A statin-dependent QTL for GATM expression is associated with statin-induced myopathy

Lara M. Mangravite1, Barbara E. Engelhardt1‡, Marisa W. Medina2, Joshua D. Smith4, Christopher D. Brown5, Daniel I. Chasman6, Brigham H. Mecham7, Bryan Howie2, Heejung Shim8, Devesh Naidoo2, QiPing Feng2, Mark J. Rieder4‡, Yi Lin8, Der I. Chen6, Jerome I. Rotter7, Paul M Kidder5, Jemma C. Hopewell9, Sarah Parish7, Jane Armitage9, Rory Collins7, Russell A. Wilke2, Deborah A. Nickerson4, Matthew Stephens2,10 and Ronald M. Krauss3

Statins are prescribed widely to lower plasma low-density lipoprotein (LDL) concentrations and cardiovascular disease risk1 and have been shown to have beneficial effects in a broad range of patients2–4. However, statins are associated with an increased risk, albeit small, of clinical myopathy5 and type 2 diabetes6. Despite evidence for substantial genetic influence on LDL concentrations6, pharmacogenomic trials have failed to identify genetic variations with large effects on either statin efficacy7–9 or toxicity10, and have produced little information regarding mechanisms that modulate statin response. Here we identify a downstream target of statin treatment by screening for the effects of in vitro statin exposure on genetic associations with gene expression levels in lymphoblastoid cell lines derived from 480 participants of a clinical trial of simvastatin treatment7. This analysis identified six expression quantitative trait loci (eQTLs) that interacted with simvastatin exposure, including rs9806699, a cis-eQTL for the gene glycine amidinotransferase (GATM) that encodes the rate-limiting enzyme in creatine synthesis. We found this locus to be associated with incidence of statin-induced myotoxicity in two separate populations (meta-analysis odds ratio = 0.60). Furthermore, we found that GATM knockdown in hepatocyte-derived cell lines attenuated transcriptional response to sterol depletion, demonstrating that GATM may act as a functional link between statin-mediated lowering of cholesterol and susceptibility to statin-induced myopathy.

Analysis of individual variation in transcriptional response to drug treatment has successfully identified regulatory genetic variants that interact with treatment in model organisms11 and human tissues12–15. Cellular transcriptional analysis may be particularly useful for investigating genetic influences on statin efficacy, as statin-induced plasma LDL lowering is controlled through sterol-response element binding genetic influences on statin efficacy, as statin-induced plasma LDL lowering is controlled through sterol-response element binding protein (SREBP)–mediated transcriptional regulation16. Therefore, to identify novel regulatory variants that interact with statin treatment, we conducted a genome-wide eQTL analysis based on comparing simvastatin exposure versus control exposure of 480 lymphoblastoid cell lines (LCLs) derived from European American participants in the Cholesterol and Pharmacogenetics (CAP) trial (http://www.clinicaltrials.gov/ct2/show/NCT00451828). LCLs have proven to be a useful model system for the study of genetic regulation of gene expression17,18. Although non-genetic sources of variation, if uncontrolled, may limit the utility of LCLs for transcriptional perturbation analyses19,20, there has been increasing use of these cells to screen for genetic variants associated with molecular response to drug intervention20. Furthermore, many features of statin-mediated regulation of cholesterol metabolism are operative in LCLs21.

Simvastatin exposure had a significant effect on gene expression levels for 5,509 of 10,195 expressed genes (54%, false discovery rate (FDR) < 0.0001). The magnitude of change in expression across all responsive genes was small (0.12 ± 0.08 mean log₂ change ± s.d., Fig. 1) with 1,952 genes exhibiting ≥10% change in expression and only 21 genes exhibiting ≥50% change in expression. Among the strongest responders were 3-hydroxy-3-methylglutaryl-CoA reductase (HMGCR), which encodes the direct target of simvastatin inhibition (0.49 ± 0.29 mean log₂ change ± s.d., P < 0.0001, n = 480), and low density lipoprotein receptor (LDLR), which encodes the receptor responsible for internalization of LDL particles (0.50 ± 0.35 mean log₂ change ± s.d., P < 0.0001). As expected, surface expression of the LDLR protein was also increased following simvastatin exposure (1.6 ± 0.11 mean log₂ change ± s.d., P < 0.0001, n = 474). Gene-set enrichment analysis showed a treatment-dependent increase in expression of genes involved in lipid metabolism and cholesterol uptake.

### Figure 1 Simvastatin treatment alters transcript expression in LCLs.

Log change in expression following simvastatin- and control-exposed lymphoblastoid cell lines (n = 480) displayed as a function of the log sum of expression traits. Grey, genes for which expression was significantly changed in response to simvastatin exposure (n = 5,509 genes, 0.12 ± 0.08 mean absolute log₂ change ± s.d., q < 0.0001); black: genes for which expression was not significantly changed (n = 4,686); red: genes in the cholesterol biosynthesis pathway, all of which exhibited significant changes in expression. A.U., arbitrary units.

