Making Sense of Genome-Wide Association Studies
Integrating Genetic Variation With Gene Expression to Derive Functional Mechanisms Underlying Disease Risk

Hsiao-Huei Chen, PhD; Alexandre F. R. Stewart, PhD

Genome-wide association studies (GWAS) have made enormous strides in identifying common genetic variants that contribute to common diseases\(^1\,2\) and disease-associated traits.\(^3\) Most of these genetic variants occur in noncoding sequences within introns of genes or in regions of the genome between genes, so called gene deserts, suggesting that they affect gene expression and regulation rather than the sequence of the expressed proteins. Genetic variants do not just affect gene expression in their vicinity but can affect gene expression on different chromosomes. This reflects the finite geometry of the genome within the nucleus, where transcription factor complexes serve multiple active templates. Thus, a variant that disrupts gene expression by affecting the recruitment of transcription factors on 1 chromosome can also have a strong effect at genes on other chromosomes.

When a genetic variant is associated with the altered expression of a messenger RNA (mRNA) of a gene on the same chromosome (in cis) or on a different chromosome (in trans), it is referred to as an expressed quantitative trait locus (eQTL). A growing number of studies sampling different cell types that have undergone genome-wide genotyping using commercial single nucleotide polymorphism (SNP) arrays as well as whole-genome mRNA expression analysis have identified many eQTLs, providing a wealth of information about the networks of genes influenced by particular genetic variants. Many of these studies are collected in online databases (http://www.ncbi.nlm.nih.gov/projects/gap/eqtl/index.cgi).

One limitation of these databases is that although they may provide information related to the effects of genetic variants on baseline expression of genes, no information is available pertaining to the clinical phenotypes of the individuals from which the tissues were obtained. Another limitation is that possible interactions of disease with specific eQTLs also cannot be ascertained using these eQTL databases. A recent study by the Framingham Heart Study provides a huge advance, marrying eQTL information from individuals with extensive phenotyping information, enabling significant improvements in the identification of eQTLs and their position within functional networks.\(^4\)

Mechanistic Insight From Mediation Analysis
An important contribution of the Yao et al\(^4\) study in this issue of Circulation is their use of mediation testing to address potential mechanisms whereby newly identified eQTLs may be contributing to particular traits, such as elevated low-density lipoprotein cholesterol or triglyceride levels. Mediation analysis is an approach that seeks to identify and account for the mechanism that underlies an observed relationship between an independent variable (eg, a SNP) and a dependent variable (eg, low-density lipoprotein cholesterol) via the inclusion of a third explanatory variable, known as a mediator variable (eg, an eQTL-dependent mRNA). For example, using this approach, they found that PCSK7 not only is a cis-regulated target of a genetic variant in Znfl259 (rs964184), but that PCSK7 mediates part of the effect of this variant on low-density lipoprotein cholesterol and triglyceride levels. Thus, PCSK7 may be a therapeutic target to lower low-density lipoprotein cholesterol and triglycerides.

The authors performed eQTL analysis using RNA extracted from blood (where the bulk of the RNA would come largely from peripheral blood mononuclear cells) of individuals whose genomes had undergone SNP genotyping in the Framingham Heart Study. Peripheral blood mononuclear cells, predominantly lymphocytes and monocytes, are a readily available source of cells from patients who give a blood sample for genetic studies, but are likely to reflect only a fraction of the cis- and transregulated eQTLs associated with specific genetic variants due to tissue selection bias. Many genetic variants are likely to have tissue-specific effects not manifest in peripheral blood but only present in vascular endothelial or smooth muscle cells, for instance. Of the 1077 SNPs linked to 21 CAD-associated phenotypes that they considered for eQTL analysis, 424 cis-eQTLs and 44 trans-eQTLs were identified (with a few having both cis and trans effects), indicating that 642 (60%) did not have a detectable effect on gene expression in blood. Nonetheless, when seeking replication for their finding in silico using the blood eQTL browser (http://genenetwork.nl/bloodeqtlbrowser/) that includes meta-analysis eQTL data from peripheral blood of 5,311 individuals with genome-wide SNP genotyping (but without phenotypic information), of the 240 cis-eQTLs and 25...
trans-eQTLs that were reported in the online database, 166 cis-eQTLs (69%) and 25 trans-eQTLs (100%) showed the same associated genes and direction of altered expression as in their own dataset. That 31% of the cis-eQTLs did not replicate the associated gene or direction of altered expression is worrisome, because it suggests that a more stringent false discovery rate (FDR) threshold may be required for cis-eQTLs. On the other hand, that all of the trans-eQTLs showed the same affected genes and direction of change indicates the trans-eQTLs more robustly reflect the functional consequence of the variant alleles than do the cis-eQTLs.

Where Is the 9p21.3 Locus? Glaringly absent from their study is the 9p21.3 risk locus, the first such common variant associated with the risk of coronary artery disease to yield to the GWAS approach. None of the cis- or trans-eQTL they identified was at the 9p21.3 risk locus. A possible reason for this absence may be that the specific cell type where the 9p21.3 variants alter gene expression has not yet been annotated for eQTLs. For example, arterial smooth muscle cells are clearly a missing eQTL dataset that might shed significant information regarding coronary atherosclerosis given that a recent breakthrough study also published in Circulation showed that vascular smooth muscle cells transdifferentiated to foam cells are the prominent cell type in human atherosclerotic lesions, rather than macrophages. In summary, in contrast to prior studies that have relied on public databases to identify eQTLs, the study by Yao et al took a direct approach measuring gene expression in peripheral blood for individuals from the Framingham Heart study for which a wide and deep phenotypic dataset is available. Thus, these authors were able to identify not only significant pathway enrichment from GWAS SNPs, but also to begin to address mechanisms through a mediation analysis. It will be exciting to apply this very robust approach in future studies that exploit eQTL databases not yet currently available for additional cell types relevant to coronary artery disease such as macrophages and arterial smooth muscle cells. It should be pointed out that the Systems Genetics Resource at the University of California Los Angeles provides eQTL information for 147 human samples of mixed cultures containing vascular endothelial cells and smooth muscle cells (http://systems.genetics.ucla.edu/data/SMC) that could have been mined in the Yao et al study. It will also be important to expand the eQTL databases to include basic, clinically important characteristics (eg, blood lipids, blood pressure, smoking history, etc) so that future analyses can replicate and elaborate new functional associations. Should the Framingham Heart Study make their blood eQTL and phenotype data available as a searchable online database, this would be a formidable resource.

Sources of Funding
Dr Chen is supported by a Mid-Career Investigator Award by the Heart and Stroke Foundation of Ontario (Canada).

Disclosures
None.

References

Downloaded from http://circ.ahajournals.org/ by guest on June 2, 2017


8. WTCCC. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007;447:661–678.


Making Sense of Genome-Wide Association Studies: Integrating Genetic Variation With Gene Expression to Derive Functional Mechanisms Underlying Disease Risk
Hsiao-Huei Chen and Alexandre F. R. Stewart

_Circulation._ 2015;131:519-521; originally published online December 22, 2014; doi: 10.1161/CIRCULATIONAHA.114.014634

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circ.ahajournals.org/content/131/6/519

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in _Circulation_ can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to _Circulation_ is online at:
http://circ.ahajournals.org//subscriptions/