frontiers

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

*** Correspondence:**
Matthias Heinig

matthias.heinig@helmholtz-muenchen.de

Keywords: eQTL, expression quantitative trait loci, genome wide association study, GWAS, cardiovascular disease. (Min.5-Max. 8)

Abstract

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.

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 (MacArthur et al. 2017). This progress has also highlighted hundreds of loci associated with cardiovascular phenotypes: the current GWAS catalog (Burdett et al. 2018) lists 249 distinct chromosomal regions associated with coronary artery disease with candidate genes and pathways at many loci summarized in (Klarin et al. 2017), 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 (Timpson et al. 2018), for instance truncating mutations in the sarcomeric protein TTN cause dilated cardiomyopathy (Siu et al. 1999; Gerull et al. 2002; Herman et al. 2012; Roberts et al. 2015). 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 (Burdett et al. 2018). 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 (ENCODE Project Consortium 2012; Roadmap Epigenomics Consortium et al. 2015). 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 (Maurano et al. 2012; Farh et al. 2015). Since regulatory elements are highly tissue specific, this information can even be used to identify the disease-relevant tissues (Maurano et al. 2012; Farh et al. 2015). 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 (Jansen and Nap 2001), 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 (Civelek and Lusis 2014; Albert and Kruglyak 2015). Biological information processing and regulation is not limited to transcription, so this approach has also been generalized towards other intermediate molecular traits such as DNA methylation (Banovich et al. 2014; Lemire et al. 2015), open chromatin (Degner et al. 2012), histone modifications (Waszak et al. 2015; Grubert et al. 2015) (Del Rosario et al. 2015), gene, exon and transcript expression levels (Montgomery et al. 2010; Pickrell et al. 2010; Lappalainen et al. 2013; GTEx Consortium et al. 2015; Battle et al. 2017) translation and protein levels (Li et al. 2016) as well as metabolites (Suhre et al. 2011; Shin et al. 2014). In particular the information from the epigenome can be used to identify regulatory variants, and to characterize their role in disease (Maurano et al. 2012; Del Rosario et al. 2015; Degner et al. 2012; Li et al. 2016).

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 (Maurano et al. 2012; Grundberg et al. 2012; Farh et al. 2015; GTEx Consortium et al. 2015). 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 (Monti et al. 2008). 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 (Heinig et al. 2017) or ischemic time of post-mortem samples (GTEx Consortium et al. 2015). Therefore additional care has to be taken in data analysis accounting for observed and hidden confounders (Stegle et al. 2010). Current reviews provide an overview of recent human eQTL studies (Albert and Kruglyak 2015; Vandiedonck 2018).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 (Battle et al. 2017). 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 resource (Franzén et al. 2016). 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 (Heinig et al. 2017). This observation is also supported by the finding that more eQTLs associated with disease SNP can be found in diseased populations (Franzén et al. 2016). 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 (Zeller et al. 2010). 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 (Kilpinen et al. 2017). 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 (Roadmap Epigenomics Consortium et al. 2015) by annotation with epigenetic markers of regulatory elements and recent developments of sequencing based methods (e.g. Hi-C) to study chromosomal architecture(Davies et al. 2017) in a wide variety of human tissues (Schmitt et al. 2016) 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 (Smemo et al. 2016).

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 500kb – 1Mb, 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 (Musunuru et al. 2010) and LIPA (Wild et al. 2011) 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 (Heinig et al. 2017) and also more generally for any GWAS trait (GTEx Consortium et al. 2015; Battle et al. 2017). 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 (Watanabe et al. 2017). For instance a recent GWAS on CAD (van der Harst and Verweij 2018) 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 (Battle et al. 2017). 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 (Heinig et al. 2017). 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 (Westra et al. 2013).

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 1 (A)). 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 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 (Davey Smith and Hemani 2014). 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 1 (A)) changes in phenotypic outcome can be attributed to the changes in gene expression.