1Sage Bionetworks, 1100 Fairview Avenue North, Seattle, Washington 98109, USA. 2Department of Human Genetics, University of Chicago, 920 East 58th Street, Chicago, Illinois 60637, USA. 3Children’s Hospital Oakland Research Institute, 5700 Martin Luther King Jr Way, Oakland, California 94609, USA. 4Department of Genome Sciences, University of Washington, 3720 15th Avenue Northeast, Seattle, Washington 98195, USA. 5Department of Genetics, University of Pennsylvania, 415 Cune Boulevard, Philadelphia, Pennsylvania 19103, USA. 6Center for Cardiovascular Disease Prevention, Division of Preventive Medicine, Brigham and Women’s Hospital, 900 Commonwealth Drive, Boston, Massachusetts 02115, USA. 7Department of Medicine, Division of Clinical Pharmacology, Vanderbilt University Medical Center, 23rd Avenue South, Nashville, Tennessee 37232, USA. 8Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, 1124 West Carson Street, Torrance, California 90502, USA. 9Clinical Trial Service Unit and Epidemiological Studies Unit, University of Oxford, Richard Doll Building, Old Road Campus, Roosevelt Drive, Oxford OX3 7LF, UK. 10Department of Statistics, Eckhart Hall, 5734 South University Avenue, University of Chicago, Chicago, Illinois 60637, USA. 11Present addresses: Bioinformatics and Bioinformatics Department and Department of Statistical Science, Duke University, Durham, North Carolina, USA (B.E.E.); Adaptive Biotechnologies, Seattle, Washington, USA (M.J.R.).

*These authors contributed equally to this work.

©2013 Macmillan Publishers Limited. All rights reserved
Steroid biosynthesis, consistent with the mechanism responsible for the lipid-lowering response to statin, and a decrease in expression of genes involved in RNA splicing, consistent with evidence for statin regulation of alternative splicing of genes involved in cellular cholesterol homeostasis (Supplementary Fig. 1).

We first identified eQTLs without considering whether they interact with simvastatin exposure. We computed Bayes factors to quantify evidence for the association between every single-nucleotide polymorphism (SNP) and the expression level of each gene, and we used permutations to estimate FDRs (see Methods). This analysis identified 4,590 genes with cis-eQTLs, defined as eQTLs within 1 megabase (Mb) of the gene’s transcription start or end site (FDR = 1%, log_{10} Bayes factor ≥ 3.24, Supplementary Table 1). Statistical power to detect eQTLs was substantially increased by controlling for known covariates and unknown confounders (represented by principal components of the gene expression data) and by testing for association with expression unknown confounders (represented by principal components of the gene expression data) and by testing for association with expression.

As in similar studies, we found many fewer differential eQTLs than stable eQTLs or SNPs with similar effects across both conditions. The finding of relatively few gene by exposure interactions, and of relatively modest effect sizes of those interactions, seems remarkably consistent across studies regardless of method (including family-based comparisons), exposure, sample size, sample source, or the number of stable eQTLs detected. We focus further analysis on our most significant differential association from the bivariate model, the GATM locus, for which we observed stronger evidence for eQTL association after statin exposure and for which there was evidence for biological relevance to pathways involved in lipoprotein metabolism and myopathy (see Supplementary data).

GATM encodes glycine aminotransferase, an enzyme that is required for the synthesis of creatine. We observed evidence for differential eQTL expression (log_{10} BF, left y axis); SNPs associated with statin-induced myopathy (red); significance threshold (dotted line) recombination rates in centimorgans (cM) per megabase (Mb) (blue, right y axis). Bottom panel, transcribed genes (green); DNase I hypersensitive (DHS) sites and transcription factor binding sites (TFBS; black); predicted chromosomal enhancers (orange) and promoters (red) as identified in hepatocyte (HepG2), lymphoblastoid (GM12878), and myocyte (HSMCM) cell lines, using ChromHMM software (see Methods).

