



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

OPEN ACCESS

Edited by:

Tanja Zeller,
Universität Hamburg, Germany

Reviewed by:

Hauke Busch,
Universität zu Lübeck, Germany
Ville-Petteri Mäkinen,
South Australian Health and Medical
Research Institute, Australia
Saikat Banerjee,
Max-Planck-Institut für
Biophysikalische Chemie, Germany

*Correspondence:

Matthias Heinig
matthias.heinig@
helmholtz-muenchen.de

Specialty section:

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

Received: 26 March 2018

Accepted: 15 May 2018

Published: 05 June 2018

Citation:

Heinig M (2018) Using Gene
Expression to Annotate
Cardiovascular GWAS Loci.
Front. Cardiovasc. Med. 5:59.
doi: 10.3389/fcvm.2018.00059

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 visceral 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 of 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).

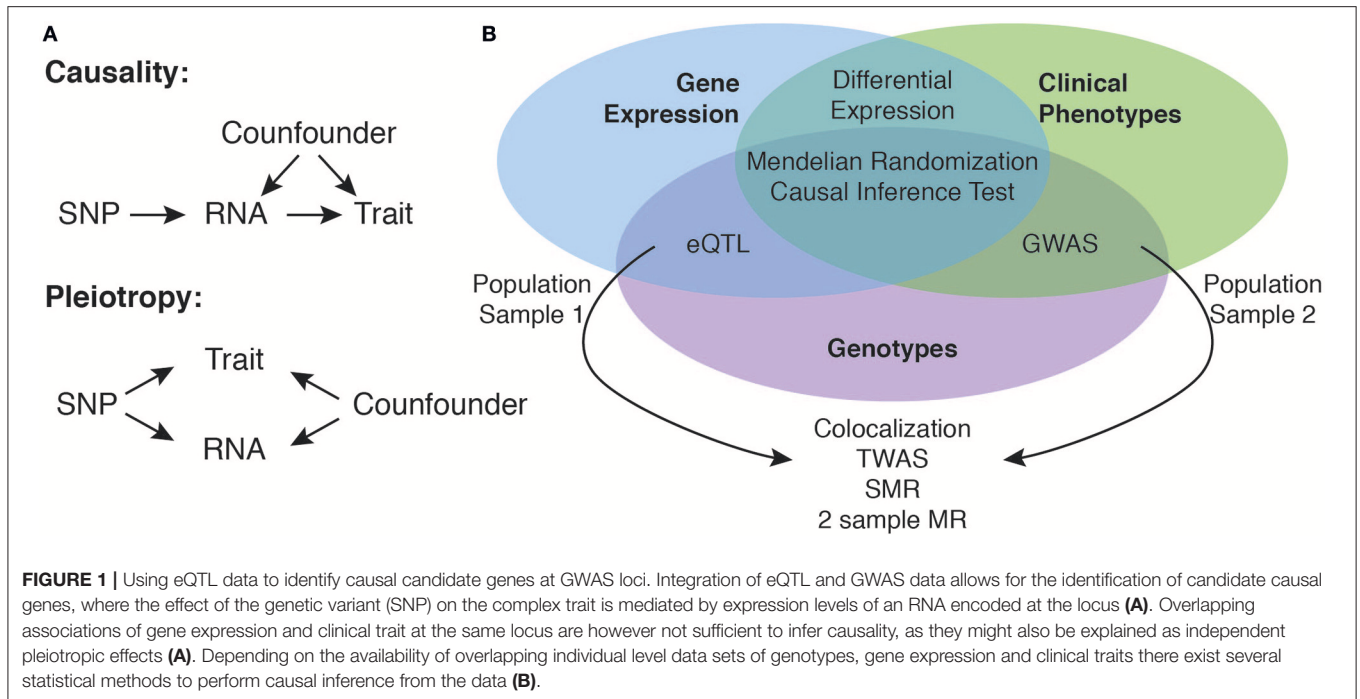
An important limitation of the overlap-based strategy is that it cannot be used to establish causality. Strictly speaking the experimental design does only allow inferring causality in a statistical sense. In genetic associations the direction of causality is always fixed (Figure 1A). To establish a 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



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.

REFERENCES

- 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/gkx1133
- Burdett T, Hall PN, Hastings E, Hindorf 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
- 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
- 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
- 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
- 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
- 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
- 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
- ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* (2012) **489**:57–74. doi: 10.1038/nature11247
- 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
- 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
- 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
- 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
- Civelek M, and Lusis AJ. Systems genetics approaches to understand complex traits. *Nat Rev Genet.* (2014) **15**:34–48. doi: 10.1038/nrg3575
- 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
- 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
- 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
- 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
- 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
- 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
- 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
- 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
- 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
- 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/nature12531
- 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
- 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
- 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
- Suhre K, Shin SY, Petersen AK, Mohny RP, Meredith D, Wägele B, et al. Human metabolic individuality in biomedical and pharmaceutical research. *Nature* (2011) **477**:54–60. doi: 10.1038/nature10354
- 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
- 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
- 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

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

FUNDING

This work was supported by funding to MH by the Federal Ministry of Education and Research (BMBF, Germany) in the projects eMed:symAtrial (01ZX1408D) and eMed:confirm (01ZX1708G).

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-Marin, 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 disease novelty 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, Séguérel 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 inflammation 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 N, 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

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.

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