Classically, MR and similar approaches to statistically establish causality (Schadt et al. 2005) (Millstein et al. 2009) require to measure all variables in the same population Figure 1 (B). 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 (Pickrell et al. 2016; Hormozdiari et al. 2016; Gusev et al. 2016; Zhu et al. 2016). A Bayesian co-localization approach based on summary statistics (Giambartolomei et al. 2014) 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 (Manor and Segal 2013) into a GWAS population (Gamazon et al. 2015; Gusev et al. 2016) 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 (Gamazon et al. 2015; Gusev et al. 2016). Alternatively the transcriptome wide association study (TWAS) method (Gusev et al. 2016) and other methods (Barbeira et al. 2017) 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 (Gusev et al. 2016). 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 (Pierce and Burgess 2013) 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 (Zhu et al. 2016). 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 (Zhong, Beaulaurier, et al. 2010; Zhong, Yang, et al. 2010) 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 (Mäkinen et al. 2014).

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 (Mäkinen et al. 2014). 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 (Wang et al. 2012; Shu et al. 2017; Mäkinen et al. 2014; Talukdar et al. 2016; Franzén et al. 2016). A study investigating shared molecular networks and their drivers between cardiovascular diseases and type 2 Diabetes applied this strategy (Shu et al. 2017). 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 (Talukdar et al. 2016) and in vitro studies (Talukdar et al. 2016; Mäkinen et al. 2014).

Conclusions

eQTL data provides first leads towards 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(Kilpinen et al. 2017) as they can potentially be differentiated into any cell type or developmental stage (Nguyen et al. 2018) 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 (Kang et al. 2018) 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 towards causation. Clearly this is the most difficult task and requires on top of rigorous statistical approaches such as MR also experimental validation.

1 Conflict of Interest

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.

2 Author Contributions

MH wrote the manuscript.

3 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).

4 References

Albert, Frank W, and Leonid Kruglyak. 2015. "The Role of Regulatory Variation in Complex Traits and Disease.." *Nature Publishing Group* 16 (4): 197–212. doi:10.1038/nrg3891.

Banovich, Nicholas E, Xun Lan, Graham McVicker, Bryce van de Geijn, Jacob F Degner, John D Blischak, Julien Roux, Jonathan K Pritchard, and Yoav Gilad. 2014. "Methylation QTLs Are Associated with Coordinated Changes in Transcription Factor Binding, Histone Modifications, and Gene Expression Levels." Edited by Timothy E Reddy. *PLoS Genetics* 10 (9): e1004663–12. doi:10.1371/journal.pgen.1004663.

Barbeira, A, S P Dickinson, J M Torres, ES Torstenson bioRxiv, 2017. 2017. "Integrating Tissue Specific Mechanisms Into GWAS Summary Results." *Biorxiv.org* , October. doi:10.1101/045260.

Battle, Alexis, Christopher D Brown, Barbara E Engelhardt, and Stephen B Montgomery. 2017. "Genetic Effects on Gene Expression Across Human Tissues." *Nature* 550 (7675). Nature Publishing Group: 204–13. doi:10.1038/nature24277.

Burdett, Tony, P N Hall, Emma Hastings, Lucia A Hindorff, Heather Junkins, Alan Klemm, Jacqueline MacArthur, et al. 2018. "The NHGRI-EBI Catalog of Published Genome-Wide Association Studies." February 12. www.ebi.ac.uk/gwas.

Civelek, Mete, and Aldons J Lusis. 2014. "Systems Genetics Approaches to Understand Complex Traits.." *Nature Reviews Genetics* 15 (1). Nature Research: 34–48. doi:10.1038/nrg3575.

Davey Smith, George, and Gibran Hemani. 2014. "Mendelian Randomization: Genetic Anchors for Causal Inference in Epidemiological Studies.." *Human Molecular Genetics* 23 (R1): R89–R98. doi:10.1093/hmg/ddu328.

Davies, James O J, A Marieke Oudelaar, Douglas R Higgs, and Jim R Hughes. 2017. "How Best to Identify Chromosomal Interactions: a Comparison of Approaches." *Nature Methods*, January. Nature Publishing Group, 1–10. doi:10.1038/nmeth.4146.

Degner, Jacob F, Athma A Pai, Roger Pique-Regi, Jean-Baptiste Veyrieras, Daniel J Gaffney, Joseph K Pickrell, Sherryl De Leon, et al. 2012. "DNase I Sensitivity QTLs Are a Major Determinant of Human Expression Variation.." *Nature* 482 (7385): 390–94. doi:10.1038/nature10808.

Del Rosario, Ricardo Cruz-Herrera, Jeremie Poschmann, Sigrid Laure Rouam, Eileen Png, Chiea

Chuen Khor, Martin Lloyd Hibberd, and Shyam Prabhakar. 2015. “Sensitive Detection of Chromatin-Altering Polymorphisms Reveals Autoimmune Disease Mechanisms..” *Nature Methods* 12 (5): 458–64. doi:10.1038/nmeth.3326.