**Figure 2** | Treatment-specific QTL associated with GATM expression. a, Association of rs9806699 with quantile normalized GATM expression levels following control exposure (left panel, not significant); simvastatin exposure (middle left panel, log_{10} Bayes factor (BF) = 5.1, effect size = −0.43); fold change (middle right panel, log_{10} BF = 5.7, effect size = −0.40); control versus simvastatin-exposed GATM expression (right panel; black, GG, n = 225; red, GA, n = 207; green, AA, n = 48). Box height and whiskers are described in Supplementary Methods. b, Top panel, SNPs associated with GATM expression (log_{10} BF, left y axis); SNPs associated with statin-induced myopathy (red); significance threshold (dotted line) recombination rates in centimorgans (cM) per megabase (Mb) (blue, right y axis). Bottom panel, transcribed genes (green); DNase I hypersensitive (DHS) sites and transcription factor binding sites (TFBS; black); predicted chromosomal enhancers (orange) and promoters (red) as identified in hepatocyte (HepG2), lymphoblastoid (GM12878), and myocyte (HSMCM) cell lines, using ChromHMM software (see Methods).
association with *GATM* (log_{10} Bayes factor > 5.1) across a group of 51 SNPs within the *GATM* locus that are in linkage disequilibrium (chromosome 15: 45627979–45740392, hg19, r^2 = 0.85–0.99, n = 587). The most significant differential eQTL association was observed with SNP rs9806699 (minor allele frequency (MAF) = 0.32), for which we observed stronger evidence for an association with *GATM* expression following simvastatin exposure (log_{10} Bayes factor = 5.1, effect size = −0.43) than following control exposure (log_{10} Bayes factor = 0.52, effect size = −0.17, Fig. 2a). SNPs at this locus also had a stable association with expression of a neighbouring gene, *SPATASLI* (differential eQTL rs9806699 log_{10} Bayes factor = −0.33, stable eQTL rs9806699 log_{10} Bayes factor = 21.75, Supplementary Fig. 4). This locus has been shown previously to be associated with reduced glomerular filtration rate (GFR)\(^{3}\) with a small effect size (<1%). This association was specific to GFR as estimated from plasma creatinine but not from cystatin C, a second biomarker of renal function, suggesting that the association was related to variation in creatinine production rather than renal elimination. We found evidence for SNP differential association with *GATM* that spans the *GATM* coding region and includes multiple SNPs located within DNase I hypersensitive sites, active promoters and several alternative *GATM* transcription start sites (Fig. 2b).

Phosphorylation of creatine, the primary downstream product of *GATM* activity, is a major mechanism for energy storage in muscle and is mediated by creatine kinase, the primary plasma biomarker of statin-induced myopathy. To test the relationship of this locus with statin-induced myopathy, we examined the association of the *GATM* differential eQTL locus with statin-induced myopathy in a population-based cohort comprised of 72 cases of myopathy and 220 matched controls (Marshfield cohort)\(^{27}\). In this cohort, we observed that the minor allele at the *GATM* differential eQTL locus was associated with reduced incidence of statin-induced myopathy (odds ratio = 0.61, 95% confidence interval = 0.39–0.95, P = 0.03; Table 1). This association was replicated in a second cohort consisting of 100 cases of myopathy and 220 matched controls (Marshfield cohort)\(^{27}\). In this cohort, we observed that the locus with this enzyme in statin-treated populations in which statin-induced myopathy was not observed. Within CAP (40 mg per day simvastatin exposure for 6 weeks), no association of rs9806699 was observed with plasma creatine kinase either before simvastatin exposure (n = 575, P = 0.83) or following exposure (n = 574, P = 0.48). This lack of association was confirmed in a second statin study (Justification for the Use of Statins in Prevention: an Intervention Trial Evaluating Rosuvastatin (JUPITER) trial, 20 mg per day rosuvastatin, median follow-up = 1.9 years; http://clinicaltrials.gov/show/NCT00239681) both before rosuvastatin exposure (n = 8,504, P = 0.54) and after treatment (n = 3,052, P = 0.83)\(^{3}\). These findings suggest that the observed association of the *GATM* locus with risk for statin-induced myopathy is independent of an association with plasma creatine kinase. Although the present studies do not address the mechanism for the link between reduced *GATM* expression and protection from statin-induced myopathy, it is thought that diminished capacity for phosphocreatine storage modifies cellular energy storage and adenosine monophosphate-activated protein kinase (AMPK) signalling\(^{28,29}\) in a manner that is protective against cellular stress as induced by glucose deprivation\(^{26}\) or, potentially, by cholesterol depletion. Given that myocellular creatine stores are predominantly derived from renal and hepatic creatine biosynthesis, these results raise the possibility that statins may predispose to muscle toxicity in part through metabolic effects in the liver, the major site of statin’s pharmacologic actions (Supplementary Fig. 5). Conversely, the finding of severe myopathy in two cases of extreme genetic *GATM* deficiency\(^{30}\) suggests that this protective effect may be overcome if creatine synthesis is insufficient to support myocellular energy needs.

Given the influence of statin exposure on regulation of *GATM* expression, we next tested whether *GATM* may modulate sterol-mediated changes in cholesterol homeostasis. Knockdown of *GATM* in hepatocyte-derived cell lines (HepG2 and HuH7) resulted in reduced upregulation of SREBP-responsive genes (HMGCR, LDLR and SREBF2) by sterol depletion (Fig. 3a). Moreover, *GATM* knockdown decreased media accumulation of apolipoprotein B (apoB), the major structural protein of LDL, in both cell lines (P < 0.05; Fig. 3b), but did not alter levels of apoAI, the major structural protein in high density lipoproteins (HDLs, LDLs and VLDLs).

