

GENOMICS AND SYSTEMS BIOLOGY APPROACHES IN THE STUDY OF LIPID DISORDERS

ALEJANDRA RODRÍGUEZ¹ AND PÄIVI PAJUKANTA^{1,2,3}¹Department of Human Genetics, David Geffen School of Medicine; ²Molecular Biology Institute and³Bioinformatics Interdepartmental Program, University of California, Los Angeles, USA

ABSTRACT

Cardiovascular disease (CVD) is a broad definition for diseases of the heart and blood vessels with high mortality and morbidity worldwide. Atherosclerosis and hypertension are the most common causes of CVD, and multiple factors confer the susceptibility. Some of the predisposing factors are modifiable such as diet, smoking, and exercise, whereas others, including age, sex, and individual's genetic variations contributing to the CVD composition traits, are non-modifiable. This latter group includes serum lipid traits. High serum lipid levels, specifically high levels of serum low-density lipoprotein cholesterol and triglycerides, are well-established key risk factors of atherosclerosis. This review will discuss genomics and systems biology approaches in the study of common dyslipidemias. The non-Mendelian forms of dyslipidemias are highly complex, and the molecular mechanisms underlying these polygenic lipid disorders are estimated to involve hundreds of genes. Interactions between the different genes and environmental factors also contribute to the clinical outcomes; however, very little is known about these interactions and their molecular mechanisms. To better address the complex genetic architecture and multiple properties leading to high serum lipid levels, networks and systems approach combining information at genomic, transcriptomics, methylomics, proteomics, metabolomics, and phenome level are being developed, with the ultimate goal to elucidate the cascade of dynamic changes leading to CVD in humans. (REV INVEST CLIN. 2018;70:217-23)

Key words: Allele specific expression. Cardiovascular disease. Coronary heart disease. Epigenomics. Genome-wide association study. High-density lipoprotein cholesterol. Low-density lipoprotein cholesterol. Metabolomics. Network. Personal medicine. Protein-protein interaction. Quantitative trait locus. Single nucleotide polymorphism. Systems genetics. Transcriptomics. Triglycerides. Type 2 diabetes. Weighted gene co-expression network analysis.

Corresponding author:

*Päivi Pajukanta

Department of Human Genetics
David Geffen School of Medicine at UCLA
Gonda Center, Room 6335B
695 Charles E. Young Drive South
Los Angeles, California 90095-7088, USA
E-mail: ppajukanta@mednet.ucla.edu

