Published in final edited form as:
Curr Opin Cardiol. 2012 May ; 27(3): 210–213. doi:10.1097/HCO.0b013e3283522198.

Genome Wide Studies of Gene Expression Relevant to Coronary Artery Disease

Jeffrey Hsu[1] and **Jonathan D. Smith**[1,2]

[1]Department of Molecular Medicine, Cleveland Clinic Lerner College of Medicine of Case Western Reserve University, Cleveland, OH 44195

[2]Department of Cell Biology, Cleveland Clinic, Cleveland, OH 44195

Abstract

Purpose of Review—Genome-wide association studies have led to the discovery of many single nucleotide polymorphisms (SNPs) associated with coronary artery disease (CAD). However, many of these SNPs are in between genes (intergenic), and presumably function through the regulation of gene expression. Microarrays that measure the expression of thousands of mRNAs have allowed investigators to study how genetic variation alters gene expression at a genome-wide level. Combining these methods have led to progress in understanding the molecular basis for the genetic susceptibility to atherosclerosis.

Recent Findings—Recent studies confirm that gene expression differences due to genetic variation play an underlying role in atherosclerosis. Expression levels of *SORT1* are negatively correlated with an intergenic risk allele on chromosome 1p13.3 that was previously associated with CAD. Increased *SORT1* expression leads to lower hepatic secretion of LDL providing a mechanistic link between a common risk variant and disease. In addition three out of thirteen newly identified CAD risk loci were found to strongly affect the expression of nearby genes. Another recent study detected variants adjacent to a newly identified atherosclerosis risk locus on chromosome 11q22 that were associated with the expression of *PDGFD*, a member of the platelet derived growth factor family.

Summary—Cataloging the genetics of gene expression provides a small but crucial molecular link between genetics and clinical phenotypes such as atherosclerosis. Thus, gene expression is an endophenotype that can lead to the discovery of the underlying genes responsible for increasing atherosclerosis risk and potential diagnostic and therapeutic targets.

Keywords

expression genomewide studies; coronary artery disease; genetics

Introduction and Definitions

In the same manner that clinical phenotypes are associated with genetic variation in genome-wide association studies (GWAS), gene transcript levels can also be associated with genotypes. These studies are called expression genome wide association studies (eGWAS)

Corresponding Author: Jonathan D. Smith, PhD, Cleveland Clinic, 9500 Euclid Avenue, NC-10, Cleveland, OH 44195, USA, Tel (216)444-2248, Fax (216)444-9404, smithj4@ccf.org.

Conflicts of interest

Jeffrey Hsu was supported by the Howard Hughes Med Into Grad Scholar program, and by the Molecular Medicine training grant from NIH National Institute of General Medical Sciences T32GM088088. Jonathan Smith is supported by NIH NHLBI RO1 HL 098193.

when both the genotyping and gene expression quantification is done on a genome-wide scale though the use of single nucleotide polymorphism (SNP) and gene expression microarrays. Transcript levels are treated as quantitative traits and a genetic locus that associates with a transcript levels is called an expression quantitative trait locus (eQTL). eQTLs that occur near the physical genomic location of the gene are known as *cis*-eQTLs, while associations occurring on far from the gene or on different chromosomes are referred to as *tran*-eQTLs. SNPs that are associated with transcript levels are commonly referred to as expression SNPs (eSNPs). The *cis*-eSNPs, vs. *trans*-eSNPs, are generally stronger, easier to replicate, and may be explained by direct regulation of adjacent gene expression. Here we highlight some recent eGWAS that illuminated genes involved in atherosclerotic CAD. RNAseq is starting to be used instead of expression microarrays, with coverage for genes and non-coding RNAs not present on the arrays, and this data can be similarly used for eQTL studies (1).

Expression studies provide mechanistic insight for CAD associated regions

A large GWAS identified common genetic variants in intergenic region located on 1p13.3 that are associated with CAD (2). Like many GWAS results, the non-coding nature of these variants made determining the mechanism of the association less-straightforward than if it were a variant within a coding region of gene. There are two plausible explanations for an intergenic GWAS finding: 1) the GWAS identified variant is in linkage disequilibrium (often co-inherited) with a functional protein coding variant (including those causing alternative splicing); or 2) the identified variant, or a variant in linkage disequilibrium with a the identified variant, functions by regulating the expression of a nearby gene. This gene regulation may be due to altering a binding site for a transcription factor or one of the various chromatin remodeling complexes, thereby altering transcription levels. Thus, eGWAS can be used to determine genetic variants that regulate gene expression, and if these same variants are associated with the clinical phenotype, then a causative link can be implied: DNA variant → gene expression change → disease susceptibility. This type of evidence is much stronger than the mere association between DNA variant and disease susceptibility, and leads to mechanistic insight.