**Table 1 | Associations of SNPs at the *GATM* locus with statin-induced myopathy.**

| Cases (n) | Controls (n) | SNP | Position | LD (r^2) | MAF (cases) | MAF (controls) | Effect size | P value |
|----------|--------------|-----|----------|----------|-------------|---------------|-------------|----------|--------|
| Marshfield 72 | 220 | rs9806699 | Chr 15: 43,527,684 | 1.0 | 0.21 | 0.30 | 0.61 (0.39–0.95) | 3.2 × 10^{-2} |
| Marshfield 72 | 220 | rs1719247 | Chr 15: 43,408,027 | 0.76 | 0.19 | 0.29 | 0.59 (0.36–0.93) | 2.4 × 10^{-2} |
| SEARCH 100 | 4,021 | rs1719247 | Chr 15: 43,408,027 | 0.80 | 0.21 | 0.29 | 0.66 (0.41–1.02) | 6.4 × 10^{-2} |
| SEARCH 100 | 4,029 | rs1346268 | Chr 15: 43,460,321 | 0.70 | 0.17 | 0.25 | 0.61 (0.42–0.88) | 1.0 × 10^{-2} |
| Meta-analysis | 72 | 220 | rs9806699 | Chr 15: 43,408,027 | 0.76 | 0.19 | 0.29 | 0.59 (0.36–0.93) | 2.4 × 10^{-2} |
| Meta-analysis | 72 | 220 | rs1346268 | Chr 15: 43,460,321 | 0.70 | 0.17 | 0.25 | 0.61 (0.42–0.88) | 1.0 × 10^{-2} |
| Meta-analysis | 72 | 220 | rs1719247 | Chr 15: 43,408,027 | 0.76 | 0.19 | 0.29 | 0.59 (0.36–0.93) | 2.4 × 10^{-2} |

- **Differential eQTL associations with *GATM* expression in CAP were:** log_{10} Bayes factor = 6.22 (rs9806699), log_{10} Bayes factor = 4.35 (rs1719247), and log_{10} Bayes factor = 5.96 (rs1346268). All SNPs were in Hardy-Weinberg equilibrium in these populations. Effect size reported as odds ratio with 95% confidence interval in parentheses. Chr, chromosome; LD, linkage disequilibrium with respect to the top differential eQTL SNP, rs9806699 based on Pearson correlation (r^2).

**Figure 3 | GATM knockdown attenuated sterol-mediated induction of expression of SREBP-responsive genes.** a. Changes in transcript concentrations following sterol depletion through 24 h exposure to lipoprotein deficient serum (LPDS)-containing media versus standard FBS-containing media in hepatocyte-derived HepG2 (left, n = 12) and HuH7 (right, n = 12) cell lines. Asterisk indicates P < 0.05 for the comparison of GATM versus non-targeting control (NTC) siRNA treated cells. b. Fold changes in accumulation of apolipoprotein B (apoB) and apolipoprotein AI (apoAI) in media after gene knockdown with GATM versus (NTC) siRNA in HepG2 cells (left, n = 6–10) or HuH7 cells (right, n = 4–6) under standard culture conditions. Experiments repeated 2–3 times with 2–8 biological replicates each. Data presented as average values. Error bars, s.e.m.
METHODS SUMMARY

Gene expression levels were measured using the Illumina Human-Ref8v3 beadarray in 480 lymphoblastoid cell lines derived from European American participants in CAP, a 6-week trial of simvastatin (40 mg per day), after 24 h exposure to 2 μM activated simvastatin or control buffer. Treatment-specific effects were modelled after adjustment for known covariates and unknown confounding variables using linear regression, and eQTLs were identified using the BIMBAM software after imputing the available genotypes. Differential eQTLs were identified using the BIMBAM software (univariate test) as well as linear models of differential association (bivariate tests). Associations with myopathy were tested in two cohorts containing cases with statin-induced myopathy (definitions of myopathy in Marshfield and SEARCH are described in the Supplementary Methods), and associations with plasma creatine kinase were tested in two statin trials that did not contain myopathy cases (CAP and JUPITER). Media accumulation of apolipoproteins was measured by enzyme-linked immunosorbent assay (ELISA) and gene expression was measured by quantitative PCR in hepatoma cell lines (HepG2 and HuH7) after GATM knockdown as achieved by 48 h transfection of Ambion Silence Select short interfering RNA (siRNA) or non-targeting control. See full Methods for complete details.

Full Methods and any associated references are available in the online version of the paper.

Received 11 April; accepted 26 July 2013.
Published online 28 August 2013.

1. Cholesterol Treatment Trialists’ (CTT) Collaborators. Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90,056 participants in 14 randomised trials of statins. Lancet 366, 1267–1278 (2005).

2. Simon, J. A. et al. Phenotypic predictors of response to simvastatin therapy among African-Americans and Caucasians: the Cholesterol and Pharmacogenetics (CAP) Study. Am. J. Cardiol. 97, 843–850 (2006).

3. Ridker, P. M. et al. Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein. N. Engl. J. Med. 359, 2195–2207 (2008).

4. Fernandez, G., Spatz, E. S., Jablonski, C. & Phillips, P. S. Statin myopathy: a common dilemma not reflected in clinical trials. Cleve. Clin. J. Med. 78, 393–403 (2011).

5. Rajpathak, S. N. et al. Statin therapy and risk of developing type 2 diabetes: a meta-analysis. Diabetes Care 32, 1924–1929 (2009).

6. Tesickov, T. M. et al. Biological, clinical and population relevance of 95 loci for blood lipids. Nature 466, 707–713 (2010).

7. Barber, M. J. et al. Genome-wide association of lipid-lowering response to statins in combined study populations. PLoS ONE 5, e9763 (2010).

8. Feskanich, D. et al. Pharmacogenetic study of statin therapy and cholesterol reduction. J. Am. Med. Assoc. 291, 2821–2827 (2004).

9. Trompet, S. et al. Replication of LDL GWAs hits in PROSPER/PHASE as validation for future (pharmaco)genetic analyses. BMC Med. Genet. 12, 131 (2011).

10. The SEARCH Collaborative Group. SLC1A1 variants and statin-induced myopathy—a genome-wide study. N. Engl. J. Med. 359, 789–799 (2006).

11. Brem, R. B., Storey, J. D., Whittle, J. & Kruglyak, L. Genetic interactions between polymorphisms that affect gene expression in yeast. Nature 436, 701–703 (2005).

12. Grundberg, E. et al. Global analysis of the impact of environmental perturbation on cis-regulation of gene expression. PLoS Genet. 7, e1001279 (2011).

13. Romaniszyn, C. E. et al. Systems genetics analysis of gene-by-environment interactions in human cells. Am. J. Hum. Genet. 86, 399–410 (2010).

14. Maranville, J. C. et al. Interactions between glucocorticoid treatment and regulatory polymorphisms contribute to cellular response phenotypes. PLoS Genet. 7, e1002152 (2011).

15. Smirnov, D. A. et al. Genetic variation in radiation-induced cell death. Genome Res. 332–339 (2011).

16. Brown, M. S. & Goldstein, J. L. The SREBP pathway: regulation of cholesterol metabolism by proteolysis of a membrane-bound transcription factor. Cell 89, 331–340 (1997).

17. Merley, M. et al. Genetic analysis of genome-wide variation in human gene expression. Nature 430, 743–747 (2004).

18. Stranger, B. E. et al. Population genomics of human gene expression. Nature Genet. 39, 1217–1224 (2007).

19. Cai, K., Muchmore, D. A., Ober, C. & Gilad, Y. The effects of EBV transformation on gene expression levels and methylations profiles. Hum. Mol. Genet. 20, 1643–1652 (2011).

20. Choy, E. et al. Genetic analysis of human traits in vitro: drug response and gene expression in lymphoblastoid cell lines. PLoS Genet. 4, e1000257 (2008).

21. Mangravite, L. M. et al. Combined influence of LDLR and HMGRQ sequence variation on lipid-lowering response to simvastatin. Arterioscler. Thromb. Vasc. Biol. 30, 1485–1492 (2010).

22. Medina, M. W. et al. Coordinately regulated alternative splicing of genes involved in cholesterol biosynthesis and uptake. PLoS ONE 6, e19420 (2011).

23. Stephens, M. & Balding, D. J. Bayesian statistical methods for genetic association studies. Nature Rev. Genet. 10, 681–690 (2009).

24. Leek, J. T. & Storey, J. D. Capturing heterogeneity in gene expression studies by surrogate variable analysis. PLoS Genet. 3, e161 (2007).

25. Innocenti, F. et al. Identification, replication, and functional fine-mapping of expression quantitative trait loci in primary human liver tissue. PLoS Genet. 7, e1002078 (2011).

26. Köttgen, A. et al. Multiple loci associated with indices of renal function and chronic kidney disease. Nature 419, 712–717 (2002).

27. Mareedu, R. K. et al. Use of an electronic medical record to characterize cases of severe and de novo statin-induced myopathy. Prev. Cardiol. 12, 32–38 (2009).

28. Choe, C. U. et al. L-arginine/glycine amidinotransferase deficiency protects from metabolic syndrome. Hum. Mol. Genet. 22, 110–123 (2013).

29. Ioe, T. et al. GAMT, a p53-inducible modulator of apoptosis, is critical for the adaptive response to nutrient stress. Mol. Cell 36, 379–392 (2009).

30. Edvardson, S. et al. L-arginine/glycine amidinotransferase (AGAT) deficiency: clinical presentation and response to treatment in two patients with a novel mutation. Mol. Genet. Metab. 101, 228–232 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements This project was funded by a grant from the US National Institutes of Health (NIH), U01 HL69757. B.E.E. was funded through the Bioinformatics Research Development Fund, supported by K. and G. Gould and NIH grant k99/R00 HG002655. M.S. was funded by NIH grant HG002855. We acknowledge the efforts of T. Kitchner and R. Mareedu for case validation in the Marshfield cohort. SEARCH was supported by the Medical Research Council, British Heart Foundation, National Health Service Genetic Knowledge Park, Centre National de Génotypage and Merck. The Heart Protection Study was funded by grants from the Medical Research Council, British Heart Foundation, Roche Vitamins and Merck. J.C.H. acknowledges support from the BHF Centre of Research Excellence, Oxford, UK. Genetic analysis in JUPITER was funded by a grant from AstraZeneca to D.I.C. and P.M.R.

Author Contributions L.M.M. designed experiment and analyses, generated samples, performed analyses, and wrote the manuscript. B.E.E. designed and performed analyses and wrote the manuscript. C.D.B. performed analyses of ENCODE data. B.H.M. designed and performed core analyses and coordinated the project. J.D.S., M.J.R. and D.A.N. performed association analyses and genotype data. M.W.M. and D.N. designed, performed and analysed functional experiments. B.H. and H.S. developed and performed the imputation methodology. R.A.W., Q.F., D.S., J.M.R. and D.A.N. collected and genotyped the myopathy cohort from the Marshfield clinic and performed association analyses. J.C.H., S.P., J.A. and R.C. collected and genotyped myopathy cohort from the SEARCH trial and performed association analyses in that cohort along with the Heart Protection Study. J.I.R. and Y.-D.I.C. measured creatine kinase in CAP, D.I.C. and P.M.R. measured creatine kinase and performed related analyses in JUPITER. M.S. supervised, designed and contributed to analyses, and participated in manuscript development. R.M.K. supervised the project and participated in experimental design and manuscript development. M.S. and R.M.K. co-directed this project.

Author Information The gene expression data have been deposited in the Gene Expression Omnibus (GEO) under accession number GSE36686 and in Synapse (https://www.synapse.org/#!Synapse:syn299510). The genotype data have been deposited in the database for genotypes and phenotypes (dbGaP, http://www.ncbi.nlm.nih.gov/gap) under accession number phs000481. The full set of eQTLs identified in our study is available at http://eqtl.uchicago.edu. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.M.M. (lara.mangravite@sagebase.org), M.S. (mstephens@uchicago.edu) or R.M.K. (rkrauss@chori.org).
METHODOLOGY

In vitro simvastatin exposure of lymphoblastoid cell lines. Lymphoblastoid cell lines (LCLs), immortalized by Epstein–Barr virus transformation of lymphocytes isolated from whole blood18, were derived from European American participants in the CAP trial, a 6-week simvastatin trial of 40 mg per day (Supplementary Table 8). Simvastatin was provided by Merck, converted to active form (beta-hydroxy simvastatin acid, SVA) and quantified by liquid chromatography–tandem mass spectrometry as described previously19. LCLs were normalized to a uniform cell density and exposed to 2 μM SVA (simvastatin-exposed) or control buffer (control-exposed) for 24 h as described previously18. This concentration was selected by assessing dose–response effects on expression profiles (n = 8 LCLs, 4 doses), wherein a more robust change in expression profiles was observed with 2 μM simvastatin exposure (7.8% of genes, q = 0.001) than lower doses (<0.1% of genes for 0.2 μM or 0.2 μM, q = 0.001, data not shown). Pre-experiment cell density was recorded as a surrogate for cell growth rate. After exposure, cells were lysed in RNAlater (Ambion), and RNA was isolated using the Qiagen miniprep RNA isolation kit with column DNase treatment.

Expression profiling and differential expression analysis. RNA quality and quantity were assessed using a Nanodrop ND-1000 spectrophotometer and an Agilent Bioanalyzer, respectively. Paired RNA samples, selected based on RNA quality and quantity, were amplified and labelled with biotin using the Illumina TotalPrep-96 RNA amplification kit, hybridized to Illumina HumanRef-8v3 bead-arrays (Illumina), and scanned using an Illumina BeadXpress reader. Data were read into GenomeStudio and samples were selected for inclusion based on quality-control criteria: signal to noise ratio (95th:5th percentiles); matched gender between sample and data; and average correlation of expression profiles within three standard deviations of the within-group mean (r = 0.99 ± 0.0093 for control-exposed and r = 0.98 ± 0.0071 for simvastatin-exposed beadarrays). In total, viable expression data were obtained from 1,040 beadarrays including 480 sets of paired samples for 10,195 genes. Genes were annotated through biomaRt from ensembl. Build 54 (http://may2009.archive.ensemble.org/martview). Treatment-specific effects were modelled from the data following adjustment for known covariates using linear regression20. False discovery rates were calculated for differentially expressed transcripts using the qvalue package21. Ontological enrichment in differentially expressed gene sets was measured using GSEA (1,000 permutations by phenotype) using gene sets representing Gene Ontology biological processes as described in the Molecular Signatures v5.0 C5 Database (10–500 genes per set)22.

Expression QTL mapping. For association mapping, we use a Bayesian approach23 implemented using the software package BIMBAM24 that is robust to poor imputation and small minor-allele frequencies25. Gene expression data were normalized as described in the Supplementary Methods for the control-treated (C480) and simvastatin-treated (T480) data and used to compute D480 = T480 − C480 and S480 = T480 + C480, where T480 represents the adjusted simvastatin-treated data and C480 represents the adjusted control-treated data. SNPs were imputed as described in the Supplementary Methods. To identify eQTLs and differential eQTLs, we measured the strength of association between each SNP and gene in each analysis (control-treated, simvastatin-treated, averaged, and difference) using BIMBAM with default parameters23. BIMBAM computes the Bayes factor for an additive or dominant response in expression data as compared with the null, which is that there is no correlation between that gene and that SNP. BIMBAM averages the Bayes factor over four plausible prior distributions on the effect sizes of additive and dominant models. We used a permutation analysis (see Supplementary Methods) to determine cutoffs for eQTLs in the averaged analysis (S480) at an FDR of 1% for cis-eQTLs (log_{10} Bayes factor > 3.24) and trans-eQTLs (log_{10} Bayes factor > 7.20). For cis-eQTLs, we considered the largest log_{10} Bayes factor above the cis-cutoff for any SNP within 1 Mb of the transcription start site or the transcription end site of the gene under consideration. For trans-eQTLs, we considered the largest log_{10} Bayes factor above the trans-cutoff for any SNP, and if that SNP was in the cis-neighbourhood of the gene being tested, we ignored any potential trans-associations; there were 6,130 genes for which the SNP with the largest log_{10} Bayes factor was not in cis with the associated gene. Correspondingly, we only considered those 6,130 genes when computing the permutation-based FDR for the trans-associations.

Differential expression QTL mapping. We define cis-SNPs as being within 1 Mb of the transcription start site or end site of that gene. To identify differential eQTLs, we first computed associations between all SNPs and the log fold change using BIMBAM as above. We then considered a larger set of models for differential eQTLs. The associations for the genes in Supplementary Fig. 3 indicate that there are a few possible patterns of differential association. Although these patterns may have a mechanistic or phenotypic interpretation, they are not distinguished by a test of log fold change. We used the interaction models introduced in another paper24 to compute the statistical support (assessed with Bayes factors) for the four alternative eQTL models described above versus the null model (no association with genotype).

These methods are based on a bivariate normal model for the treated data (T) and control-treated data (U). Note that simply quantile transforming T and U to a standard normal distribution is not sufficient to ensure that they are jointly bivariate normal, and so we used the following more extensive normalization procedure. Let D = qT − qU and S = qT + qU, where q indicates that the vector following it has been quantile normalized. We then quantile normalize and scale D and S to produce S = (σ_uqS) and D = (σ_qqD), where σ_u, σ_q are robust estimates of the standard deviations of S and D, respectively (specifically, they are the median absolute deviation multiplied by 1.4826). Note that this transformation ensures that S and D are univariate normal. Furthermore, they are independent, which ensures that they are also bivariate normal. Finally, let U = 0.5(S − D) and T = 0.5(S + D).

The Bayes factor when the eQTL effect is identical in the two conditions (model 1) uses the linear model L(S − D + g), where g is the vector of genotypes at a single SNP. The Bayes factor when the eQTL is only present in the control-treated samples (model 2) uses the model L(U − T + g). The Bayes factor when the eQTL is only present in the simvastatin-treated samples (model 3) uses the model L(T − U + g). The Bayes factor when the eQTL effect is in the same direction but unequal in strength (model 4) uses the model L(D − S + g). We averaged each Bayes factor for each gene and each cis-SNP over four plausible effect size priors (0.05, 0.1, 0.2 and 0.4).

Association with statin-induced myopathy. For the Marshfield Cohort27, cases of myopathy were identified from electronic medical records of patients treated at the Marshfield Clinic (Wisconsin, USA) using a combination of automated natural language processing and manual review as described previously27. Seventy-two cases of incipient myopathy (creatine kinase concentrations greater than 3-fold normal concentrations, with evidence in the charts of muscle complaints) were identified for which patients were not also undergoing treatment with concomitant drugs known to increase incidence of statin-induced myopathy (fibrate or niacin). Controls were matched based on statin exposure, age and gender. This study was approved by the Marshfield Clinic institutional review board. The study population included residents living in Central and Northern Wisconsin, served by the Marshfield Clinic, a large multi-specialty group practice28. For the SEARCH and Heart Protection Study Collaborative Groups29, a total of 100 myopathy cases were identified from participants with genotyping data in the SEARCH trial, including 39 definite myopathy cases (creatine kinase > 10 × upper limit of normal (ULN) with muscle symptoms) and 61 incipient myopathy cases (defined as creatine kinase > 5 times baseline value and alanine transaminase ≥ 1.7 times baseline value and creatine kinase > 3×ULN). Genotypes were available from the Illumina Human610-Quad Beadchip for 25 myopathy cases (12% of which had definite myopathy) and from the Illumina HumanHap300-Duo BeadChip for 75 myopathy cases (48% of which had definite myopathy). Genotypes for rs9806699 were only available for the 25 cases genotyped on the Illumina Human610-Quad Beadchip, so proxy SNPs were used for analyses in Table 1. Analyses of rs9806699 are provided in Supplementary Table 9. All myopathy cases were compliant with statin therapy (95 myopathy cases occurred while the patient was taking simvas- tatin 80 mg daily, and 5 cases while taking simvasatin 20 mg daily). Controls were identified using the SEARCH Study as well as from the Heart Protection Study (where considerably more participants had been genotyped). Controls from the Heart Protection Study had similar baseline characteristics to those in the SEARCH Study and inclusion of this large number of additional controls improved statistical power. Multi-centre ethics approval was obtained from the South East Research Ethics Committee for the SEARCH study, and from the local ethics committees covering each of the 69 UK hospitals involved in the Heart Protection Study. Genetic associations were determined by chi-squared analysis using an additive model. A meta-analysis was performed using a random effects model and, for the Bayesian analysis, we used an expected effect size of 0.2. Associations of rs9806699 with plasma creatine kinase in the CAP30 and JUPITER31 trials were also assessed using linear regression. The CAP trial (ClinicalTrials.gov number NCT00451828) was approved by the institutional review boards located at Children’s Hospital Oakland Research Institute (Oakland, California) and all enrollment sites. The JUPITER trial (ClinicalTrials.gov number NCT00239681) was approved by the Institutional Review Board of Brigham and Women’s Hospital. Informed consent was obtained from all participants in all trials.

©2013 Macmillan Publishers Limited. All rights reserved
Genomic characterization of GATM locus. Cis-regulatory elements were downloaded from the ChromHMM (ref. 39) track of the UCSC Genome Browser (ref. 40) and aggregated manually.

**Functional analysis of candidate genes.** GATM knockdown was achieved by 48 h transfection of Ambion Silence Select siRNA or non-targeting control into 80,000 HepG2 or Huh7 cells per well in 12-well plates. To assess the influence of sterol depletion, cell culture medium was replaced with medium containing 10% lipo-protein deficient serum (Hyclone) or fetal bovine serum (Omega Scientific) at 24 h after transfection. All samples were harvested 48 h post transfection. Transcript levels were quantified by quantitative PCR and normalized to 3LPTM. Cell culture medium was taken from all samples at the time of collection, and ApoB (MP Biomedicals) and ApoAI (Meridian Life Sciences) were quantified in triplicate by sandwich-style ELISA. Samples with a coefficient of variation greater than 15% were subjected to repeat measurement.

31. Pressman, S. & Rotter, J. I. Epstein–Barr virus transformation of cryopreserved lymphocytes: prolonged experience with technique. Am. J. Hum. Genet. 49, 467 (1991).

32. Mecham, B. H., Nelson, P. S. & Storey, J. D. Supervised normalization of microarrays. Bioinformatics 26, 1308–1315 (2010).

33. Storey, J. D. A direct approach to false discovery rates. J. R. Stat. Soc. 64, 479–498 (2002).

34. Subramanian, A., Kuehn, H., Gould, J., Tamayo, P. & Mesirov, J. P. GSEA-P: a desktop application for Gene Set Enrichment Analysis. Bioinformatics 23, 3251–3253 (2007).

35. Servin, B. & Stephens, M. Imputation-based analysis of association studies: candidate regions and quantitative traits. PLoS Genet. 3, e114 (2007).

36. Guan, Y. & Stephens, M. Practical issues in imputation-based association mapping. PLoS Genet. 4, e1000279 (2008).

37. Wilke, R. A. et al. Identifying genetic risk factors for serious adverse drug reactions: current progress and challenges. Nature Rev. Drug Discov. 6, 904–916 (2007).

38. Hopewell, J. C. et al. Impact of common genetic variation on response to simvastatin therapy among 18 705 participants in the Heart Protection Study. Eur. Heart J. 34, 982–992 (2013).

39. