/B*. As described in chapter 2, other work has shown that recruitment of the Polycomb complexes may account also for how *ANRIL* regulates genes *in trans* on a genome-wide level: Overexpression of linear *ANRIL* isoforms in cultured cells was found to promote pro-atherogenic cell functions, such as proliferation and reduced apoptosis, and to trigger the differential expression of hundreds of genes, in this case without affecting *CDKN2A/B* suppressors. Results from that study therefore questioned whether *ANRIL* regulated these tumor suppressor genes *in cis* at all (36, 77).

How does circular *ANRIL*, whose abundances is reduced in CAD patients, fit into this model? Both in human peripheral blood T-lymphocytes, as well as in PBMCs, whole blood and endatherectomy plaque tissue, *circANRIL* isoforms were found to be downregulated in samples from CAD patients carrying the Chr9p21 risk allele (22, 36). In an initial model, it was suggested that the production of *circANRIL* from central *ANRIL* exons would shorten the linear *ANRIL* lncRNA and, thereby, impaired linear *ANRIL*'s function in epigenetic control of target genes (22). In a second study, a more primary role was found for *circANRIL* that was, furthermore, independent of linear *ANRIL* (36). Here, *circANRIL* was found to be 10-fold more abundant than linear *ANRIL*. Mass-spectrometric analysis of proteins interacting with *circANRIL* showed that it bound to PES1 protein, a member of the evolutionarily conserved PeBoW complex. This complex is essential for proper rRNA-processing, that is the excision of RNA spacer elements from pre-ribosomal rRNA precursors. *CircANRIL* inhibited the activity of the PeBoW complex, as deduced from the accumulation of unsufficiently processed (and non-functional) 26S and 32S pre-rRNA intermediates when *circANRIL* was overexpressed (36). A deficit in rRNA maturation caused nucleolar stress and p53 activation, culminating in inhibition of cell proliferation and in an increase in apoptosis. Notably, the observed functions of *circANRIL* were inverse to that of linear *ANRIL* and, as shown by genomic knockout of linear *ANRIL* exons, independent from the presence of these lncRNA isoforms. Thus, experimental evidence from expression analysis *in vivo* and from genetic experiments both indicated

that *circANRIL* was anti-atherogenic. Together, linear *ANRIL* confers overproliferation, and circular *ANRIL* protects from overproliferation, suggesting that the genotype of Chr9p21 is important to determine the balance of linear and circular *ANRIL* levels in SMCs and macrophages, and that a dominance of linear *ANRIL* in this ratio, even when small, over decades skews for CAD (36) (**Figure 1B**).

Whether suppressing linear *ANRIL* or boosting circularization is sufficient to protect from atherosclerotic cues *in vivo* is matter of ongoing research. The fact that *ANRIL* RNA is not conserved beyond primates complicates the functional *in vivo* analysis of the Chr9p21 locus. So far, insight on how CAD is controlled by Chr9p21 through genetic modeling in mouse mutants is fragmented. The genetic elements of Chr9p21 and their relative positioning are overall syntenically conserved in mouse chromosome 4. So far, only one study has investigated, if deletion of a 70 kb long portion of mouse Chr4 corresponding to the CAD haplotype block in humans had an effect on atherosclerosis *in vivo* (78). This region contains a multi-exon lncRNA, *AK148321*, which is, however, likely not corresponding to human *ANRIL*. Mutant mice (78) developed tumors, reminiscent of tumorigenesis associated with mutation in the Chr9p21 region. But despite some metabolic changes in the mutant mice and enhanced platelet activation, no significant change in atherosclerotic fatty lesion formation was observed (78), putting in question the validity of this mouse model for studying *ANRIL*-driven atherogenesis. On the other hand, the mutants did develop more vascular aneurysms (79), supporting that some aspects of CAD were indeed contained in the noncoding mouse sequence.

Overall, the picture is not yet fully clear. While the genetic data from mice support the importance of individual noncoding genetic elements and of some of the protein-coding tumor suppressors for regulation of atherosclerosis and other CAD entities, whether the lncRNA encoded in the locus regulates CAD mechanistically via epigenetically regulating the neighboring tumor suppressors *in cis* has not been determined. Nevertheless, mouse genetics remains an interesting research avenue to explore some aspects of Chr9p21 biology, at least relating to aneurysm, cancer, and glaucoma formation.