In order to gain insight into the CAD association on chromosome 1p13.3, Schadt et al. obtained SNP genotypes and profiled gene expression from 400 human liver samples (3). This led to the identification of strong associations between expression levels of a set of genes located within 120kb of the best associated CAD SNP on chromosome 1p13.3. *CELSR2* and *PSRC1* were the genes closest to the CAD SNP, but *SORT1* expression was best associated with the CAD SNP (3). The more common allele (major allele) of the CAD SNP is associated with a greater risk for CAD and lower expression of *SORT1* and *CELSR2*. Using eQTL and LDL-C data from inbred mouse strains, Schadt et al. were able to rule out *PSCR1* as a candidate gene as only *SORT1* and *CELSR2 levels* were inversely correlated with LDL-C levels. This strongly suggested that *PSCR1* was not the causative gene. Work done by Musunuru et al (4) went on to show that over expression of *SORT1* decreased LDL-C, while *SORT1* knockdown increased LDL-C, and these changes were mediated via alteration of VLDL secretion from the liver. They also identified the causal genetic variant, rs12740374, that regulates *SORT1* expression by sequentially testing the nearby SNPs and observing their effects on gene expression in a reporter gene transfection assay. The minor allele of rs12740374 creates a binding site for the C/EBP transcription factor that increases hepatic expression of *SORT1*. It must be noted, however, that a study by Kolby et. al (5) however showed an opposite effect of *SORT1* expression with LDL-C levels in Ldlr −/− and Sort −/− mice, with Sort1 levels seemingly increasing LDL-C levels. However the directionality of the LDL-C and SORT1 levels in Musunuru et. al paper is consistent with

the findings of Linsel-Nitschke et. al (6) (albeit the latter suggests a role in LDL-C uptake due to *SORT1* rather than hepatic VLDL secretion), human genetic associations, and eGWAS experiments. A commentary by Alan Tall and Ding Ai (7) addresses in depth the potential reasons for the differences in these findings. We preformed a bioinformatic lookup study of SORT1 eQTLs in publicly available monocyte gene expression data, and we were not able to find a significant SORT1 eQTL for the CAD associated SNP, demonstrating the hepatic tissue specificity for this eQTL. Together, these studies elegantly show how a common variants found in GWAS studies can mechanistically affect gene expression and alter a disease phenotype through a previously unknown component of LDL processing and secretion.

The risk locus on chromosome 1p13.3 affects a traditional risk factor, plasma LDL levels. Gene expression studies also potentially offer guidance to explaining risk loci that do not correlate with known CAD risk factors. A recent meta-analysis of more than 100,000 samples found 13 new susceptibility loci for CAD (8). Of these 13 new loci, three of the loci, on chromosomes 6q23.2, 17p11.2 and 17q21.32 were not associated with traditional risk factors, but contained SNPs that upon a bioinformatic lookup had previously been shown to be associated with the expression of nearby genes in liver, omental fat, subcutaneous fat, monocytes or blood (9). *TCF21* gene expression in liver and omental fat correlated positively with the CAD risk allele on chromosome 6q23.2. *PEMT* and *RASD1* expression in monocytes was negatively and positively correlated, respectively, with the CAD risk allele on chromosome 17p11.2. *PEMT* encodes for phosphatidyl ethanolamine methyltransferase an enzyme that is responsible for part of the hepatic secretion of phosptidylcholine, a major and essential component of VLDL. It was recently shown that mice deficient in *PEMT* have decreased plasma cholesterol and triglycerides on the *APOE* null background (10). Interestingly, no hepatic eQTL for *PEMT* was found in multiple eGWAS studies nor does the risk allele at this locus associate with LDL levels, potentially suggesting an alternate pathway of action. *UBE2Z* expression in blood was positively correlated with the CAD risk allele on chromosome 17q21.32.

A GWAS study using a multiethnic cohort identified five novel CAD loci, and of these, two eQTLs were characterized (11). First, a CAD associated SNP was located within an intron in the *LIPA* gene, encoding the lysosomal acid lipase, and a bioinformatic lookup revealed that this same SNP was the strongest eSNP associated with *LIPA* mRNA levels in circulating monocytes and liver (12). Since lysosomal acid lipase plays a role in the hydrolysis of cholesterol ester stored in lipid droplets of foam cells (9), this eSNP may provide a direct mechanistic link between macrophage cholesterol metabolism and CAD. Second, a CAD associated SNP located 117 kb downstream of the *PDGFD* gene (11) was found to be a strong eQTL for *PDGFD* expression in aortic media, aortic adventitia and mammary artery (13). The CAD risk allele was positively correlated with *PDGFD* expression in only these three tissues, but not in liver, whole blood, or circulating monocytes. Despite these associations, much work is still needed to elucidate the mechanism of how these genes affect atherosclerosis.