ENCODE Project Consortium. 2012. “An Integrated Encyclopedia of DNA Elements in the Human Genome..” *Nature* 489 (7414). Nature Publishing Group: 57–74. doi:10.1038/nature11247.

Farh, Kyle Kai-How, Alexander Marson, Jiang Zhu, Markus Kleinewietfeld, William J Housley, Samantha Beik, Noam Shoresh, et al. 2015. “Genetic and Epigenetic Fine Mapping of Causal Autoimmune Disease Variants..” *Nature* 518 (7539): 337–43. doi:10.1038/nature13835.

Franzén, Oscar, Raili Ermel, Ariella Cohain, Nicholas K Akers, Antonio Di Narzo, Husain A Talukdar, Hassan Foroughi Asl, et al. 2016. “Cardiometabolic Risk Loci Share Downstream Cis- and Trans-Gene Regulation Across Tissues and Diseases..” *Science* 353 (6301). American Association for the Advancement of Science: 827–30. doi:10.1126/science.aad6970.

Gamazon, Eric R, Heather E Wheeler, Kaanan P Shah, Sahar V Mozaffari, Keston Aquino-Michaels, Robert J Carroll, Anne E Eyler, et al. 2015. “A Gene-Based Association Method for Mapping Traits Using Reference Transcriptome Data..” *Nature Genetics* 47 (9): 1091–98. doi:10.1038/ng.3367.

Gerull, Brenda, Frenneaux, Michael, Michael Gramlich, John Atherton, Mark McNabb, Karoly Trombitás, Sabine Sasse-Klaassen, et al. 2002. “Mutations of TTN, Encoding the Giant Muscle Filament Titin, Cause Familial Dilated Cardiomyopathy..” *Nature Genetics* 30 (2): 201–4. doi:10.1038/ng815.

Giambartolomei, Claudia, Damjan Vukcevic, Eric E Schadt, Lude Franke, Aroon D Hingorani, Chris Wallace, and Vincent Plagnol. 2014. “Bayesian Test for Colocalisation Between Pairs of Genetic Association Studies Using Summary Statistics..” Edited by Scott M Williams. *PLoS Genetics* 10 (5). Public Library of Science: e1004383. doi:10.1371/journal.pgen.1004383.

Grubert, Fabian, Judith B Zaugg, Maya Kasowski, Oana Ursu, Damek V Spacek, Alicia R Martin, Peyton Greenside, et al. 2015. “Genetic Control of Chromatin States in Humans Involves Local and Distal Chromosomal Interactions..” *Cell* 162 (5): 1051–65. doi:10.1016/j.cell.2015.07.048.

Grundberg, Elin, Kerrin S Small, Åsa K Hedman, Alexandra C Nica, Alfonso Buil, Sarah Keildson, Jordana T Bell, et al. 2012. “Mapping Cis- and Trans-Regulatory Effects Across Multiple Tissues in Twins..” *Nature Genetics* 44 (10): 1084–89. doi:10.1038/ng.2394.

GTEx Consortium, K G Ardlie, F A Wright, and E T Dermitzakis. 2015. “Human Genomics. the Genotype-Tissue Expression (GTEx) Pilot Analysis: Multitissue Gene Regulation in Humans..” *Science* 348 (6235): 648–60. doi:10.1126/science.1262110.

Gusev, Alexander, Arthur Ko, Huwenbo Shi, Gaurav Bhatia, Wonil Chung, Brenda W J H Penninx, Rick Jansen, et al. 2016. “Integrative Approaches for Large-Scale Transcriptome-Wide Association Studies.” *Nature Genetics* 48 (3). Nature Publishing Group: 245–52. doi:10.1038/ng.3506.

Heinig, Matthias, Michiel E Adriaens, Sebastian Schafer, Hanneke W M van Deutekom, Elisabeth M Lodder, James S Ware, Valentin Schneider, et al. 2017. “Natural Genetic Variation of the Cardiac Transcriptome in Non-Diseased Donors and Patients with Dilated Cardiomyopathy..” *Genome Biology* 18 (1). BioMed Central: 170. doi:10.1186/s13059-017-1286-z.

Herman, Daniel S, Lien Lam, Matthew R G Taylor, Libin Wang, Polakit Teekakirikul, Danos Christodoulou, Lauren Conner, et al. 2012. “Truncations of Titin Causing Dilated Cardiomyopathy..” *New England Journal of Medicine* 366 (7): 619–28. doi:10.1056/NEJMoa1110186.

Hormozdiari, Farhad, Martijn van de Bunt, Ayellet V Segrè, Xiao Li, Jong Wha J Joo, Michael Bilow, Jae Hoon Sul, Sriram Sankararaman, Bogdan Pasaniuc, and Eleazar Eskin. 2016. “Colocalization of GWAS and eQTL Signals Detects Target Genes.” *The American Journal of Human Genetics* 99 (6). American Society of Human Genetics: 1245–60.

doi:10.1016/j.ajhg.2016.10.003.
Jansen, R C, and J P Nap. 2001. “Genetical Genomics: the Added Value From Segregation..” *Trends Genet*.
Kang, Hyun Min, Meena Subramaniam, Sasha Targ, Michelle Nguyen, Lenka Maliskova, Elizabeth McCarthy, Eunice Wan, et al. 2018. “Multiplexed Droplet Single-Cell RNA-Sequencing Using Natural Genetic Variation..” *Nature Biotechnology* 36 (1). Nature Publishing Group: 89–94. doi:10.1038/nbt.4042.
Kilpinen, Helena, Angela Goncalves, Andreas Leha, Vackar Afzal, Kaur Alasoo, Sofie Ashford, Sendu Bala, et al. 2017. “Common Genetic Variation Drives Molecular Heterogeneity in Human iPSCs..” *Nature* 546 (7658). Nature Research: 370–75. doi:10.1038/nature22403.
Klarin, Derek, Qiuyu Martin Zhu, Connor A Emdin, Mark Chaffin, Steven Horner, Brian J McMillan, Alison Leed, et al. 2017. “Genetic Analysis in UK Biobank Links Insulin Resistance and Transendothelial Migration Pathways to Coronary Artery Disease.” *Nature Genetics* 49 (9). Nature Publishing Group: 1392–97. doi:10.1038/ng.3914.
Lappalainen, Tuuli, Michael Sammeth, Marc R Friedländer, Peter A C t Hoen, Jean Monlong, Manuel A Rivas, Mar Gonzàlez-Porta, et al. 2013. “Transcriptome and Genome Sequencing Uncovers Functional Variation in Humans..” *Nature* 501 (7468): 506–11. doi:10.1038/nature12531.
Lemire, Mathieu, Syed H E Zaidi, Maria Ban, Bing Ge, Dylan Aïssi, Marine Germain, Irfahan Kassam, et al. 2015. “Long-Range Epigenetic Regulation Is Conferred by Genetic Variation Located at Thousands of Independent Loci..” *Nature Communications* 6 (February). Nature Publishing Group: 6326. doi:10.1038/ncomms7326.
Li, Yang I, Bryce van de Geijn, Anil Raj, David A Knowles, Allegra A Petti, David Golan, Yoav Gilad, and Jonathan K Pritchard. 2016. “RNA Splicing Is a Primary Link Between Genetic Variation and Disease..” *Science* 352 (6285). American Association for the Advancement of Science: 600–604. doi:10.1126/science.aad9417.
MacArthur, Jacqueline, Emily Bowler, Maria Cerezo, Laurent Gil, Peggy Hall, Emma Hastings, Heather Junkins, et al. 2017. “The New NHGRI-EBI Catalog of Published Genome-Wide Association Studies (GWAS Catalog)..” *Nucleic Acids Research* 45 (D1): D896–D901. doi:10.1093/nar/gkw1133.
Manor, Ohad, and Eran Segal. 2013. “Robust Prediction of Expression Differences Among Human Individuals Using Only Genotype Information..” *PLoS Genetics* 9 (3): e1003396. doi:10.1371/journal.pgen.1003396.
Maurano, Matthew T, Richard Humbert, Eric Rynes, Robert E Thurman, Eric Haugen, Hao Wang, Alex P Reynolds, et al. 2012. “Systematic Localization of Common Disease-Associated Variation in Regulatory DNA..” *Science* 337 (6099): 1190–95. doi:10.1126/science.1222794.
Mäkinen, Ville-Petteri, Mete Civelek, Qingying Meng, Bin Zhang, Jun Zhu, Candace Levian, Tianxiao Huan, et al. 2014. “Integrative Genomics Reveals Novel Molecular Pathways and Gene Networks for Coronary Artery Disease..” Edited by Alan Attie. *PLoS Genetics* 10 (7): e1004502. doi:10.1371/journal.pgen.1004502.
Millstein, Joshua, Bin Zhang, Jun Zhu, and Eric E Schadt. 2009. “Disentangling Molecular Relationships with a Causal Inference Test..” *BMC Genetics* 10 (1). BioMed Central Ltd: 23. doi:10.1186/1471-2156-10-23.
Montgomery, Stephen B, Micha Sammeth, Maria Gutierrez Arcelus, Radoslaw P Lach, Catherine Ingle, James Nisbett, Roderic Guigó, and Emmanouil T Dermitzakis. 2010. “Transcriptome Genetics Using Second Generation Sequencing in a Caucasian Population..” *Nature* 464 (7289): 773–77. doi:10.1038/nature08903.
Monti, J, J Fischer, S Paskas, M Heinig, and H Schulz. 2008. “Soluble Epoxide Hydrolase Is a Susceptibility Factor for Heart Failure in a Rat Model of Human Disease.” *Nature* 40 (5): 529–

37. doi:10.1038/ng.129.
Musunuru, Kiran, Alanna Strong, Maria Frank-Kamenetsky, Noemi E Lee, Tim Ahfeldt, Katherine V Sachs, Xiaoyu Li, et al. 2010. "From Noncoding Variant to Phenotype via SORT1 at the 1p13 Cholesterol Locus.." *Nature* 466 (7307): 714–19. doi:10.1038/nature09266.
Nguyen, Quan, Samuel Lukowski, Han Chiu, Clayton Friedman, Anne Senabouth, Liam Crowhurst, Timothy Bruxmer, Angelika Christ, Nathan Palpant, and Joseph Powell. 2018. "Determining Cell Fate Specification and Genetic Contribution to Cardiac Disease Risk in hiPSC-Derived Cardiomyocytes at Single Cell Resolution." *bioRxiv*, February, 1–37. doi:10.1101/229336.
Pickrell, Joseph K, John C Marioni, Athma A Pai, Jacob F Degner, Barbara E Engelhardt, Everlyne Nkadori, Jean-Baptiste Veyrieras, Matthew Stephens, Yoav Gilad, and Jonathan K Pritchard. 2010. "Understanding Mechanisms Underlying Human Gene Expression Variation with RNA Sequencing.." *Nature* 464 (7289): 768–72. doi:10.1038/nature08872.
Pickrell, Joseph K, Tomaz Berisa, Jimmy Z Liu, Laure Ségurel, Joyce Y Tung, and David A Hinds. 2016. "Detection and Interpretation of Shared Genetic Influences on 42 Human Traits.." *Nature Genetics* 48 (7). Nature Research: 709–17. doi:10.1038/ng.3570.
Pierce, Brandon L, and Stephen Burgess. 2013. "Efficient Design for Mendelian Randomization Studies: Subsample and 2-Sample Instrumental Variable Estimators.." *American Journal of Epidemiology* 178 (7): 1177–84. doi:10.1093/aje/kwt084.
Roadmap Epigenomics Consortium, Anshul Kundaje, Wouter Meuleman, Jason Ernst, Misha Bilenky, Angela Yen, Alireza Heravi-Moussavi, et al. 2015. "Integrative Analysis of 111 Reference Human Epigenomes.." *Nature* 518 (7539): 317–30. doi:10.1038/nature14248.
Roberts, Angharad M, James S Ware, Daniel S Herman, Sebastian Schafer, John Baksi, Alexander G Bick, Rachel J Buchan, et al. 2015. "Integrated Allelic, Transcriptional, and Phenomic Dissection of the Cardiac Effects of Titin Truncations in Health and Disease.." *Science Translational Medicine* 7 (270). American Association for the Advancement of Science: 270ra6–270ra6. doi:10.1126/scitranslmed.3010134.
Schadt, Eric E, John Lamb, Xia Yang, Jun Zhu, Steve Edwards, Debraj GuhaThakurta, Solveig K Sieberts, et al. 2005. "An Integrative Genomics Approach to Infer Causal Associations Between Gene Expression and Disease.." *Nature Genetics* 37 (7): 710–17. doi:10.1038/ng1589.
Schmitt, Anthony D, Ming Hu, Inkyung Jung, Zheng Xu, Yunjiang Qiu, Catherine L Tan, Yun Li, et al. 2016. "A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome." *Cell Reports* 17 (8). The Authors: 2042–59. doi:10.1016/j.celrep.2016.10.061.
Shin, So-Youn, Eric B Fauman, Ann-Kristin Petersen, Jan Krumsiek, Rita Santos, Jie Huang, Matthias Arnold, et al. 2014. "An Atlas of Genetic Influences on Human Blood Metabolites.." *Nature Genetics* 46 (6). Nature Research: 543–50. doi:10.1038/ng.2982.
Shu, Le, Kei Hang K Chan, Guanglin Zhang, Tianxiao Huan, Zeyneb Kurt, Yuqi Zhao, Veronica Codoni, et al. 2017. "Shared Genetic Regulatory Networks for Cardiovascular Disease and Type 2 Diabetes in Multiple Populations of Diverse Ethnicities in the United States.." Edited by Tuuli Lappalainen. *PLoS Genetics* 13 (9). Public Library of Science: e1007040. doi:10.1371/journal.pgen.1007040.
Siu, B L, H Niimura, J A Osborne, D Fatkin, C MacRae, S Solomon, D W Benson, J G Seidman, and C E Seidman. 1999. "Familial Dilated Cardiomyopathy Locus Maps to Chromosome 2q31.." *Circulation* 99 (8): 1022–26.
Smemo, Scott, Juan J Tena, Kyoung-Han Kim, Eric R Gamazon, Noboru J Sakabe, Carlos Gómez-Marín, Ivy Aneas, et al. 2016. "Obesity-Associated Variants Within FTO Form Long-Range Functional Connections with IRX3." *Nature* 507 (7492). Nature Publishing Group: 371–75. doi:10.1038/nature13138.
Stegle, Oliver, Leopold Parts, Richard Durbin, and John Winn. 2010. "A Bayesian Framework to Account for Complex Non-Genetic Factors in Gene Expression Levels Greatly Increases Power

in eQTL Studies..” Edited by Aviv Regev. *PLoS Computational Biology* 6 (5). Public Library of Science: e1000770. doi:10.1371/journal.pcbi.1000770.

Suhre, Karsten, So-Youn Shin, Ann-Kristin Petersen, Robert P Mohney, David Meredith, Brigitte Wägele, Elisabeth Altmaier, et al. 2011. “Human Metabolic Individuality in Biomedical and Pharmaceutical Research..” *Nature* 477 (7362): 54–60. doi:10.1038/nature10354.

Talukdar, Husain A, Hassan Foroughi Asl, Rajeev K Jain, Raili Ermel, Arno Ruusalepp, Oscar Franzén, Brian A Kidd, et al. 2016. “Cross-Tissue Regulatory Gene Networks in Coronary Artery Disease..” *Cell Systems* 2 (3): 196–208. doi:10.1016/j.cels.2016.02.002.

Timpson, Nicholas J, Celia M T Greenwood, Nicole Soranzo, Daniel J Lawson, and J Brent Richards. 2018. “Genetic Architecture: the Shape of the Genetic Contribution to Human Traits and Disease..” *Nature Publishing Group* 19 (2). Nature Publishing Group: 110–24. doi:10.1038/nrg.2017.101.

van der Harst, Pim, and Niek Verweij. 2018. “Identification of 64 Novel Genetic Loci Provides an Expanded View on the Genetic Architecture of Coronary Artery DiseaseNovelty and Significance.” *Circulation Research* 122 (3): 433–43. doi:10.1161/CIRCRESAHA.117.312086.

Vandiedonck, C. 2018. “Genetic Association of Molecular Traits: a Help to Identify Causative Variants in Complex Diseases..” *Clinical Genetics* 93 (3): 520–32. doi:10.1111/cge.13187.

Wang, I-Ming, Bin Zhang, Xia Yang, Jun Zhu, Serguei Stepaniants, Chunsheng Zhang, Qingying Meng, et al. 2012. “Systems Analysis of Eleven Rodent Disease Models Reveals an Inflammatome Signature and Key Drivers..” *Molecular Systems Biology* 8 (1). EMBO Press: 594–94. doi:10.1038/msb.2012.24.

Waszak, Sebastian M, Olivier Delaneau, Andreas R Gschwind, Helena Kilpinen, Sunil K Raghav, Robert M Witwicki, Andrea Orioli, et al. 2015. “Population Variation and Genetic Control of Modular Chromatin Architecture in Humans..” *Cell* 162 (5): 1039–50. doi:10.1016/j.cell.2015.08.001.

Watanabe, Kyoko, Erdogan Taskesen, Arjen van Bochoven, and Danielle Posthuma. 2017. “Functional Mapping and Annotation of Genetic Associations with FUMA..” *Nature Communications* 8 (1). Nature Publishing Group: 1826. doi:10.1038/s41467-017-01261-5.

Westra, Harm-Jan, Marjolein J Peters, Tonu Esko, Hanieh Yaghootkar, Claudia Schurmann, Johannes Kettunen, Mark W Christiansen, et al. 2013. “Systematic Identification of Trans eQTLs as Putative Drivers of Known Disease Associations..” *Nature Genetics* 45 (10). Nature Research: 1238–43. doi:10.1038/ng.2756.

Wild, Philipp S, Tanja Zeller, Arne Schillert, Silke Szymczak, Christoph R Sinning, Arne Deiseroth, Renate B Schnabel, et al. 2011. “A Genome-Wide Association Study Identifies LIPA as a Susceptibility Gene for Coronary Artery Disease..” *Circulation. Cardiovascular Genetics* 4 (4). American Heart Association, Inc.: 403–12. doi:10.1161/CIRCGENETICS.110.958728.

Zeller, Tanja, Philipp Wild, Silke Szymczak, Maxime Rotival, Arne Schillert, Raphaele Castagne, Seraya Maouche, et al. 2010. “Genetics and Beyond--the Transcriptome of Human Monocytes and Disease Susceptibility..” *PLoS ONE* 5 (5): e10693. doi:10.1371/journal.pone.0010693.

Zhong, Hua, John Beaulaurier, Pek Yee Lum, Cliona Molony, Xia Yang, Douglas J MacNeil, Drew T Weingarth, et al. 2010. “Liver and Adipose Expression Associated SNPs Are Enriched for Association to Type 2 Diabetes..” Edited by Trudy F C Mackay. *PLoS Genetics* 6 (5): e1000932. doi:10.1371/journal.pgen.1000932.

Zhong, Hua, Xia Yang, Lee M Kaplan, Cliona Molony, and Eric E Schadt. 2010. “Integrating Pathway Analysis and Genetics of Gene Expression for Genome-Wide Association Studies..” *American Journal of Human Genetics* 86 (4): 581–91. doi:10.1016/j.ajhg.2010.02.020.

Zhu, Zhihong, Futao Zhang, Han Hu, Andrew Bakshi, Matthew R Robinson, Joseph E Powell, Grant W Montgomery, et al. 2016. “Integration of Summary Data From GWAS and eQTL Studies Predicts Complex Trait Gene Targets.” *Nature Genetics* 48 (5): 481–87. doi:10.1038/ng.3538.

459
460

5 Tables

Table 1. Recent cardiovascular eQTL resources.

Ref	Tissue	Sample size	Population
(Sigurdsson et al. 2017)	Left Atrial wall	62	European
(Heinig et al. 2017)	Left Ventricle	205	European
(Christophersen et al. 2017)	Left Atria	329	European / African American
(Koopmann et al. 2014)	Left Ventricle	129	European
(Battle et al. 2017)	Atrial Appendage	264	European / African American
(Battle et al. 2017)	Left Ventricle	272	European / African American
(Battle et al. 2017)	Aorta	267	European / African American
(Battle et al. 2017)	Tibial artery	388	European / African American
(Battle et al. 2017)	Coronary artery	152	European / African American
(Battle et al. 2017)	Adipose - Subcutaneous	385	European / African American
(Battle et al. 2017)	Adipose - Visceral	313	European / African American
(Battle et al. 2017)	Liver	153	European / African American
(Franzen et al. 2016)	Mammary artery	600	European
(Franzen et al. 2016)	Atherosclerotic aortic root	600	European
(Franzen et al. 2016)	Visceral abdominal fat	600	European
(Franzen et al. 2016)	Skeletal muscle	600	European
(Franzen et al. 2016)	Liver	600	European

Figure legends

Figure 1: *Using eQTL data to identify causal candidate gene 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).