SUMMARY

Starting from a GWAS signal for CAD in a "gene desert" on Chr9p21 in 2007, research in the last decade has firmly established this region as strongest genetic factor of human atherosclerosis and has contributed to a better understanding of the underlying pathophysiology. The picture has emerged that one of the major routes how this locus controls atherosclerosis risk is through regulating the expression of the lncRNA *ANRIL* *in cis*, where the risk allele leads to high levels of linear *ANRIL* but decreases circular *ANRIL* expression. Linear *ANRIL* has been established as molecular scaffold guiding epigenetic protein complexes and promoting pro-atherogentic cells functions. On the contrary, circularization shifts *ANRIL's* function toward controlling ribosomal RNA processing and controlling protein translation thereby promoting athero-protection (**Figure 1B**). The molecular mechanisms of how the ratio of linear and circular *ANRIL* is controlled by the genotype at the locus are currently not resolved and it will be important to determine which gene regulatory elements within the *ANRIL* gene are disturbed by causal CAD risk SNPs. Experimentally exploring details of the molecular effector mechanisms for linear *ANRIL* and for circular *ANRIL* will be paramount, but this task will not be trivial because linear and circular *ANRIL* isoforms always co-exist and in part share the same sequence. Not last, more nuanced relations between Chr9p21 genotype and gene expression output can be expected to be found in the future if, for example, analyses were to take into account cell type-specific and context (stress, inflammation, senescence)-specific effects, aspect that whole tissue expression profiling is currently missing. Additionally, although it is early days, measuring the levels of *circANRIL/linear ANRIL*, might offer a prognostic value and help improve CAD risk stratification or allow to better monitor treatment response or disease recurrence.Yet, since *circANRIL* levels are reduced in plaque tissue, and since *circANRIL* has been found to be anti-atherogenic with or without co-existing linear *ANRIL*, increasing *circANRIL* abundance in patients could also be of therapeutic relevance. Expressing *circANRIL* levels in the cells of the vasculature in CAD disease models might, therefore, be a promising next step to exploit the accumulated knowledge on the Chr9p21 CAD risk locus.

REFERENCES

1. 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

2. 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

3. Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, Mayer B, et al. Genomewide association analysis of coronary artery disease. *N Engl J Med.* (2007) 357:443–53. doi: 10.1056/NEJMoa072366

4. 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/nature 05911

5. Larson MG, Atwood LD, Benjamin EJ, Cupples LA, D'Agostino RBSr, Fox CS, et al. Framingham Heart Study 100K project: genome-wide associations for cardiovascular disease outcomes. *BMC Med Genet.* (2007) 8 Suppl 1:S5. doi: 10.1186/1471-2350-8-S1-S5

6. Matarin M, Brown WM, Singleton A, Hardy JA, Meschia JF, Investigators I. Whole genome analyses suggest ischemic stroke and heart disease share an association with polymorphisms on chromosome 9p21. *Stroke* (2008) 39:1586–9. doi: 10.1161/STROKEAHA.107.502963

7. Wahlstrand B, Orho-Melander M, Delling L, Kjeldsen S, Narkiewicz K, Almgren P, et al. The myocardial infarction associated CDKN2A/CDKN2B locus on chromosome 9p21 is associated with stroke independently of coronary events in patients with hypertension. *J Hypertens.* (2009) 27:769–73. doi: 10.1097/HJH.0b013e328326f7eb

8. Helgadottir A, Thorleifsson G, Magnusson KP, Gretarsdottir S, Steinthorsdottir V, Manolescu A, et al. The same sequence variant on 9p21 associates with myocardial infarction, abdominal aortic aneurysm and intracranial aneurysm. *Nat Genet.* (2008) 40:217–24. doi: 10.1038/ng.72

9. Gschwendtner A, Bevan S, Cole JW, Plourde A, Matarin M, Ross-Adams H, et al. Sequence variants on chromosome 9p21.3 confer risk for atherosclerotic stroke. *Ann Neurol.* (2009) 65:531–9. doi: 10.1002/ana.21590

10. International Stroke Genetics C, Wellcome Trust Case Control C, Bellenguez C, Bevan S, Gschwendtner A, Spencer CC, et al. Genome-wide association study identifies a variant in HDAC9 associated with large vessel ischemic stroke. *Nat Genet.* (2012) 44:328–33. doi: 10.1038/ng.1081