This eQTL for *PDGFD* expression was characterized in a tissue specific manner in arterial tissue. The tissue specificity of the correlation offers guidance in what tissues to target in follow-up studies and potentially provides a clue to the role of the particular tissue in disease. However, because not every conceivable tissue can be analyzed by genome wide expression studies, the presence or absence of a QTL in a tissue is not fully informative. To understand the dynamics of sample size effects and tissue specificity in eQTLs, Dorbin et al. performed an *in silico* analysis using expression data from human liver, subcutaneous fat, and omental fat (14). They found that ~90% of *cis*-eQTLs that are identified in a small population are replicated in a larger sample set, showing the low extent of false positive *cis*-

eQTLs. However, larger sample size studies can identify additional and weaker *cis*-eQTLs than smaller studies, showing that small studies tend to have high levels of false negative findings. Although it is advantageous to start with a large sample size with more power to detect *cis*-eQTLs, it still may be worthwhile to obtain expression data in tissues not previously studied even if the sample size is modest, as the determined *cis*-eQTLs are likely to be real and tissue specificity may be important for the discovery of some eQTL. Despite valid evidence for the occurrence of tissue specific eQTLs, the claim of tissue specificity must be evaluated with caution, as small sample sizes with high false negative rates will underestimate the cross-tissue nature of most eQTLs. The eQTL found for *PDGFD* is an example where a novel eQTL was found in a tissue that was not previously characterized. However, the lack of an eQTL at a GWAS locus in a particular tissue does not preclude it from having a role in a phenotype as Dorbin et al demonstrate that with larger sample sizes, many originally identified tissue specific eQTLs are found to really be cross-tissue eQTLs. A possible explanation for this is that replication follows simple scaling laws; more power is required to detect an eQTL in an alternate tissue. Another possibility for some eQTLs is that the genotypic effects on transcript levels could be genuinely stronger in specific tissues. For instance, the transcription of a gene may be more dependent on a particular enhancer element in the liver vs. the monocyte. The implication of this is that at small sample sizes an eQTL at a CAD risk locus found in one tissue but not in another tissue does not necessarily eliminate the tissue lacking an eQTL as possibly responsible for the correlation between genotype and disease. Nonetheless, the effect size of an eQTL could be informative to its physiological importance. As atherosclerosis is a disease involving many tissues, such as circulating monocytes, hepatocytes, smooth muscle cells, and endothelial cells, cataloguing the genetics of gene expression in as many tissues possible would greatly facilitate a better understanding of the disease.

In one of the largest eGWAS studies published with 1490 samples, Zeller et al tried to assess whether circulating monocyte gene expression was informative for a set of clinical phenotypes (12). Using published GWAS results for lipids, body mass index, and blood pressure, they found that most identified GWAS loci for these phenotypes had no nearby genes expressed in monocytes. They concluded that monocyte gene expression plays no role in these traits. However, this study was informative for the chromosome 9p21 CAD risk locus, the strongest CAD risk locus identified in many GWAS, which overlaps with a long non-coding RNA, and is adjacent to two cyclin dependent kinase genes, *CDKN2B* and *CDKN2A* (12). The eSNPs at this locus that associate with *CDKN2B* expression in monocytes are not the same SNPs that associate with CAD risk (12). This is consistent with findings from other large studies in which the CAD associated SNPs also did not correlate with monocyte *CDKN2B* expression (14,17), though others have found a correlation in peripheral blood T-cells (18) and whole blood (19). Although monocytes may not be the relevant tissue for the CAD phenotype, this data suggests that the CAD risk alleles mediate disease susceptibility by a mechanism not dependent upon *CDKN2B* expression in monocytes. Harismendy et al. suggest that the CAD associated SNPs at 9p21 fall within STAT1 binding sites in an enhancer that controls expression of the non-coding RNA via interferon gamma signaling in vascular endothelial cells (20). These 9p21 gene expression studies are very controversial and the subject of much ongoing research.

Conclusion

Expression genome wide association studies provide complementary evidence to genome-wide association studies. As more CAD associated loci are discovered via GWAS, genome-wide assays of gene expression and the identification of eQTLs/eSNPs will be increasingly important in helping to determine relevant tissues and to prioritize gene targets for follow-up studies.

References and Recommended Reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

* of special interest

** of outstanding interest

1. Pickrell JK, Marioni JC, Pai AA, Degner JF, Engelhardt BE, Nkadori E, Veyrieras JB, Stephens M, Gilad Y, Pritchard JK. Understanding mechanisms underlying human gene expression variation with RNA sequencing. Nature. 2010; 464:768–772. [PubMed: 20220758]
2. Samani N, Erdmann J, Hall A, Hengstenberg C, Mangino M, Mayer B, et al. Genomewide association analysis of coronary artery disease. New England Journal of Medicine. 2007; 357:443–453. [PubMed: 17634449]
3. Schadt EE, Molony C, Chudin E, Hao K, Yang X, Lum PY, et al. Mapping the genetic architecture of gene expression in human liver. PLoS Biol. 2008; 6:e107. [PubMed: 18462017]

4**. 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; 466:714–719. Musunuru K et al. show that a common intergenic variant affects gene expression and blood lipid levels. [PubMed: 20686566]

5. Kjolby M, Andersen OM, Breiderhoff T, Fjorback AW, Pedersen KM, Madsen P, Jansen P, Heeren J, Willnow TE, Nykjaer A. Sort1, encoded by the cardiovascular risk locus 1p13.3, is a regulator of hepatic lipoprotein export. Cell Metabolism. 2010; 12:213–223. [PubMed: 20816088]
6. Linsel-Nitschke P, Heeren J, Aherrahrou Z, Bruse P, Gieger C, Illig T, Prokisch H, Heim K, Doering A, Peters A, Meitinger T, Wichmann HE, Hinney A, Reinehr T, Roth C, Ortlepp JR, Soufi M, Sattler AM, Schaefer J, Stark K, Hengstenberg C, Schaefer A, Schreiber S, Kronenberg F, Samani NJ, Schunkert H, Erdmann J. Genetic variation at chromosome 1p13.3 affects sortilin mRNA expression, cellular LDL-uptake and serum LDL levels which translates to the risk of coronary artery disease. Atherosclerosis. 2010; 208:183–189.
7. Tall A, Ai D. Sorting Out Sortilin. Circulation Research. 2011; 108:158–160. [PubMed: 21252148]

8**. 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. In addition to identifying novel CAD risk loci in a large meta-analysis, Schunkert et al. show that several of the newly identified risk loci affect adjacent gene expression in three of the novel loci. [PubMed: 21378990]

9*. 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(5):e1000932. Zhong et. al show that eSNPs in adipose and liver tissues are significantly enriched for SNPs that have been shown to associate with type 2 diabetes, and suggests a strategy using gene networks to identifying disease causing pathways. [PubMed: 20463879]

10. Cole LK, Dolinsky VW, Dyck JRB, Vance DE. Impaired phosphatidylcholine biosynthesis reduces atherosclerosis and prevents lipotoxic cardiac dysfunction in ApoE−/− Mice. Circ Res. 2011 Mar 18.108:686–94. [PubMed: 21273556]

11**. The 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 Jan.43:339–44. The Coronary Artery Disease Genetics Consortium in a meta-analysis of a multiethnic European and South-Asian cohort found eSNPs near the *PDGFD* and *LIPA* genes that also associated with coronary artery disease. [PubMed: 21378988]

12. 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 Jan.5:e10693. [PubMed: 20502693]
13. Ouimet M, Franklin V, Mak E, Liao X, Tabas I, Marcel YL. Autophagy Regulates Cholesterol Efflux from Macrophage Foam Cells via Lysosomal Acid Lipase. Cell Metab. 2011 Jun 8.13:655–67. [PubMed: 21641547]

14. Folkersen L, van't Hooft F, Chernogubova E, Agardh HE, Hansson GK, Hedin U, et al. Association of genetic risk variants with expression of proximal genes identifies novel susceptibility genes for cardiovascular disease. Circ Cardiovas Genet. 2010; 3:365–373.

15*. Dobrin R, Greenawalt DM, Hu G, Kemp DM, Kaplan LM, Schadt EE, et al. Dissecting Cis Regulation of Gene Expression in Human Metabolic Tissues. PLoS ONE. 2011; 6:e23480. Dorbin et al. through simulations on their data in three tissues show that eQTLs found in modestly powered studies remain highly replicable in larger samples sizes. [PubMed: 21912597]

16. Cunnington MS, Keavney B. Genetic mechanisms mediating atherosclerosis susceptibility at the chromosome 9p21 locus. Curr Atheroscler Rep. 2011; 13:193–201. [PubMed: 21487702]

17. 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; 30:620–7.

18. 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. [PubMed: 19343170]

19. Jarinova O, Stewart AFR, 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.

20. 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; 470:264–268. [PubMed: 21307941]

Key Points

- Genome-wide expression studies have greatly aided the understanding and elucidation of the genetic effects on clinical traits such as coronary artery disease.
- eQTLs in a diverse set of tissues will be necessary to capture all the genetic effects on gene expression.
- The utility and power of eSNPs to localize genes of interest in coronary artery disease has been demonstrated and will be of increasing importance in future studies.