Received for publication: 02-05-2018

Approved for publication: 13-06-2018

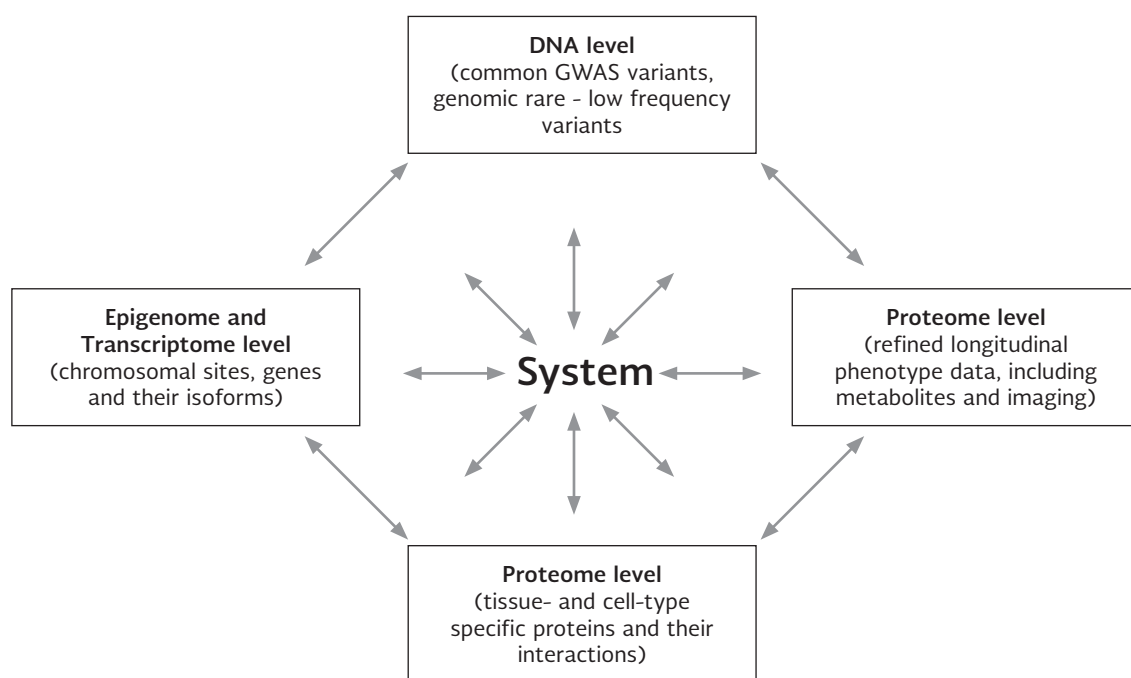
doi: 10.24875/RIC.18002576

INTRODUCTION

The Human Genome Project¹, HapMap Project², and 1000 Genomes Project³ laid the foundation to high-throughput genomics approaches to search for DNA sequence variants and their molecular and transcriptional characteristics, contributing to serum lipid levels and cardiovascular disease (CVD). In genome-wide association studies (GWAS) of single traits, hundreds of thousands of individuals have been genotyped for millions of genotyped and imputed single nucleotide polymorphisms (SNPs) throughout the human genome to search for lipid-associated variants at the DNA level. GWAS have been highly successful in identifying novel common variants for serum lipid levels, CVD, and other disorders⁴. Interestingly, one of the earlier GWAS meta-analysis for coronary heart disease (CHD) already supported the diverse pathophysiology of CVD, because only 3 of the 13 new loci identified in that study were associated with traditional CHD risk factors, such as serum lipid levels⁵. Taken together the common variants identified in lipid GWAS exhibit small effects on lipid phenotypes and typically explain ~10-20% of the entire phenotypic variance of which some 50% is estimated to be heritable⁶.

Overall, the GWAS results with common variants suggest that other sources of variation such as rare and low-frequency variants; gene-gene and gene-environment interactions, and above DNA factors, such as epigenetic changes, each explains part of the remaining variance. Furthermore, the fact that single variants do not alone explain a large proportion of lipid levels or CVD clearly indicates that networks and systems approaches are needed to complement the single variant analysis and better model the complex interactive properties involved in the development and etiology of dyslipidemias⁷. There are several different types of biological networks such as protein-protein interaction (PPI) networks, metabolic networks, and transcript regulatory networks (reviewed in 8). These biological networks represent the functional or physical connectivity between genes⁸ and thus help identify closely connected genes, interaction patterns, and enrichment in functional categories. Systems genetics is based on the principle that a complex system has intrinsic features and interactions that cannot be simply derived from the additive effects of the individual parts (Fig. 1)⁷. Accordingly, the complex architecture and molecular mechanisms of polygenic dyslipidemias could be tackled using integrative systems approaches that better model the

Figure 1. Complex systems have intrinsic features and interactions that cannot be simply derived from the additive effects of the individual parts.



complex genetic responses to environmental factors, such as diet and lifestyle.

HIGH-THROUGHPUT SEQUENCING SEARCHES FOR VARIATION AT THE DNA AND RNA LEVEL

Some lipid GWAS loci overlapped with genes implicated for monogenic forms of dyslipidemias, suggesting that in general GWAS loci could be excellent targets for resequencing to identify novel disease-causing rare variants^{9,10}. Targeted sequencing of GWAS loci has indeed identified rare coding variants with stronger functional effects, for example, in hypertriglyceridemic subjects¹¹. Furthermore, exome sequencing and whole genome sequencing using efficient and large-scale massive parallel sequencing approaches enable the identification of the full spectrum of variants, including rare and low-frequency variants, which may exhibit stronger effects than the common variants on lipids. Consequently, these variant types may also be easier to link to the underlying molecular mechanisms than common variants with subtle effects, especially when utilizing family-based sequencing, in which cosegregation of the variants with the lipid trait can be monitored and combined with quantitative trait locus (QTL) analysis, linkage analysis, and/or identity-by-descent sharing among the family members^{12,13}. Mining longitudinal electronic health records and Biobank data should further facilitate GWAS discoveries¹⁴, especially beyond the European populations that have been disproportionately represented in GWAS scans, which have left many minority populations with high dyslipidemia susceptibility, such as Latinos, under-investigated in GWAS. Thus, there is still a pressing need to explore the effects of common variants in large enough non-European discovery cohorts to identify common risk variants with population-based differences in allele frequencies.

Exome sequencing has rapidly become the state-of-the-art golden standard method in gene identification of monogenic forms of rare dyslipidemias and other diseases^{15,16}, whereas its feasibility in unrelated individuals with complex common forms of dyslipidemias has been less successful. This suggests that much larger sample sizes, likely in the range of high hundreds of thousands – few millions, are required for

successful whole genome sequencing studies of dyslipidemias to discover new rare and low-frequency risk variants.

RNA sequencing provides a link from genome to transcriptome when investigating the genetics of dyslipidemias¹⁷. It produces information not only about the variants of genes expressed in the particular tissue but also allows us to interrogate transcriptomes at unprecedented resolution in quantification of gene expression levels through the entire length of the transcript; characterization of known and novel splice forms; investigation of the currently largely unknown allele-specific expression; and joint analyses of these data for sequence variation and disease phenotypes¹⁷. A careful quality control is, however, essential in RNA sequencing to avoid errors due to lane, strand, and position biases¹⁸.

Meta-analysis of multiple RNA sequencing cohorts has remained challenging, and this lack of large RNA sequencing cohorts has so far hindered especially the transcriptomics analyses that require extensive sample sizes, such as trans eQTL analysis, which inherently involves a large multiple testing issues. Methods imputing the cis component of expression, such as the transcriptome-wide association study approach¹⁹, have helped leverage the sparse RNA sequencing data by testing association between gene expression and phenotypes using GWAS summary statistics alone through imputing the heritable CIS component of gene expression. This helps investigators to impute expression data into larger study cohorts, and thus facilitate the integration of DNA level data with transcriptomics analysis.

Cellular heterogeneity of human tissues forms another pressing challenge to the interpretation of bulk RNA sequencing data. Thus, there is an urgent need to develop methods for decomposition of cell-type specific proportions in gene expression. An even harder task is to perform the actual deconvolution, i.e., to measure the cell-type specific expression of each gene, which could then be used in downstream cell-type specific *cis* and *trans* eQTL, differential gene expression, and network analyses. Single-cell RNA sequencing can identify cell populations with distinct expression patterns whereas bulk RNA sequencing provides a dynamic view of gene expression across all cells. Thus, single-cell RNA-sequencing data could

greatly help deconvolute bulk RNA sequencing data to enable these highly desirable cell-type-specific transcriptomics analyses to identify cell-type specific expression changes underlying dyslipidemias and other clinical conditions.

METABOLOMICS NETWORKS IDENTIFY NOVEL GENES FOR DYSLIPIDEMIAS

In genetic studies, it is crucial to investigate the phenotypes with the closest links to the genotypes. Man-made diagnostic criteria, cumulative disease endpoints, and outcomes may not best reflect the critical underlying genotype-phenotype link. Recent advances in measuring several hundred - thousands of metabolomics phenotypes, including composition of different lipid particles and from serum samples, for example, by nuclear magnetic resonance²⁰ have made it possible to search for both the heritability estimates of these parameters and their correlations with traditional lipid measurements, such as triglycerides (TGs), high-density lipoprotein cholesterol (HDL-C), total cholesterol (TC), low-density lipoprotein cholesterol (LDL-C), and the outcome, CHD. Re-investigation of lipid GWAS data using heritable lipid metabolomics phenotypes measured by nuclear magnetic resonance has demonstrated how common SNPs explain more of the trait variance of lipid metabolomics phenotypes than of the serum composite lipid measurements, TGs, HDL-C, TC, and LDL-C²⁰.

A previous study introduced metabolomics networks as an approach to identify novel genes for atherosclerosis, combining existing GWAS data, metabolomics networks, and gene expression data²¹. The authors investigated ~6,600 individuals with serum metabolomic profiles and GWAS data and identified 11 metabolic networks for a multivariate genome-wide association analysis using the correlation structure of 130 metabolites. Overall, they detected 34 genome-wide significant loci, of which 7 were new. When the authors compared these data to univariate tests, they noticed that multivariate association analysis identified almost twice as many significant associations as univariate tests²¹. These data were also further supported by multi-tissue gene expression studies, suggesting that metabolomics networks help identify additional signals underlying CVD.

SOME GWAS VARIANTS REGULATE TRANSCRIPTION

Systems genetics utilizes systems analysis of phenotypes and genetic variants at multiple levels, and it has been an especially useful approach to follow-up variants identified in GWAS (reviewed in 7, 8). For example, a sequence variant can influence the expression of a gene nearby (cis-eQTL) or in a distant locus (trans-eQTL). As only a few of the common variants in the GWAS studies reside in the coding regions, many GWAS studies have explored whether the non-coding genome-wide significant variants have regulatory effects by investigating them in a cis-eQTL analysis of relevant tissues^{9,22}. Active development of methods in the area of colocalization of GWAS and eQTL signals has also facilitated these efforts^{23,24}. Kathiresan et al. originally identified several genome-wide significant lipid SNPs that altered liver expression of a regional gene not previously known to be involved in CVD⁹. These observations were made in more than 900 human liver RNA samples, suggesting a mechanistic role for the novel variants and genes in a tissue highly relevant to lipid metabolism.

Variants regulating epigenomic sites and molecular phenotypes also contribute to gene expression as was earlier demonstrated by a pioneering study that utilized DNase I sequencing to measure chromatin accessibility in 70 HapMap lymphoblastoid cell lines²⁵. The authors showed that DNase I sensitivity QTLs (dsQTL) are important determinants of human expression variation as more than 50% of eQTL, SNPs are also dsQTLs²⁵. Similarly, multiple molecular phenotypes, including histone marks and chromosomal interactions at promoters, enhancers, and other chromosomal annotation sites, have been measured by various sequencing techniques such as chromatin immunoprecipitation-seq, assay for transposase-accessible chromatin-seq, and chromatin conformation capture (Hi-C)-seq, in a large cascade of recent studies to elucidate the functional role of these sites in explaining GWAS signals (reviewed recently in ²⁶). These techniques provide epigenomics data that can be integrated with gene expression data to identify variant-specific molecular effects of GWAS variants and other risk/protective variants²⁶. Splice eQTL studies have also shown that regulation of splicing is an important functional mechanism of GWAS variants²⁷. Despite these methodological

advances to assess gene expression and epigenomic marks, the major challenge of GWAS studies has been the relatively slow conversion of the identified association signals to new functional knowledge and mechanistic insight²⁶. This is an inherent problem of a DNA-level genome-wide scan in which most of the screened common variants are either intronic or intergenic and often belong to extended linkage disequilibrium (LD) intervals with multiple regional variants in tight LD.

A pioneering previous study utilized a systems genetics approach for the further investigation of the 48 genes implicated by GWAS for type 2 diabetes (T2D) to link GWAS variants to pathways and biological networks²⁸. The authors explored coexpression patterns of these 48 genes with genes expressed in human islets from 63 donors, of which 9 had T2D. The authors identified gene coexpression and PPI networks associated with islet insulin secretion and HbA1c. When they integrated the data to form a rank list of putative T2D genes, they observed that expression variation of the top 20 genes explained 24% of the variance in HbA1c. The study demonstrated how systems genetics can be utilized to connect T2D-associated SNPs to pathways relevant for human islet function and pathogenesis of T2D²⁸.

TRANSCRIPT NETWORKS IDENTIFY NOVEL CONNECTIONS BETWEEN TRANSCRIPTS, PHENOTYPES, VARIANTS, AND METHYLATION SITES

Conventional differential gene expression analysis between two study groups suffers from correcting for multiple testing of thousands of transcripts. Few signals survive or exhibit strong enough fold changes between the tested groups, especially when investigating cases and controls of complex traits. Pathway analysis of the most differentially expressed lists of genes can partially circumvent this issue (reviewed in 8). In addition, gene set enrichment analysis has been widely used to identify enriched gene sets in transcriptomics^{8,29}. Transcript networks can also be built to explore transcript profiles from a systems perspective. Several methods have been developed for this purpose (reviewed in 7,8). One such method is the weighted gene coexpression network analysis (WGCNA)^{30,31} that is designed to identify modules of

densely interconnected genes by searching for genes with similar patterns of connectivity. The coexpression networks (modules) are constructed by correlation measures followed by hierarchical clustering and dynamic tree cut for module detection^{30,31}. The gene coexpression modules summarize the main patterns of variation. The first principal component represents the summary of the module and is referred to as the module eigengene. Thus, WGCNA can be utilized to identify gene coexpression modules summarizing the main trends in a variation on the transcriptome data and subsequently the coexpression modules can be linked to clinical traits or binary outcomes and DNA/RNA variants (Fig. 2). Relating network modules instead of genes to a clinical trait is a major advantage of WGCNA. This module-based analysis will significantly reduce the burden of multiple testing to the number of modules and examined phenotypic traits instead of accounting for tens of thousands of multiple comparisons as is done in conventional differential expression analyses. Furthermore, the network modules that are significantly correlated with the disease status, phenotypic traits, and/or DNA/RNA sequence variants may subsequently be analyzed using pathway tools, such as DAVID^{32,33} to identify biologically meaningful functional enrichment categories. WGCNA also allows for investigation of module preservation between study samples and across tissues^{30,31}. Furthermore, WGCNA can utilize gene expression data derived from both microarrays and RNA sequencing³⁴.

Recently, WGCNA has also been utilized for genomic analysis of methylation data using the quantitative methylation intensity values (beta values) as estimates of methylation³⁵. These methylation intensity values are conceptually equivalent to the expression intensity since unmethylated promoters are associated with genes that are up-regulated and methylated promoters with genes that are downregulated. Importantly, gene expression and methylation networks can be compared across tissues; furthermore, the coexpression and comethylation modules can be statistically correlated with phenotypic traits and sequence variant data. Thus, network approaches such as WGCNA allow a combination of data at multiple levels, including phenotype, DNA variation, transcript, and methylation data. There are several examples of WGCNA analysis identifying trait- and disease-associated transcript networks in animal and human studies^{34,36,37}.

Figure 2. Weighted gene coexpression analysis identifies coexpression network modules that can be correlated with phenotypic traits. Correlations of module eigengenes (ME) are shown with three phenotypic traits. The rows are labeled by the ME color, and the columns are labeled by the clinical trait. The correlation coefficients are shown for each cell, and *p*-value for the significance of the correlation is depicted in parentheses.

Network module-trait-relationships			
MEgreen	-0.22 (7e-05)	-0.18 (0.001)	-0.17 (0.002)
MEblack	-0.5 (4e-22)	-0.49 (8e-22)	-0.5 (6e-23)
MEred	-0.41 (8e-15)	-0.37 (4e-12)	-0.39 (2e-13)
MEblue	-0.11 (0.04)	-0.057 (0.3)	-0.078 (0.2)
MEpink	-0.2 (2e-04)	-0.2 (2e-04)	-0.25 (4e-06)
MEmagenta	0.18 (8e-04)	0.092 (0.09)	0.17 (0.002)
MEyellow	-0.029 (0.6)	-0.14 (0.01)	-0.052 (0.3)
MEpurple	0.13 (0.02)	0.11 (0.04)	0.17 (0.001)
MEbrown	0.3 (1e-08)	0.3 (3e-08)	0.33 (6e-10)
MEturquoise	0.21 (1e-04)	0.18 (7e-04)	0.22 (6e-05)
ME grey	-0.077 (0.2)	-0.17 (0.002)	-0.13 (0.02)
	Trait 1	Trait 2	Trait 3

TOWARD PERSONALIZED OMICS

Networks and systems genetics approaches lay a foundation for personalized medicine that aims to combine and integrate biological high-throughput data at multiple levels throughout an individual's life. A pioneering study on this topic followed one individual over a 14-month period by collecting genomic, transcriptomic, proteomic, metabolomic, and autoantibody profiles from the individual to build an integrative Personal Omics Profile³⁸. Interestingly, the authors discovered extensive, dynamic changes in different molecular components and biological pathways across healthy and diseased conditions. Specifically, the authors detected extensive heteroallelic changes during healthy and diseased states and found 497 and 1,047 genes with allele-specific expression differences during the two viral infections caused by

HRV and RSV the person endured. They also observed, for example, how glucose levels increased after the RSV infection for several months³⁸, demonstrating this for the first time a link between an RSV infection and the onset of T2D in human. Similar longitudinal Omics studies in a larger number of individuals will likely reveal highly useful new information about the epigenomic and transcriptional mechanisms how humans react to environmental challenges in a context- and allele-specific ways to modify their gene expression and protein levels properly as a response to environmental cues.

In summary, high-throughput data produced at the genome, transcriptome, proteome, metabolome, and phenome level enable network and systems-based analysis of dyslipidemias. Each level is important, and refined, detailed data are needed for a proper

representation of a level, emphasizing especially the importance of careful clinical phenotyping. Recent systems-based follow-ups of GWAS variants using metabolomics, transcriptomics, and epigenomics data have started to elucidate the usefulness of the network and systems approaches. Interpretation of the system's data is clearly more challenging than that of single trait and variant data, and there is also a high demand to develop rigorous follow-up, and biology-based replication approaches for the network approaches. Despite these challenges, the network and systems approaches have a high potential to lead us toward personalized medicine of cardiometabolic and other disorders.

REFERENCES

1. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature*. 2004;431:931-45.
2. The International HapMap Consortium. A haplotype map of the human genome. *Nature*. 2005;437:1299-320.
3. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature*. 2010;467:1061-73.
4. Visscher PM, Wray NR, Zhang Q, et al. 10 years of GWAS discovery: biology, function, and translation. *Am J Hum Genet*. 2017;101:5-22.
5. Schunkert H, König IR, Kathiresan S, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet*. 2011;43:333-8.
6. Weissglas-Volkov D, Pajukanta P. Genetic causes of high and low serum HDL-cholesterol. *J Lipid Res*. 2010;51:2032-57.
7. MacLellan WR, Wang Y, Lusis AJ. Systems-based approaches to cardiovascular disease. *Nat Rev Cardiol*. 2012;9:172-84.
8. Sun YV. Integration of biological networks and pathways with genetic association studies. *Hum Genet*. 2012;131:1677-86.
9. Kathiresan S, Willer CJ, Peloso GM, et al. Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat Genet*. 2009;41:56-65.
10. Teslovich TM, Musunuru K, Smith AV, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*. 2010;466:707-13.
11. Johansen CT, Wang J, McIntyre AD, et al. Excess of rare variants in non-genome-wide association study candidate genes in patients with hypertriglyceridemia. *Circ Cardiovasc Genet*. 2012;5:66-72.
12. Civelek M, Wu Y, Pan C, et al. Genetic regulation of adipose gene expression and cardio-metabolic traits. *Am J Hum Genet*. 2017;100:428-43.
13. Belbin GM, Odgis J, Sorokin EP, et al. Genetic identification of a common collagen disease in Puerto Ricans via identity-by-descent mapping in a health system. *Elife* 2017;6:e25060.
14. Hoffmann TJ, Theusch E, Haldar T, et al. A large electronic-health-record-based genome-wide study of serum lipids. *Nat Genet*. 2018;50:401-13.
15. Ng SB, Buckingham KJ, Lee C, et al. Exome sequencing identifies the cause of a Mendelian disorder. *Nat Genet*. 2010;42:30-5.
16. Reddy MV, Iatan I, Weissglas-Volkov D, et al. Exome sequencing identifies 2 rare variants for low high-density lipoprotein cholesterol in an extended family. *Circ Cardiovasc Genet*. 2012;5:538-46.
17. Pickrell JK, Marioni JC, Pai AA, et al. Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature*. 2010;464:768-72.
18. Pickrell JK, Gilad Y, Pritchard JK. Comment on "Widespread RNA and DNA sequence differences in the human transcriptome". *Science*. 2012;335:1302.
19. Gusev A, Ko A, Shi H, et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet*. 2016;48:245-52.
20. Tukiainen T, Kettunen J, Kangas AJ, et al. Detailed metabolic and genetic characterization reveals new associations for 30 known lipid loci. *Hum Mol Genet*. 2012;21:1444-55.
21. Inouye M, Ripatti S, Kettunen J, et al. Novel loci for metabolic networks and multi-tissue expression studies reveal genes for atherosclerosis. *PLoS Genet*. 2012;8:e1002907.
22. International Consortium for Blood Pressure Genome-Wide Association Studies, Ehret GB, Munroe PB, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature*. 2011;478:103-9.
23. Hormozdizadeh F, van de Bunt M, Segrè AV, et al. Colocalization of GWAS and eQTL signals detects target genes. *Am J Hum Genet*. 2016;99:1245-60.
24. Kichaev G, Roytman M, Johnson R, et al. Improved methods for multi-trait fine mapping of pleiotropic risk loci. *Bioinformatics*. 2017;33:248-55.
25. Degner JF, Pai AA, Pique-Regi R, et al. DNase I sensitivity QTLs are a major determinant of human expression variation. *Nature*. 2012;482:390-94.
26. Roman TS, Mohlke KL. Functional genomics and assays of regulatory activity detect mechanisms at loci for lipid traits and coronary artery disease. *Curr Opin Genet Dev*. 2018;50:52-9.
27. Li YI, van de Geijn B, Raj A, et al. RNA splicing is a primary link between genetic variation and disease. *Science*. 2016;352:600-4.
28. Taneera J, Lang S, Sharma A, et al. A systems genetics approach identifies genes and pathways for Type 2 diabetes in human islets. *Cell Metab*. 2012;16:122-34.
29. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102:15545-50.
30. Zhang B, Horvath S. A general framework for weighted gene coexpression network analysis. *Stat Appl Genet Mol Biol*. 2005;4:Article: 17.
31. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559.
32. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4:44-57.
33. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*. 2009;37:1-13.
34. Haas BE, Horvath S, Pietiläinen KH, Cantor RM, Nikkila E, Weissglas-Volkov D, et al. Adipose co-expression networks across finns and Mexicans identify novel triglyceride-associated genes. *BMC Med Genomics*. 2012;5:61.
35. van Eijk KR, de Jong S, Boks MP, et al. Genetic analysis of DNA methylation and gene expression levels in whole blood of healthy human subjects. *BMC Genomics*. 2012;13:636.
36. Voineagu I, Wang X, Johnston P, et al. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature*. 2011;474:380-4.
37. Fuller TF, Ghazalpour A, Aten JE, et al. Weighted gene co-expression network analysis strategies applied to mouse weight. *Mamm Genome*. 2007;18:463-72.
38. Chen R, Mias GI, Li-Pook-Than J, et al. Personal omics profiling reveals dynamic molecular and medical phenotypes. *Cell*. 2012;148:1293-307.