11. Dichgans M, Malik R, Konig IR, Rosand J, Clarke R, Gretarsdottir S, et al. Shared genetic susceptibility to ischemic stroke and coronary artery disease: a genome-wide analysis of common variants. *Stroke* (2014) 45:24–36. doi: 10.1161/STROKEAHA.113.002707

12. Cluett C, McDermott MM, Guralnik J, Ferrucci L, Bandinelli S, Miljkovic I, et al. The 9p21 myocardial infarction risk allele increases risk of peripheral artery disease in older people. *Circ Cardiovasc Genet.* (2009) 2:347–53. doi: 10.1161/CIRCGENETICS.108.825935

13. 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

14. Murabito JM, White CC, Kavousi M, Sun YV, Feitosa MF, Nambi V, et al. Association between chromosome 9p21 variants and the ankle-brachial index identified by a meta-analysis of 21 genome-wide association studies. *Circ Cardiovasc Genet.* (2012) 5:100–12. doi: 10.1161/CIRCGENETICS.111.961292

15. Bilguvar K, Yasuno K, Niemela M, Ruigrok YM, von Und Zu Fraunberg M, van Duijn CM, et al. Susceptibility loci for intracranial aneurysm in European and Japanese populations. *Nat Genet.* (2008) 40:1472–7. doi: 10.1038/ng.240

16. Yasuno K, Bilguvar K, Bijlenga P, Low SK, Krischek B, Auburger G, et al. Genome-wide association study of intracranial aneurysm identifies three new risk loci. *Nat Genet.* (2010) 42:420–5. doi: 10.1038/ng.563

17. Nelson CP, Goel A, Butterworth AS, Kanoni S, Webb TR, Marouli E, et al. Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nat Genet.* (2017) 49:1385–91. doi: 10.1038/ng.3913

18. 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

19. 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

20. 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

21. 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

22. Burd CE, Jeck WR, Liu Y, Sanoff HK, Wang Z, Sharpless NE. Expression of linear and novel circular forms of an INK4/ARF-associated non-coding RNA correlates with atherosclerosis risk. *PLoS Genet.* (2010) 6:e1001233. doi: 10.1371/journal.pgen.1001233

23. Zeller T, Wild P, Szymczak S, Rotival M, Schillert A, Castagne R, et al. Genetics and beyond–the transcriptome of human monocytes and disease susceptibility. *PLoS ONE* (2010) 5:e10693. doi: 10.1371/journal.pone.0010693

24. Schunkert H, Konig IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet.* (2011) 43:333–8. doi: 10.1038/ng.784

25. Holdt LM, Sass K, Gabel G, Bergert H, Thiery J, Teupser D. Expression of Chr9p21 genes CDKN2B (p15(INK4b)), CDKN2A (p16(INK4a), p14(ARF)) and MTAP in human atherosclerotic plaque. *Atherosclerosis* (2011) 214:264– 70. doi: 10.1016/j.atherosclerosis.2010.06.029

26. Kim JB, Deluna A, Mungrue IN, Vu C, Pouldar D, Civelek M, et al. Effect of 9p21.3 coronary artery disease locus neighboring genes on atherosclerosis in mice. *Circulation* (2012) 126:1896–906. doi: 10.1161/CIRCULATIONAHA.111.064881

27. Pilbrow AP, Folkersen L, Pearson JF, Brown CM, McNoe L, Wang NM, et al. The chromosome 9p21.3 coronary heart disease risk allele is associated with altered gene expression in normal heart and vascular tissues. *PLoS ONE* (2012) 7:e39574. doi: 10.1371/journal.pone.0039574

28. Zhuang J, Peng W, Li H, Wang W, Wei Y, Li W, et al. Methylation of p15INK4b and expression of ANRIL on chromosome 9p21 are associated with coronary artery disease. *PLoS ONE* (2012) 7:e47193. doi: 10.1371/journal.pone.0047193

29. Congrains A, Kamide K, Oguro R, Yasuda O, Miyata K, Yamamoto E, et al. Genetic variants at the 9p21 locus contribute to atherosclerosis through modulation of ANRIL and CDKN2A/B. *Atherosclerosis* (2012) 220:449–55. doi: 10.1016/j.atherosclerosis.2011.11.017

