Genetic associations with expression for genes implicated in GWAS studies for atherosclerotic cardiovascular disease and blood phenotypes

Xiaoling Zhang1,2, Andrew D. Johnson1,2, Audrey E. Hendricks1,2, Shih-Jen Hwang1,2, Kahraman Tanriverdi3, Santhi K. Ganesh4, Nicholas L. Smith5,6, Patricia A. Peyser7, Jane E. Freedman3 and Christopher J. O’Donnell1,2,8,*

1Division of Intramural Research, National Heart, Lung and Blood Institute, Bethesda, MD, USA 2The NHLBI’s Framingham Heart Study, Framingham, MA, USA 3Department of Medicine, University of Massachusetts Medical School, Worcester, MA, USA 4Division of Cardiovascular Medicine, Department of Internal Medicine, University of Michigan Health Care System, Ann Arbor, MI, USA 5Department of Epidemiology, University of Washington, Seattle, WA, USA 6Group Health Research Institute, Group Health Cooperative, Seattle Epidemiologic Research and Information Center, VA Office of Research and Development, Seattle, WA, USA 7School of Public Health, University of Michigan, Ann Arbor, MI, USA and 8Division of Cardiology, Massachusetts General Hospital, Framingham, MA, USA

Received December 3, 2012; Revised September 10, 2013; Accepted September 17, 2013

Genome-wide association studies (GWAS) have uncovered many genetic associations for cardiovascular disease (CVD). However, data are limited regarding causal genetic variants within implicated loci. We sought to identify regulatory variants (cis- and trans-eQTLs) affecting expression levels of 93 genes selected by their proximity to SNPs with significant associations in prior GWAS for CVD traits. Expression levels were measured by qRT–PCR in leukocytes from 1846 Framingham Heart Study participants. An additive genetic model was applied to 2.5 million imputed SNPs for each gene. Approximately 45% of genes (N = 38) harbored at least one cis-eSNP after a regional multiple-test adjustment. Applying a more rigorous significance threshold (P < 5 × 10^-8), we found the expression level of 10 genes was significantly associated with more than one cis-eSNP. The top cis-eSNPs for 7 of these 10 genes exhibited moderate-to-strong association with ≥1 CVD clinical phenotypes. Several eSNPs or proxy SNPs (r^2 = 1) were replicated by other eQTL studies. After adjusting for the lead GWAS SNPs for the 10 genes, expression variances explained by top cis-eSNPs were attenuated markedly for LPL, FADS2 and C6orf184, suggesting a shared genetic basis for the GWAS and expression trait. A significant association between cis-eSNPs, gene expression and lipid levels was discovered for LPL and C6orf184. In conclusion, strong cis-acting variants are localized within nearly half of the GWAS loci studied, with particularly strong evidence for a regulatory role of the top GWAS SNP for expression of LPL, FADS2 and C6orf184.

INTRODUCTION

Cardiovascular disease (CVD), particularly coronary artery disease (CAD) and stroke, is the leading cause of death in the United States (1) and other industrialized countries. Atherosclerosis is the main pathophysiology underlying CVD, and the burden of subclinical atherosclerosis in major arteries such as the coronary and carotid arteries predicts future risk of CVD (2,3). Beyond well-established traditional atherosclerosis risk factors, such as elevated levels of LDL cholesterol and blood pressure, there is evidence that biomarkers of hemostasis and thrombosis as well as blood cell counts may be important risk markers. For example, elevated levels of coagulation factors including fibrinogen, factor VII (FVII), factor VIII (FVIII) and von Willebrand Factor (vWF) are associated with increased risk of thrombosis and of CVD (4–6). Variation in the blood-group

*To whom correspondence should be addressed at: Division of Intramural Research, National Heart, Lung and Blood Institute, the NHLBI’s Framingham Heart Study, 73 Mt. Wayte Avenue Suite #2, Framingham, MA 01702, USA. Tel: +1 5089353435; Fax: +1 5088722678; Email: odonnellc@nhlbi.nih.gov

Published by Oxford University Press 2013.
This work is written by (a) US Government employee(s) and is in the public domain in the US.
antigen ABO is a known risk factor for venous thrombosis and may have a role in clinically apparent CAD (7). Erythrocyte disorders such as anemia and erythrocytosis are associated with hypertension and other CVDs (8), and variation in erythrocyte measures even within normal ranges has been related to mortality and prognosis in clinically apparent CAD (9). Genome-wide association studies (GWAS) have uncovered many genetic loci that are strongly associated with CVD traits (10), including clinically apparent CAD (11–13), subclinical atherosclerosis, traditional risk factors such as lipids, and blood biomarkers of hemostasis/thrombosis and hematology (8–10,14–20).

However, despite the abundance of SNP association findings, data remain sparse regarding causal genes and genetic variants within implicated loci, limiting the rapid translation of genetic findings into clinical applications. Variants discovered in GWAS, or nearby variants in linkage disequilibrium (LD), may act via various regulatory mechanisms that act transcriptionally or post-transcriptionally (21). Recently, expression quantitative trait locus (eQTL) studies in a variety of cells and tissues have identified genetic variants that strongly affect gene regulation and are often reproduced across tissues and datasets (22–24). However, aside from eQTL studies in monocytes (25) and whole blood (26), there are limited additional studies of eQTLs for CVD and its risk factors, particularly within samples of single large population.

Our study objective was to identify genetic regulatory mechanisms that may contribute to biological pathways underlying GWAS associations with CVD phenotypes. For this objective, we selected 93 genes for gene expression profiling based upon GWAS associations with CVD phenotypes. For this objective, isms that may contribute to biological pathways underlying

RESULTS

Gene expression measurements of genes at loci for clinically apparent CAD, subclinical atherosclerosis and blood risk factors

We identified a total of 93 genes from top loci in published GWAS studies associated with CAD and blood risk factor traits including lipid levels, hemostatic factors and red blood cell (RBC) indices. In addition, three housekeeping genes were measured (B2M, GAPDH and ACTB) for normalization purposes, as previously described (28). The mean gene expression levels (cycle thresholds) for all measured genes are shown in Supplementary Material, Table S1 along with a description of the rationale for gene selection. The majority of genes targeted (83/93 = 90%) showed moderate-to-high expression levels in large numbers of leukocyte samples, with a few exceptions (N = 10), namely ADH4, LPA, APOA4, APOA5, APOC4, APOB, EDNRA, FGB, HNF1A and PHACTR1. After excluding these 10 genes with low levels of expression, a total of 83 non-housekeeping genes were available for subsequent analysis.

cis-eQTL associations

Applying a regionally adjusted P-value based on the effective number of independent tests at each locus (29), we found that the expression levels of 38 genes are associated with at least one cis-eSNP as shown in Table 1. Greater than 45% of genes (38/83) harbor at least one cis-eSNP that survives a regional multiple-test correction. The top cis-eQTL results for other 45 genes not passing this threshold are shown in Supplementary Material, Table S2.

Ten genes have more than one cis-eSNP strongly associated (P < 5 × 10−8) including ABO (P < 1.12 × 10−9), FADS2 (P < 5.72 × 10−73), LPL (P < 1.21 × 10−72), APOC2 (P < 1.98 × 10−36), FNTB (P < 1.90 × 10−22), CDKN2B (P < 1.86 × 10−13), FAM117B (P < 2.93 × 10−11), C6orf184 (P < 7.07 × 10−11), CDKN2BAS–ANRIL (P < 4.58 × 10−10) and SCAR5 (P < 3.70 × 10−8) (Table 1). Regional plots and expression by genotype plots are shown in Figure 1A and C for LPL. Figure 1B and D for FADS2 and Figure 2A and C for ABO as examples. Similar plots for the other seven top genes with P < 5 × 10−8 are presented in Supplementary Material, Figs S1 and S2. Among these 10 genes, 6 show down-regulation of expression with the minor allele (negative beta coefficient), suggesting potential reduced function (Table 1). The magnitude of effect sizes ranged from 0.32 to 1.83 ΔCt units per allele, with the largest (3.6-fold) change per allele being observed for the SNP rs6993414, associated with LPL expression. The distance of the top eSNP from the index cis gene, among the 38 genes surviving multiple-test correction, ranged from 0 bases (15 eSNPs within the regulated gene region) to 94 kb (MYBPHL), with a clear trend toward gene-centricity of significant findings.

Among the top 10 cis-eQTL associations, the strongest cis-eSNPs explained between ∼1 and 17% of the expression variance (Table 1), which is generally larger than the trait variance explained by GWAS SNPs (30). For example, the cis-eSNP rs8176731 accounts for 17.7% of APOB expression variance. In order to distinguish between the contribution of the clinical/GWAS SNP and the eSNP to the gene expression variance, we performed conditional analyses for each of the top loci (Table 2). An eQTL association that disappears/attenuates markedly after conditioning upon the GWAS SNP implies that the two traits might share the same genetic basis. After adjusting for the lead GWAS SNP, the expression variance explained by top cis-eSNPs generally was not markedly attenuated, with the exception of LPL and FADS2 [associated with lipid levels (27)] and C6orf184 [associated with CAD (11)] (Table 2). Furthermore, in order to fine-map the top 10 eQTL regions, we conducted the same cis-eQTL analysis using 1000G imputation SNP data for the top 10 genes in Table 2. This strategy identified four times more cis-eSNPs at a threshold of P < 5 × 10−8 (Supplementary Material, Table S3a), which is not surprising because 1000G imputation data contain many more SNPs than 2.5 million HapMap data.

As shown in Supplementary Material, Table S3b, the top cis-eSNP result using HapMap 2.5 million SNP imputation data is different from that of using 1000 Genomes Phase 1 imputation data except for FADS2. Regional plots for LPL, FADS2 and C6orf184 are shown in Supplementary Material, Fig. S3.
<table>
<thead>
<tr>
<th>Gene</th>
<th>Clinical phenotype</th>
<th>#SNP cis-eSNP</th>
<th>Chr</th>
<th>Physical location</th>
<th>MAF</th>
<th>Major &lt; minor</th>
<th>Beta (s.e.)</th>
<th>P cis-eSNP</th>
<th>Distance from gene (bp)</th>
<th>Variance explained by top cis-eSNP (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABO*</td>
<td>Hemostatic, CAD, others</td>
<td>190</td>
<td>40</td>
<td>rs8176731</td>
<td>0.36</td>
<td>T&gt;C</td>
<td>-1.43 (0.07)</td>
<td>1.12 x 10^-3</td>
<td>intron</td>
<td>17.73</td>
</tr>
<tr>
<td>FADS2*</td>
<td>Lipid, CAD</td>
<td>144</td>
<td>44</td>
<td>rs869567</td>
<td>0.15</td>
<td>C&gt;T</td>
<td>-1.53 (0.08)</td>
<td>5.72 x 10^-3</td>
<td>1.19 x 10^-3</td>
<td>149</td>
</tr>
<tr>
<td>LPL*</td>
<td>Lipid, CAD</td>
<td>281</td>
<td>41</td>
<td>rs6993414</td>
<td>0.10</td>
<td>A&gt;G</td>
<td>-1.83 (0.12)</td>
<td>1.21 x 10^-2</td>
<td>+78148</td>
<td>5.63</td>
</tr>
<tr>
<td>APOC2*</td>
<td>Lipid, CAD</td>
<td>85</td>
<td>31</td>
<td>rs2288912</td>
<td>0.45</td>
<td>C&gt;G</td>
<td>0.82 (0.07)</td>
<td>1.98 x 10^-3</td>
<td>1.19 x 10^-3</td>
<td>-44</td>
</tr>
<tr>
<td>FNTB*</td>
<td>RBC traits</td>
<td>218</td>
<td>35</td>
<td>rs714184</td>
<td>0.07</td>
<td>G&gt;A</td>
<td>-1.00 (0.10)</td>
<td>1.90 x 10^-2</td>
<td>+18239</td>
<td>1.09</td>
</tr>
<tr>
<td>CDKN2B*</td>
<td>CAD, CAC</td>
<td>151</td>
<td>30</td>
<td>rs598664</td>
<td>0.11</td>
<td>T&gt;C</td>
<td>0.50 (0.07)</td>
<td>1.86 x 10^-3</td>
<td>+18239</td>
<td>1.09</td>
</tr>
<tr>
<td>FAM117B*</td>
<td>CAD</td>
<td>93</td>
<td>22</td>
<td>rs1971739</td>
<td>0.30</td>
<td>G&gt;C</td>
<td>0.32 (0.05)</td>
<td>2.93 x 10^-3</td>
<td>-15277</td>
<td>0.35</td>
</tr>
<tr>
<td>C6orf184*</td>
<td>CAD</td>
<td>250</td>
<td>30</td>
<td>rs7773213</td>
<td>0.53</td>
<td>C&gt;T</td>
<td>-0.36 (0.06)</td>
<td>7.07 x 10^-3</td>
<td>-13621</td>
<td>0.88</td>
</tr>
<tr>
<td>CDKN2B4S/ANRIL*</td>
<td>CAC, CAD</td>
<td>254</td>
<td>49</td>
<td>rs1360590</td>
<td>0.48</td>
<td>C&gt;T</td>
<td>-0.36 (0.06)</td>
<td>4.58 x 10^-3</td>
<td>intron</td>
<td>2.63</td>
</tr>
</tbody>
</table>

*Significant after correcting for the total effective number of independent SNPs tested (N = 3692). Twenty genes pass a global multiple-test adjusted P-value cut off (P < 1.35 x 10^-5), which was calculated as dividing 0.05 by the total effective number of independent SNPs tested across 83 genes.

The qTL positive control genes (PEX6, OAS1 and MTRR) are excluded even though they pass multiple-test corrections. Distances from genes are from the maximal boundaries of all transcript isoforms; negative values indicate upstream (5') from genes and positive values reflect positions downstream (3') from genes.

SNP, single-nucleotide polymorphism; #Eff, effective number of; Chr, chromosome; MAF, minor allele frequency; Beta, beta-coefficient; bp, base pair; CAD, clinically apparent coronary artery disease; RBC, red blood cell; CAC, coronary artery calcium; FVII, coagulation factor 7; vWF, von Willebrand factor.
Figure 1. Regional plots and box plots for gene LPL and FADS2 top cis-SNPs whose signals were attenuated markedly after adjusting for the lead GWAS SNPs. Top panel: observed $-\log_{10}(P)$ and recombination rates by chromosomal position for all associated SNPs near rs6993414 near LPL on chromosome 8 (A) and rs968567 near FADS2 on chromosome 11 (B). Association plots were conducted with the use of R. Top SNP of interest and $P$-values in each region are indicated (red diamonds). Color coding indicates the strength of LD of each SNP with the top SNP in each region (measured by $r^2$). The light blue line represents the estimated recombination rates. Gene annotations are shown as dark green lines. Middle panel: linear regression analysis of normalized log transformed $\Delta C_T$ expression values ($\Delta C_T = C_{T\text{gene}} - \text{mean CT\text{housekeeping\,genes}}$) for gene LPL with genotypes of SNP rs6993414, giving a $P$-value of $1.21 \times 10^{-52}$ and beta-coefficient of $-1.83$ (C), and for gene FADS2 with genotypes of SNP rs968567, giving a $P$-value of $5.72 \times 10^{-73}$ and $\beta$-coefficient of $-1.53$ (D). Bottom panel: regional plots for eQTL associations (red triangles) and GWAS clinical traits associations (blue circles). Colors indicate LD with the peak eSNPs (red diamonds). LPL cis-associations versus. GWAS HDL results (E), and FADS2 cis-associations versus. GWAS HDL results (F).
Co-occurrence of cis-eQTLs with GWAS results for CVD and related risk factors

We assessed the co-occurrence of eSNPs and GWAS associations with CAD and related phenotypes, by checking the associations of eSNPs within respective GWAS scans for coronary artery calcium (CAC), lipids, RBC traits, hemostatic factors and blood pressure (Table 3). Several eSNPs exhibit strong association with gene expression levels and moderate-to-strong association with CAD or one or more related traits. These eSNPs include SNPs associated with ABO (for ABO, eQTL $P < 1.12 \times 10^{-91}$, vWF $P < 1.94 \times 10^{-8}$, LDL $P < 3.45 \times 10^{-10}$), FADS2 (eQTL $P < 5.72 \times 10^{-73}$, LDL $P < 3.87 \times 10^{-77}$), LPL (eQTL $P < 1.21 \times 10^{-52}$, TG $P < 2.91 \times 10^{-104}$), APOC2 (eQTL $P < 1.98 \times 10^{-36}$, TG $P < 1.38 \times 10^{-6}$), CDKN2B (eQTL $P < 1.86 \times 10^{-13}$, LDL $P < 5.3 \times 10^{-7}$), FAM117B (eQTL $P < 2.93 \times 10^{-11}$, TC $P < 4.5 \times 10^{-13}$), C6orf184 (eQTL $P < 7.07 \times 10^{-11}$, HDL $P < 1.7 \times 10^{-3}$, MCV $P < 2.9 \times 10^{-6}$) and CDKN2BAS-S (eQTL $P < 4.58 \times 10^{-10}$, CAC $P < 1.78 \times 10^{-5}$). SNPs located in 9p21 are strongly cis associated with the expression of two genes within the same locus (CDKN2B and CDKN2BAS-S). These associations suggest functional genetic

Figure 2. Regional plots and box plots for gene ABO top cis-SNPs whose signal was not attenuated after adjusting for the lead GWAS SNPs. (A) Observed $-\log_{10}(P)$ and recombination rates by chromosomal position for all associated SNPs near rs8176731 near ABO on chromosome 9. Top SNP of interest and $P$-values in each region are indicated (red diamonds). Color coding indicates the strength of LD of each SNP with the top SNP in each region (measured by $r^2$). The light blue line represents the estimated recombination rates. Gene annotations are shown as dark green lines. (B) Linear regression analysis of normalized log transformed $\Delta$CT expression values ($\Delta$CT = CT$_{\text{gene}}$ - mean CT$_{\text{housekeeping genes}}$) for gene ABO with genotypes of SNP rs8176731, giving a $P$-value of $1.12 \times 10^{-91}$ and $\beta$-coefficient of $-1.43$. (C) ABO cis-associations versus GWAS vWF results. Regional plots for eQTL associations (red triangles) and GWAS clinical traits associations (blue circles). Colors indicate LD with the peak eSNPs (red diamonds).
Table 2. Top cis-eSNP associations after adjustment for previously reported GWAS SNPs

<table>
<thead>
<tr>
<th>Gene</th>
<th>Clinical phenotype</th>
<th>Chr</th>
<th>Top 1, cis-eSNP</th>
<th>P, cis-eSNP</th>
<th>#cis-eSNP</th>
<th>Clinical SNP used for adjustment (clinSNP)</th>
<th>R²(eSNP versus clinSNP)</th>
<th>GWAS clinical traits</th>
<th>P, clinSNP in GWAS</th>
<th>P, cis-eSNP after adjust for clinSNP</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABO</td>
<td>Hemostatic, CAD, others</td>
<td>9</td>
<td>rs8176731</td>
<td>1.12 × 10⁻⁹¹</td>
<td>120</td>
<td>rs867621</td>
<td>0.04</td>
<td>vWF</td>
<td>&lt;5.00 × 10⁻³²⁴</td>
<td>6.61 × 10⁻⁸⁷</td>
</tr>
<tr>
<td>FADS2</td>
<td>Lipid, CAD</td>
<td>11</td>
<td>rs968567*</td>
<td>5.72 × 10⁻⁷³</td>
<td>64</td>
<td>rs687289</td>
<td>0.04</td>
<td>FVIII</td>
<td>&lt;5.00 × 10⁻³²⁴</td>
<td>6.35 × 10⁻⁸⁷</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>rs579459</td>
<td>0.10</td>
<td>CAD</td>
<td>4.00 × 10⁻¹⁴</td>
<td>8.84 × 10⁻⁸¹</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>rs174546</td>
<td>0.47</td>
<td>TG, HDL, TC, LDL</td>
<td>5.40 × 10⁻²⁴</td>
<td>6.71 × 10⁻²²</td>
</tr>
<tr>
<td>LPL</td>
<td>Lipid, CAD</td>
<td>8</td>
<td>rs6993414</td>
<td>1.21 × 10⁻⁵²</td>
<td>127</td>
<td>rs174626b</td>
<td>0.32</td>
<td>CAD</td>
<td>5.30 × 10⁻³⁴</td>
<td>3.00 × 10⁻³⁴</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>rs1267891f</td>
<td>1.00</td>
<td>TG,HDL</td>
<td>1.50 × 10⁻¹⁵</td>
<td>7.80 × 10⁻¹⁵</td>
</tr>
<tr>
<td>APOC2</td>
<td>Lipid, CAD</td>
<td>19</td>
<td>rs2288912</td>
<td>1.98 × 10⁻³⁶</td>
<td>32</td>
<td>rs3779788</td>
<td>0.24</td>
<td>CAD</td>
<td>2.40 × 10⁻⁷</td>
<td>7.40 × 10⁻⁷</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>rs4420638</td>
<td>0.00</td>
<td>LDL, TC</td>
<td>8.72 × 10⁻¹⁴</td>
<td>1.20 × 10⁻¹⁴</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>rs4420638d</td>
<td>0.00</td>
<td>CAD</td>
<td>2.14 × 10⁻⁴</td>
<td>3.20 × 10⁻⁴</td>
</tr>
<tr>
<td>FNTB</td>
<td>RBC traits</td>
<td>14</td>
<td>rs7148144</td>
<td>1.90 × 10⁻²²</td>
<td>7</td>
<td>rs4466998</td>
<td>0.06</td>
<td>MCV</td>
<td>4.90 × 10⁻⁸</td>
<td>1.14 × 10⁻²⁰</td>
</tr>
<tr>
<td>CDKN2B</td>
<td>CAD, CAC</td>
<td>9</td>
<td>rs598664</td>
<td>1.86 × 10⁻¹³</td>
<td>18</td>
<td>rs4977574</td>
<td>0.08</td>
<td>CAD, CAC</td>
<td>1.35 × 10⁻²²</td>
<td>9.49 × 10⁻³⁴</td>
</tr>
<tr>
<td>FAM137B</td>
<td>CAD</td>
<td>2</td>
<td>rs1971739</td>
<td>2.93 × 10⁻¹¹</td>
<td>9</td>
<td>rs6705330</td>
<td>0.27</td>
<td>CAD</td>
<td>8.98 × 10⁻¹¹</td>
<td>4.17 × 10⁻¹¹</td>
</tr>
<tr>
<td>C6orf184</td>
<td>CAD</td>
<td>6</td>
<td>rs7773213</td>
<td>7.07 × 10⁻¹¹</td>
<td>17</td>
<td>rs9374080</td>
<td>0.68</td>
<td>MCV</td>
<td>3.70 × 10⁻⁸</td>
<td>3.20 × 10⁻²</td>
</tr>
<tr>
<td>CDKN2BAS-S*</td>
<td>CAD, CAC</td>
<td>9</td>
<td>rs1360590</td>
<td>4.58 × 10⁻¹⁰</td>
<td>18</td>
<td>rs4977574</td>
<td>0.43</td>
<td>CAD, CAC</td>
<td>1.35 × 10⁻²²</td>
<td>1.85 × 10⁻⁶</td>
</tr>
<tr>
<td>SCARA5</td>
<td>Hemoatotic</td>
<td>8</td>
<td>rs7003622</td>
<td>3.70 × 10⁻⁸</td>
<td>1</td>
<td>rs2726953</td>
<td>0.01</td>
<td>vWF</td>
<td>1.30 × 10⁻¹⁶</td>
<td>2.80 × 10⁻⁸</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>rs964133</td>
<td>0.00</td>
<td>FVIII</td>
<td>4.40 × 10⁻¹⁵</td>
<td>2.71 × 10⁻⁸</td>
</tr>
</tbody>
</table>

*rs968567 was also reported to be significantly associated with the expression level of FADS2 in other blood-related eQTL studies (Supplementary Material, Table S3a)

rs174626 is not found in GWAS catalog, and FADS2 is not found to be associated with CAD in GWAS catalog.

rs12678919, a lead SNP in TG and HDL GWAS, is also reported to be significantly associated with the expression level of LPL in lymphocytes (Supplementary Material, Table S3a)

rs4420638 was associated with C-reactive protein (P = 9 × 10⁻¹³⁹) and LDL cholesterol (P = 2 × 10⁻³⁰⁸) in GWAS catalog (APOC2 is not found to be associated with CAD in GWAS catalog).

C6orf184 is an alias of gene CCDC162P. CDKN2BAS-S is an alias for a short transcript isoform of ANRIL.

All previously reported GWAS SNPs reach genome-wide significant (P < 5 × 10⁻⁵) except 3 SNPs associated with CAD (12): rs174626 (P = 5.30 × 10⁻³), rs3779788 (P = 2.40 × 10⁻⁷), and rs4420638 (P = 2.14 × 10⁻⁴).

Chr, chromosome; SNP, single-nucleotide polymorphism; clinSNP, SNP from clinical genome-wide association study used for adjustment in conditional analyses; GWAS, genome-wide association study; CAD, clinically apparent coronary artery disease; RBC, red blood cell; CAC, coronary artery calcium.
Table 3. Comparison of GWAS results for clinical CVD traits with eQTL results

<table>
<thead>
<tr>
<th>Gene</th>
<th>Blood pressure</th>
<th>Fasting glucose</th>
<th>Total cholesterol</th>
<th>Triglycerides</th>
<th>HDL cholesterol</th>
<th>LDL cholesterol</th>
<th>TG</th>
<th>FVIII</th>
<th>vWF</th>
<th>C6orf184</th>
<th>CDKN2B</th>
<th>SCARA5</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAD rs1971739</td>
<td>2.93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FAM117B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAD rs7773213</td>
<td>7.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C6orf184</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDKN2B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCARA5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Association between gene expression and clinical lipid traits

We examined the association of the expression signal of the top 10 genes in Table 2 (cis-eSNP P < 5 × 10⁻⁷) with four clinical lipid traits measured at the same blood draw as the RNA sample collection (Examination 8): HDL, triglycerides (TG), LDL and total cholesterol (TC). The transcription abundance for LPL, a gene known to be associated with HDL, is moderately associated with HDL level (P = 0.02). Another gene, C6orf184, is more significantly associated with HDL level (P = 0.002). As the dosage of the minor allele increases, gene expression increases and HDL level decreases, as observed in the Global Lipid HDL GWAS results (27). From this targeted examination of LPL and C6orf184, we identify a possible causal molecular role of these cis-eSNPs in the regulation of lipid metabolism.

Comparison of eQTL results with publicly available GWAS results

Beyond comparing our results to available full results of GWAS scans of CAD and blood risk factor traits, we also examined the top GWAS results reported in the NHGRI GWAS catalog to assess whether our most significant cis-eSNPs are the same as SNPs reported in prior GWAS to be associated with other clinical traits. Among 377 cis-eSNPs with P < 5 × 10⁻⁷ in the current study (120 associated with ABO, 64 with FADS2, 127 with LPL), 17 cis-eSNPs were significantly associated in prior GWAS (Supplementary Material, Table S4). For example, seven eSNPs associated with ABO gene expression in our study were also associated in prior GWAS with multiple traits including regulation of the genes at these loci by the same SNPs that are associated with related biomarkers.

While these concordant associations may reflect a true causal component, they could also reflect partial LD with the actual causal variants at these loci. To address this, we plotted regional eQTL associations on the same plot with regional GWAS (clinical or biomarker traits) to visualize the concordance of signals (Figs 1 and 2). LD was also assessed between top cis-eQTLs and top GWAS SNPs, conditional analysis of eQTL associations was conducted adjusting for the top GWAS SNP, and eQTL analysis was repeated for evidence of attenuation of the top signal. As shown in Table 2, the top cis-eSNP associated with ABO expression is completely independent from the other three GWAS SNPs known to be strongly associated with CAD, vWF and FVIII, respectively, indicating that SNPs associated with expression and clinical phenotypes are mutually independent (Fig. 2C). Similar results are also observed for several other top genes (APOC2, FNTB, CDKN2B and SCARA5). However, after controlling for the TG/HDL GWAS SNP (rs174546), the signal of the top cis-eSNP for FADS2 (rs968567) is dramatically attenuated (from 5.72 × 10⁻¹⁷ to 6.71 × 10⁻²²). rs968567 is moderately associated with LDL (P < 3.87 × 10⁻⁷) and TG (P < 4.96 × 10⁻¹⁰) (Table 3 and Fig. 1F). These two SNPs are partially correlated (r² = 0.47), suggesting that this cis-eSNP, rs968567, could play a regulatory role for clinical lipid traits, although future research is warranted to confirm the functional molecular mechanism of this eSNP. Similar evidence for attenuation after controlling for the clinical GWAS SNPs was noted for LPL, FAM117B, C6orf184 and CDKN2BAS-S (Table 2).
hematological and biochemical traits, plasma E-selectin levels, pancreatic cancer and CAD. Extensive pleiotropy for the ABO gene region has previously been described (31). One eSNP, rs1535 associated with FADS2 expression level \( (P = 4.02 \times 10^{-46}) \) is associated with metabolic syndrome \( (P = 4.0 \times 10^{-7}) \) in the GWAS catalog. Out of 127 SNPs associated with LPL expression level \( (P < 5 \times 10^{-8}) \), nine have been reported to be associated with hypertriglyceridemia, HDL cholesterol, TG or metabolic syndrome at a genome-side significant \( P \)-value (Supplementary Material, Table S4). We employed a conservative threshold \( (P < 5 \times 10^{-8}) \) for selecting cis-eSNPs; however, as shown in Supplementary Material, Table S5, a greater number of cis-eSNPs are identified as the significance threshold is loosened, but proportion of SNPs that overlap between cis-eSNPs and GWAS SNPs decreases as the significance threshold becomes less restrictive (Chi-squared test result for difference across \( P \)-value categories \( P < 1.1 \times 10^{-11} \)).

**trans-eQTL associations**

In analyses to detect important trans-eQTL associations, we found SNPs strongly associated with the expression level of target genes on different chromosomes \( (P < 5 \times 10^{-8}) \) (Table 4). Expression levels of VWF, transcribed from chromosome 12, are associated with the SNP rs1354034 on chromosome 3 \( (P = 5.43 \times 10^{-20}) \) proximal to ARHGEF3. The ARHGEF3 locus (rs1354034) is also associated in trans with expression levels of COL4A2 on chromosome 13 \( (P = 7.01 \times 10^{-11}) \), a target locus that was reported in a recent large GWAS analysis for CAD (11), and also within a large analysis of CAC levels (20). Two loci on chromosome 8 (rs4909812 and rs727582) are associated with the expression of CITED2 and CCND3, respectively. Both of these target genes were found within top loci in RBC traits GWAS.

The expression level of LPL is trans-associated with rs7085130 on chromosome 10 \( (P = 3.59 \times 10^{-10}) \) located within C10orf107 region. Notably, the C10orf107 locus was associated with blood pressure levels in recent GWAS (33–35). Another interesting trans-eSNP is rs2815067, located in a trans-expression locus containing the KIF6 gene and associated with expression levels of CDKN2A \( (P = 5.06 \times 10^{-4}) \). CDKN2A is located at the 9p21.3 CAD locus, and the KIF6 region was previously postulated to be a CAD gene, although replication attempts have failed (11,36).

Among seven genes in Table 4, the expression levels of six genes are also associated with a cis-eSNP that survives a regional multiple-test correction (VWF, COL4A2, CITED2, LPL, EPO and PCCB). For LPL, COL4A2 and PCCB, it is notable that the associated cis-eSNP nearly reaches genome-wide significance \( (P < 7 \times 10^{-7}) \) (Table 1 and Supplementary Material, Fig. S4).

**In silico validation of top cis- and trans-eQTLs**

Each top cis-eSNP at \( P < 5 \times 10^{-8} \) (Table 1) was mapped to previously published eQTL results. As shown in Supplementary Material, Table S6a, of 10 genes there are expression levels with evidence of significant cis-effects \(< 100 \text{ kb}\) by top eSNPs or their proxy SNPs \( (r^2 = 1) \) in at least one other blood-related
DISCUSSION

With many complex trait GWAS results published and additional genomic data available, including data on expression, transcription factor binding and protein–protein interactions, there is growing interest in identifying genes and gene-regulatory networks that may predict the function of disease-associated loci by integrating various sources of data (43–46). In our cis-eQTL analysis of 83 cardiovascular trait-related genes within a large population sample, >45% of genes harbor eQTLs that surpass a regional multiple-test correction, suggesting that a high proportion of disease-associated loci harbor cis-acting regulatory variants. These results are consistent with recent publications (47,48) and broad surveys of allele-specific SNP expression in biologically relevant human tissues that indicate 30–50% of genes are subject to cis-acting variation (49). Notably, at least one cis-eSNP ($P < 5 \times 10^{-8}$) is significantly associated with the expression level of 10 genes from the total of 83 GWAS loci tested, indicating the advantage of combining prior GWAS knowledge with eQTL study data to efficiently distinguish potentially functional associations. This approach was successfully adopted in recent reports applying eQTL analysis to GWAS studies followed by the use of in vivo knock-out mice models to help identify the functional regulatory mechanism of 9p21 SNPs for CAD (50) and a potential drug target (SORT1) for LDL lowering (51). In our cis-eQTL results, ABO and VWF could be candidates for follow-up studies of von Willebrand factor, LPL, and FADS2 for dysregulated levels of lipids and CAD, and C6orf184 for abnormal RBC-related traits and CAD.

Up to 17% of expression variance is explained by top cis-eSNPs, an upper bound of variance that is typically larger than the variance explained by SNPs in GWAS. We performed conditional analysis to distinguish the contribution to the gene expression variance of GWAS clinical SNPs versus cis-eSNPs. After adjusting for top SNPs selected from clinical biomarker GWAS results, we found that the expression variance explained by top cis-eSNPs is generally not dramatically attenuated except for two genes associated with lipid levels (FADS2 and LPL) and gene C6orf184 associated with CAD, suggesting that cis-eSNPs are often independent of clinical GWAS SNPs or may contribute to secondary association signals at these loci.

On the other hand, several genes including C6orf184, APOC2, FADS2 and LPL showed strong eQTL associations for SNPs that are themselves among the strongest for a GWAS trait of interest, suggesting coinciding signals. rs7773213 is associated with the expression of C6orf184 (also known as CCDC162, a pseudo gene) and is also strongly associated with RBC traits. Polymorphisms associated with expression levels of APOC2, FADS2 and LPL are also strongly associated with lipid levels in the largest lipid GWAS study to date (52) and thus well powered to detect concordance between eSNPs and associations with lipid levels.

In our data, top associated SNPs in RBC GWAS results are not strongly cis-associated with their proximal gene expression. One possible reason is that the gene expression may be tissue/cell type specific, and we studied leukocyte gene expression rather than proerythroblast or erythroid cell lines. Similar observations have been reported by other studies. Folkers et al. (52) identified a variant rs6725887 affecting the expression level of a well-known myocardial infarction gene NBEAL1 in aorta media ($P = 4.25 \times 10^{-7}$), but in our data, the signal is only moderate after adjusting for multiple testing ($P = 0.08$). This provides support that using tissues/cells appropriate to the disease/trait of interest may be necessary to detect eQTL association. Data from the Genotype-Tissue Expression (GTEx) project may be helpful for such future studies (available at: http://www.broadinstitute.org/gtex/).

When we examined the NHGRI GWAS catalog (available at: www.genome.gov/gwasstudies), we found the most significant cis-eSNPs ($P < 5 \times 10^{-8}$) were significantly associated with CAD risk factors and other disease traits. For example, eSNPs associated with ABO gene expression in our study are also associated with multiple traits including hematological and biochemical traits, plasma E-selectin levels, pancreatic cancer and CAD. Further, significant ABO eSNPs are also reported by other genome-wide cis-eQTL studies in monocyte (25) and CD4+ cells (37). These strong ABO associations in both GWAS and eQTL studies indicate multiple phenotype
influences of the ABO locus (pleiotropy) that may be because of one or a few similar underlying molecular mechanisms. 

trans-eQTLs have moderate effect sizes and may be difficult to detect and replicate (53). In our trans-eQTL analyses, we found many SNPs located on different chromosomes and strongly associated with the expression level of our target genes ($P < 5 \times 10^{-8}$). Several of these trans-associations have also been reported by other studies. Using blood cells, Fehrmann et al. (26) reported a MPV lead SNP rs12485738 on 3p26 (42), which independently affects some blood coagulation genes in trans, including VWF. This trans-association is also found in our data ($P = 7.67 \times 10^{-13}$), and this SNP is also in close LD with the top trans-eSNP for VWF in our data rs1354034 ($r^2 = 0.61$). Furthermore, ARHGEF3 is a gene nearby these two trans-eSNPs, whose expression is significantly cis affected (data not shown).

There are several genes whose expression tends to be regulated by both cis- and trans-eQTLs, including LPL and VWF, both in our data (Supplementary Material, Fig. S4) and other studies. For example, we identified a trans-eSNP affecting VWF expression, and Fehrmann et al. (26) reported a cis-SNP rs4764482 affecting VWF expression in blood ($P = 2.80 \times 10^{-40}$), consistent with the result in our data ($P = 7.72 \times 10^{-6}$) for the same SNP rs4764482. However, when we conducted an interaction analysis between all the significant cis-SNPs and trans-SNPs ($P < 5.0 \times 10^{-8}$) for seven genes, there is no significant cis–trans interaction effect observed (data not shown). The absence of an interaction might occur because our current trans-eQTL analyses are not comprehensive, as we only measured 93 target genes rather than all genes in the genome.

Our study has several strengths and limitations. Previous eQTL analyses have been conducted using microarray gene expression data and genotyped/selected SNPs. In this study, we applied qRT–PCR, a gold standard for measuring mRNA expression data and genotyped/selected SNPs. In this study, we applied qRT–PCR, a gold standard for measuring mRNA expression levels, to measure expression levels in a single well-characterized large population of 1846 FHS participants.

We have selected only a subset of genes to assay at key cardiovascular risk loci; thus, we may have missed important functional and cis-regulatory roles that are not captured by the eQTL networks. Furthermore, for all but one gene (CDKN2BAS), we have selected an assay that detects an exon that is common in all transcripts for each gene; thus, we cannot address issues of alternative transcript isoforms and their regulation in most cases. Although we observed concordant associations among SNPs, gene expression and blood lipid traits, there was a temporal difference in sample collection for the majority of the CVD-related biomarkers and the collection of RNA samples for gene expression measurements. Therefore, we cannot conduct systematic correlation studies between gene expression and many of the clinically relevant CVD biomarkers. Thus, we are limited in our ability to definitively establish genetic regulatory variation as the causal molecular mechanism of outcomes and biomarkers. More experimental validation and mechanistic studies are necessary to pinpoint specific regulatory variants. However, this study may provide incremental functional information for several key GWAS loci where regulatory variation is strongly suspected to play a role in disease etiology.

In the future, with improvements in data quantity and quality for key regulatory pathways from ENCODE (54) and other large projects, integrating genomic, expression, epigenomic and protein interaction data may allow a more mature systems biology approach for insights into the networks and key functions that underlie CVD. This may in turn lead to novel molecular approaches to prevention, diagnosis and treatment.

MATERIALS AND METHODS

Study population and sample collection

Participants of the FHS Offspring cohort who attended examination 8 and had blood samples available for gene expression were included ($n = 1846$). The FHS started in 1948 with 5209 randomly ascertained participants from Framingham, MA, USA, who have undergone biannual examinations to investigate CVD and its risk factors (55). In 1971, the Offspring cohort (56, 57) (comprised of 5124 children of the original cohort and the children’s spouses) and in 2002, the third generation (consisting of 4095 children of the offspring cohort), were recruited (58). FHS participants in this study are mainly of European ancestry. The FHS was reviewed by the Boston University Medical Center Institutional Review Board, and all participants gave written informed consent. Among the FHS Offspring cohort participants included in this study, 723 are male and 890 are female. The mean age at time of blood sampling was 66.6 years with a standard deviation of 8.9 years.

Samples were processed and assayed within 2 months of completion of Offspring cohort participant visits. The PBMCs were isolated on site at the FHS center. After isolation, PBMCs were transferred into a clean tube, washed with phosphate-buffered saline, pelleted by centrifugation and lysed with RNA lysis solution (RLT; Qiagen, Germantown, MD, USA). PBMC lysates were kept at –80°C for further RNA isolations.

Expression profiling

At the Boston University School of Medicine, lysed cells were thawed and total RNA was isolated in both platelets and PBMCs using RNaseasy Mini Kits (Qiagen). RNA concentration and quality was assessed using a ND-1000 NanoDrop Spectrophotometer (NanoDrop, Wilmington, DE, USA). RNA samples were stored at –80°C until qRT–PCR.

After cDNA conversion and preamplification, quantitative reverse-transcriptase polymerase chain reaction (qRT–PCR) was performed with a high-throughput RT–PCR instrument (BioMark; Fluidigm, San Francisco, CA, USA). Preamplified cDNA samples were loaded into the DynamicArray 48.48 chips (Fluidigm). TaqMan Gene Expression Assays (Applied Biosystems, Foster City, CA, USA) were diluted and pipetted out into the DynamicArray chips. Then, the DynamicArray was placed into the NanoFlex controller. All qRT–PCR reactions were performed in the BioMark Real-Time PCR system. Protocol details were described previously (28). To minimize the batch effect in this high-throughput experiment, three repeated housekeeping genes were targeted in each plate/run to serve as the internal controls. As Life Technologies does not disclose the exact probe and primer sequences, for each gene, we provided the context sequence to which the TaqMan probe binds, and the Life Technologies Assay ID as shown in Supplementary Material, Table S7.
A total of 93 genes were chosen from top loci in published GWAS studies associated with CAD(11), CAC(20) and blood risk factor traits including hemostatic factors (17,18), lipid levels (27) and RBC traits (9) (Supplementary Material, Table S1 for the gene list). In general, a single candidate gene was selected for most loci based upon proximity to the top associated SNP, though more than one gene were targeted for a few loci [e.g., at 9p21.3: CDKN2A, CDKN2B, CDKN2BAS-S (short isoform), CDKN2BAS-L (long isoform)]. After excluding 10 genes with low levels of expression, a total of 83 non-housekeeping genes were available for subsequent analysis.

Genotyping
Genotyping was carried out as a part of the SNP Health Association Resource project (SHArE) using the Affymetrix 500K mapping array (250K Nsp and 250K Stu arrays) and the Affymetrix 50K supplemental gene focused array on 9274 individuals. Genotyping resulted in 503 551 SNPs with successful call rate >95% and HWE $P > 10^{-6}$. Finally, out of a total 9274 participants, 8481 individuals with call rate >97% were remained. Imputation of ~2.5 million autosomal SNPs in HapMap with reference to release 22 CEU samples and imputation of ~36 million SNPs (autosomal + chromosome X) using 1000 Genomes Phase 1 data were conducted using the algorithm implemented in MACH (59). There were 1846 Offspring cohort participants with genotype imputation and expression measurements available for analysis. The deposition of single-nucleotide polymorphisms imputed SNPs in dbGaP for the FHS have been are described on the dbGaP website (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?Study_id=phs000342.v4.p6).

Statistical analysis
For each gene, we excluded samples that were >5 SD away from the mean of the transformed $\Delta C_T$ expression values ($\Delta C_T = CT_{\text{gene}} - \text{mean CT}_3$ housekeeping genes). This left us to 1594 samples out of a total 1846 included in the downstream GWAS analysis.

A GWAS analysis for each of the 83 genes was conducted on 2.5 million imputed SNPs assuming an additive genetic model and adjusting for age, sex, and the mean expression level of three housekeeping genes ($B2M, GAPDH,$ and $ACTB$). SNPs positively associated and within 100 kb of each gene, respectively, were defined as cis-eQTLs, otherwise they were labelled as potential trans-eQTLs. The gene boundary for each gene is based on NCBI human reference genome builds 36/hg18, and was calculated as the maximum union of overlapping transcripts isoforms based on RefSeq genes. Between 6 ($MIA3$) and 1017 ($PRKCE$) SNPs were genotyped or imputed and within 100 kb of the genes measured (median of 183 SNPs) (Supplementary Material, Fig. S1).

Missing expression measurements for a given gene and sample were defined by measurements that were beyond detection limits (30 CT). Genes with detectable expression levels for at least 20% of the samples ($n = 83$) were analyzed as continuous outcomes using an efficient score statistic. Only four genes ($PCSK9, APOC3, EPO, SCO2$) had missing values of >50%. The efficient score statistic employed here uses a robust measure of the score variance that makes the statistic robust to departure from the normality assumption on the continuous outcome. Family structure is controlled for in a random effects term by including the matrix of the kinship coefficients in the genetic covariance matrix. Specific details are described further in Dupuis et al. (60). All models were run on expression measured from leukocyte cells. A regional multiple-test correction method by computing the effective number of independent SNPs was applied to account for multiple testing (29).

Analysis for testing the possible batch effect of gene expression data
To eliminate a possible impact on our qRT–PCR expression data from batch effects, we have first calculated the Principle Components (PCs) using our rrPCR expression data, as shown in Supplementary Material, Table S8. For each of the top 10 genes in Table 2, we recomputed the $P$-value for the top eQTL after adjusting for seven top PCs. As shown in Supplementary Material, Table S8, the adjusted $P$-value is in general slightly attenuated from the unadjusted $P$-value, but all of the $P$-values remain highly statistically significant ($P < 5.0 \times 10^{-8}$) except for $SCARA5$. These results provide further reassurance that batch effects are not playing a major role for our highly significant results.

Analysis for testing the possible impact of population stratification effects
Although all FHS participants in this study are of European ancestry, we still tested a possible impact of population stratification effects using the 10 top PCs derived from 503 551 genotyping SNPs, as shown in Supplementary Material, Table S9. For each of these top 10 PCs, we first checked for evidence of association with the expression level of each of the top 10 genes in Table 2. We found no strong evidence for population stratification effects for any of the genes (no evidence for population stratification at $P < 0.01$). For six genes, expression levels are nominally associated with one or two PCs ($0.01 < P < 0.05$). For these six genes, we recomputed the $P$-value for their top eQTLs after adjusting for those PCs nominally associated with the respective genes. As shown in Supplementary Material, Table S9, the adjusted $P$-value for each of these genes is not materially different from the unadjusted $P$-value.

Comparison of expression findings to GWAS results for disease and clinical biomarkers
GWAS results were reviewed from the CARDioGRAM (Coronary ARtery Disease Genome-wide Replication And Meta-Analysis) Consortium of CAD results including 22 233 cases and 64 762 controls of European descent (11); the Global Lipids Genetics Consortium of >100 000 individuals of European ancestry (27); the ICBP (International Consortium for Blood Pressure) GWAS comprising ~200 000 individuals of European descent (35); and GWAS for CAC ($n = 15 993$) (20), RBC traits ($n = 24 167$) (9), hemostatic factors including FVII, FVIII and vWF ($n = 23 608$) (18), and circulating fibrinogen levels ($n = 22 096$) (17) from the CHARGE (Cohorts for Heart and Aging Research in Genome Epidemiology) Consortium.
Association analysis of gene expression and clinical lipid traits

Regression analyses were performed to look for the association between gene expression measured by qRT–PCR, as the outcome, and lipid traits including HDL, TG, LDL collected at Examination 8, as predictors. To be consistent with the eQTL analysis, a linear mixed effects model was used to control for family structure, and other covariates such like age, sex and the mean of the three housekeeping genes were also included in the model. Subjects with lipid treatments at Examination 8 were excluded from the analysis.

In silico validation of cis- and trans-eQTLs

Twelve genome-wide eQTL datasets from distinct tissues including whole blood (26,41), monocytes (25), CD4+ lymphocytes cells (37), lymphocytes (38), PBMCs (39), leukocytes (40), brain (61), liver, stomach and subcutaneous adipose tissue (48) were collected from nine published articles. The sample size, definition of cis- and trans-eQTLs, and the number of SNPs and genes tested in the current publications are shown in Supplementary Material, Table S10. This catalog is limited by the prior methods and platforms of the source eQTL studies and the availability of partial, top results. The selected boundaries for defining cis-eQTLs, and coverage of SNPs, genes and trans-eQTLs varied across studies (Supplementary Material, Table S10). cis- and trans-eSNPs from the current study were cross-referenced with significant eSNPs reported in the twelve eQTL datasets directly by matching SNP rsids.

Accession numbers

The expression levels for these 93 genes and three housekeeping genes were deposited in the NCBI database of genotypes and phenotypes (dbGaP) under the Dataset Name (I-matrans_ex08_1_0552 s) and Dataset Accession (pht002077.v2,p6).

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG online.

ACKNOWLEDGEMENTS

This research was conducted in part using data and resources from the Framingham Heart Study of the National Heart Lung and Blood Institute of the National Institutes of Health and Boston University School of Medicine. The analyses reflect intellectual input and resource development from the Framingham Heart Study investigators participating in the SNP Health Association Resource (SHARE) project. The authors acknowledge the essential role of the Cohorts for Heart and Aging Research in Genome Epidemiology (CHARGE) Consortium in development and support of this manuscript, especially acknowledging members from the CHARGE Consortium Hematological, Hemostasis, and Subclinical CAC Working Groups for making their GWAS data fully available (Supplementary Material, Acknowledgement).

Conflict of Interest statement. None declared.

FUNDING

This work was supported by the National Heart, Lung and Blood Institute, Division of Intramural Research.

REFERENCES