30. Zollbrecht C, Grassl M, Fenk S, Hocherl R, Hubauer U, Reinhard W, et al. Expression pattern in human macrophages dependent on 9p21.3 coronary artery disease risk locus. *Atherosclerosis* (2013) 227:244–9. doi: 10.1016/j.atherosclerosis.2012.12.030

31. 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

32. Johnson AD, Hwang SJ, Voorman A, Morrison A, Peloso GM, Hsu YH, et al. Resequencing and clinical associations of the 9p21.3 region: a comprehensive investigation in the Framingham heart study. *Circulation* (2013) 127:799–810. doi: 10.1161/CIRCULATIONAHA.112.111559

33. 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

34. Shanker J, Arvind P, Jambunathan S, Nair J, Kakkar V. Genetic analysis of the 9p21.3 CAD risk locus in Asian Indians. *Thromb Haemost.* (2014) 111:960–9. doi: 10.1160/TH13-08-0706

35. Miller CL, Pjanic M, Wang T, Nguyen T, Cohain A, Lee JD, et al. Integrative functional genomics identifies regulatory mechanisms at coronary artery disease loci. *Nat Commun.* (2016) 7:12092. doi: 10.1038/ncomms 12092

36. 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

37. Franzen O, Ermel R, Cohain A, Akers NK, Di Narzo A, Talukdar HA, et al. Cardiometabolic risk loci share downstream cis- and trans-gene regulation across tissues and diseases. *Science* (2016) 353:827–30. doi: 10.1126/science.aad6970

38. Liu Y, Sanoff HK, Cho H, Burd CE, Torrice C, Mohlke KL, et al. INK4/ARF transcript expression is associated with chromosome 9p21 variants linked to atherosclerosis. *PLoS ONE* (2009) 4:e5027. doi: 10.1371/journal.pone.0005027

39. Sarkar D, Oghabian A, Bodiyabadu PK, Joseph WR, Leung EY, Finlay GJ, et al. Multiple isoforms of ANRIL in melanoma cells: structural complexity suggests variations in processing. *Int J Mol Sci.* 18:1378. doi: 10.3390/ijms18071378

40. Jeck WR, Siebold AP, Sharpless NE. Review: a meta-analysis of GWAS and age-associated diseases. *Aging Cell* (2012) 11:727–31. doi: 10.1111/j.1474-9726.2012.00871.x

41. Nakaoka H, Gurumurthy A, Hayano T, Ahmadloo S, Omer WH, Yoshihara K, et al. Allelic imbalance in regulation of ANRIL through chromatin interaction at 9p21 endometriosis risk locus. *PLoS Genet.* (2016) 12:e1005893. doi: 10.1371/journal.pgen.1005893

42. Schaefer AS, Richter GM, Groessner-Schreiber B, Noack B, Nothnagel M, El Mokhtari NE, et al. Identification of a shared genetic susceptibility locus for coronary heart disease and periodontitis. *PLoS Genet.* (2009) 5:e1000378. doi: 10.1371/journal.pgen.1000378

43. Musunuru K, Post WS, Herzog W, Shen H, O'Connell JR, McArdle PF, et al. Association of single nucleotide polymorphisms on chromosome 9p21.3 with platelet reactivity: a potential mechanism for increased vascular disease. *Circ Cardiovasc Genet.* (2010) 3:445–53. doi: 10.1161/CIRCGENETICS.109.923508

44. 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

45. Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Mit LU, Novartis Institutes of BioMedical R, Saxena R, Voight BF, Lyssenko V, 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

46. Erdmann J, Grosshennig A, Braund PS, Konig IR, Hengstenberg C, Hall AS, et al. New susceptibility locus for coronary artery disease on chromosome 3q22.3. *Nat Genet.* (2009) 41:280–2. doi: 10.1038/ng.307

47. Myocardial Infarction Genetics C, Kathiresan S, Voight BF, Purcell S, Musunuru K, Ardissino D, et al. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat Genet.* (2009) 41:334–41. doi: 10.1038/ng.327

48. Tregouet DA, Konig IR, Erdmann J, Munteanu A, Braund PS, Hall AS, et al. Genome-wide haplotype association study identifies the SLC22A3-LPAL2-LPA gene cluster as a risk locus for coronary artery disease. *Nat Genet.* (2009) 41:283–5. doi: 10.1038/ng.314

49. Consortium IKC. Large-scale gene-centric analysis identifies novel variants for coronary artery disease. *PLoS Genet.* (2011) 7:e1002260. doi: 10.1371/journal.pgen.1002260

50. Wang Z, Jacobs KB, Yeager M, Hutchinson A, Sampson J, Chatterjee N, et al. Improved imputation of common and uncommon SNPs with a new reference set. *Nat Genet.* (2011) 44:6–7. doi: 10.1038/ng.1044

51. 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

52. 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

53. Myocardial Infarction G, Investigators CAEC, Stitziel NO, Stirrups KE, Masca NG, Erdmann J, et al. Coding variation in ANGPTL4, LPL, and SVEP1 and the risk of coronary artery disease. *N Engl J Med.* (2016) 374:1134–44. doi: 10.1056/NEJMoa1507652

54. Nioi P, Sigurdsson A, Thorleifsson G, Helgason H, Agustsdottir AB, Norddahl GL, et al. Variant ASGR1 associated with a reduced risk of coronary artery disease. *N Engl J Med.* (2016) 374:2131–41. doi: 10.1056/NEJMoa1508419

55. Webb TR, Erdmann J, Stirrups KE, Stitziel NO, Masca NG, Jansen H, et al. Systematic evaluation of pleiotropy identifies 6 further loci associated with coronary artery disease. *J Am Coll Cardiol.* (2017) 69:823–36. doi: 10.1016/j.jacc.2016.11.056

56. Howson JMM, Zhao W, Barnes DR, Ho WK, Young R, Paul DS, et al. Fifteen new risk loci for coronary artery disease highlight arterial-wall-specific mechanisms. *Nat Genet.* (2017) 49:1113–9. doi: 10.1038/ng.3874

57. Klarin D, Zhu QM, Emdin CA, Chaffin M, Horner S, McMillan BJ, et al. Genetic analysis in UK Biobank links insulin resistance and transendothelial migration pathways to coronary artery disease. *Nat Genet.* (2017) 49:1392–7. doi: 10.1038/ng.3914

58. 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

59. Salzman J, Gawad C, Wang PL, Lacayo N, Brown PO. Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS ONE* (2012) 7:e30733. doi: 10.1371/journal.pone.0030733

60. Barrett SP, Salzman J. Circular RNAs: analysis, expression and potential functions. *Development* (2016) 143:1838–47. doi: 10.1242/dev.128074

61. Holdt LM, Kohlmaier A, Teupser D. Molecular roles and function of circular RNAs in eukaryotic cells. *Cell Mol Life Sci.* (2018) 75:1071–98. doi: 10.1007/s00018-017-2688-5

62. 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

63. Mehta NN. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Circ Cardiovasc Genet.* (2011) 4:327–9. doi: 10.1161/CIRCGENETICS.111.960443

64. Mumbach MR, Satpathy AT, Boyle EA, Dai C, Gowen BG, Cho SW, et al. Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nat Genet.* (2017) 49:1602–12. doi: 10.1038/ng.3963

65. Svensson PA, Wahlstrand B, Olsson M, Froguel P, Falchi M, Bergman RN, et al. CDKN2B expression and subcutaneous adipose tissue expandability: possible influence of the 9p21 atherosclerosis locus. *Biochem Biophys Res Commun.* (2014) 446:1126–31. doi: 10.1016/j.bbrc.2014.03.075

66. Li WQ, Pfeiffer RM, Hyland PL, Shi J, Gu F, Wang Z, et al. Genetic polymorphisms in the 9p21 region associated with risk of multiple cancers. *Carcinogenesis* (2014) 35:2698–705. doi: 10.1093/carcin/bgu203

67. Kong Y, Sharma RB, Ly S, Stamateris RE, Jesdale WM, Alonso LC. CDKN2A/B T2D genome-wide association study risk SNPs impact locus gene expression and proliferation in human islets. *Diabetes* (2018) 67:872–84. doi: 10.2337/db17-1055

68. Almontashiri NA, Antoine D, Zhou X, Vilmundarson RO, Zhang SX, Hao KN, et al. 9p21.3 coronary artery disease risk variants disrupt TEAD transcription factor-dependent transforming growth factor beta regulation of p16 expression in human aortic smooth muscle cells. *Circulation* (2015) 132:1969–78. doi: 10.1161/CIRCULATIONAHA.114.015023

69. Congrains A, Kamide K, Katsuya T, Yasuda O, Oguro R, Yamamoto K, et al. CVD-associated non-coding RNA, ANRIL, modulates expression of atherogenic pathways in VSMC. *Biochem Biophys Res Commun.* (2012) 419:612–6. doi: 10.1016/j.bbrc.2012.02.050

70. Zhou X, Han X, Wittfeldt A, Sun J, Liu C, Wang X, et al. Long non-coding RNA ANRIL regulates inflammatory responses as a novel component of NF-kappaB pathway. *RNA Biol.* (2016) 13:98–108. doi: 10.1080/15476286.2015.1122164

71. Bochenek G, Hasler R, El Mokhtari NE, Konig 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

72. Bai Y, Nie S, Jiang G, Zhou Y, Zhou M, Zhao Y, et al. Regulation of CARD8 expression by ANRIL and association of CARD8 single nucleotide polymorphism rs2043211 (p.C10X) with ischemic stroke. *Stroke* (2014) 45:383–8. doi: 10.1161/STROKEAHA.113.003393

73. Lai F, Orom UA, Cesaroni M, Beringer M, Taatjes DJ, Blobel GA, et al. Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* (2013) 494:497–501. doi: 10.1038/nature11884

74. 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

75. 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

76. Kojima Y, Downing K, Kundu R, Miller C, Dewey F, Lancero H, et al. Cyclin-dependent kinase inhibitor 2B regulates efferocytosis and atherosclerosis. *J Clin Invest.* (2014) 124:1083–97. doi: 10.1172/JCI 70391

77. Sato K, Nakagawa H, Tajima A, Yoshida K, Inoue I. ANRIL is implicated in the regulation of nucleus and potential transcriptional target of E2F1. *Oncol Rep.* (2010) 24:701–7. doi: 10.3892/or_00000910

78. Visel A, Zhu Y, May D, Afzal V, Gong E, Attanasio C, et al. Targeted deletion of the 9p21 non-coding coronary artery disease risk interval in mice. *Nature* (2010) 464:409–12. doi: 10.1038/nature 08801

79. Loinard C, Basatemur G, Masters L, Baker L, Harrison J, Figg N, et al. Deletion of chromosome 9p21 noncoding cardiovascular risk interval in mice alters Smad2 signaling and promotes vascular aneurysm. *Circ Cardiovasc Genet.* (2014) 7:799–805. doi: 10.1161/CIRCGENETICS.114.000696

Check for updates

Into the Wild: GWAS Exploration of Non-coding RNAs

Hector Giral [1,2], Ulf Landmesser [1,2,3] and Adelheid Kratzer [1,2]*

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

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 gencode ID (Aliases)[a]	Cytogenetic region	Genes in locus	SNP variant[b]	Disease or phenotype traits	NHGRI Catalog[c]	Other sources[d]	GTEx eQTLs[e]
CDKN2B-AS1 ENSG00000240498 (ANRIL)	9p21.3 3.4kb 3' CDKN2B	CDKN2A, CDKN2B, DMRTA1	rs2891168 (intronic)	myocardial infarction CAD	GCST003117 (13) 5×10^{-75} GCST005194 (12) 5×10^{-204}		none
	9p21.3p 13kb 3' CDKN2B	CDKN2A, CDKN2B, DMRTA1	rs4977574 (intronic) rs10811661 (intergenic)	CAD Type 2 Diabetes	GCST005195 (12) 9×10^{-223} GCST004894 (22) 2×10^{-68}		none none
MIAT ENSG00000225783 (lnc-CRYBA4)	22q12.1		rs2331291 (intronic)	myocardial infarction		GWASdb v2 (23) 5×10^{-7}	MIAT (10hits)
H19 ENSG00000130600 (LINC00008)	11p15.5	HOTS, MRPL23	rs217727 (exonic)	Systolic blood pressure CAD	GCST006188 (24) 7×10^{-15}	Meta-analysis (25) 1×10^{-7} Case-control genotyping (26) 1×10^{-3}	none
LOC157273 ENSG00000248538 (AC022784.1; RP11-10A14.4)	8p23.1	PPP1RB	rs9987289 (intronic) rs4841132 (exonic)	HDL cholesterol Coronary artery calcification Fasting blood insulin	GCST002223 (27) 2×10^{-44} GCST001639 (28) 2×10^{-9} GCST005185 (29) 6×10^{-15}		none
KCNQ1OT1 ENSG00000269821 (KCNQ1-AS2)	11p15.5	KCNQ1	rs231362 (exonic)	Type 2 Diabetes	GCST000712 (30) 3×10^{-13}		none
LINC00243 ENST00000419357 (lnc-IER3)	6p21.33	ZNF507	rs886424 (exonic)	Type 1 Diabetes Total cholesterol Total TG	GCST002876 (31) 3×10^{-14}	Meta-analysis (32) 2×10^{-9} 8×10^{-11}	(IER3, HLAs…) 153 hits
LINC02137 ENSG00000260186 (RP11-481J2.2)	16q21		rs4784934 (intronic)	QT Interval	GCST002500 (33) 6×10^{-9}		LINC02137 (4 hits) NDRG4 (1 hit)
LOC400684 ENSG00000267213	19q13.11	ZNF507	rs12976411 (intronic)	CAD	GCST003116 (13) 1×10^{-14}		ZNF507 (1hit)
LINC00310 ENSG00000227456	21q22.11	MRPS6, SLC5A3	rs28451064 (intronic)	CAD myocardial infarction	GCST005195 (12) 1×10^{-33} GCST003117 (13) 6×10^{-12}		none

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.

[a]Official name; Gencode ID (GrCh38.p12); aliases

[b]Official variant ID; variant location in the locus

[c]NHGRI GWAS Catalog study ID described; reference of the study; p-value

[d]alternative sources such as GWASdb catalog or meta-analysis studies; reference of the study; p-value

[e]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-α21 (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).
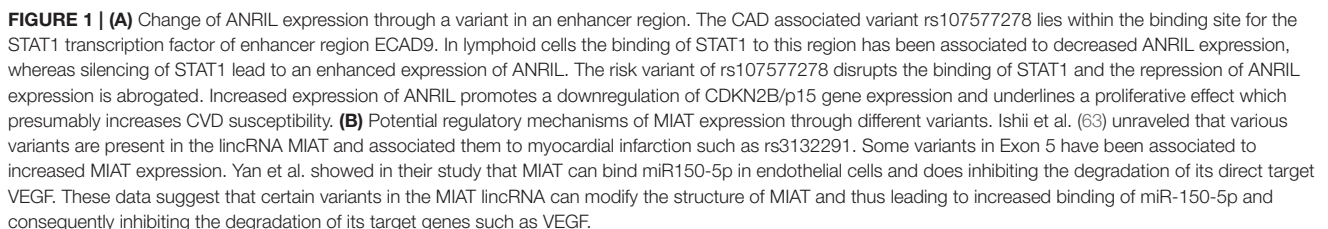
Only five of the CAD candidate variants are located in exons of CDKN2B-AS1 but none of them are located in conserved elements, questioning the likeliness to affect functional domains (59). However, numerous splice isoforms have been identified for CDKN2B-AS1 (14 isoforms, Genbank; 21 isoforms, GENCODE) 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 rs107577278 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 rs107577278 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 lncRNA 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.

AUTHOR CONTRIBUTIONS

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

ACKNOWLEDGMENTS

REFERENCES

1. 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
2. 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
3. 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
4. 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
5. 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
6. 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
7. 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
8. 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
9. 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
10. Erdmann J, Grosshennig A, Braund PS, Konig 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
11. 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
12. 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
13. 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
14. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. *Nature* (2009) 461:747–53. doi: 10.1038/nature08494
15. Chasman DI. Peering into the future of CAD genomics. *Circ Res.* (2018) 122:391–3. doi: 10.1161/CIRCRESAHA.117.312502
16. 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

17. 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
18. 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
19. 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.bbadis.2014.03.011
20. 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
21. 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
22. 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
23. 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
24. 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
25. 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
26. 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
27. 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
28. Inouye M, Ripatti S, Kettunen J, Lyytikainen 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
29. 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
30. 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, Yaghootkar 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, Burtt 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, Konig 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 Metab.* (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, Liau 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, Mazor 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, Bilenky M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. *Nature* (2015) 518:317–30. doi: 10.1038/nature14248

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership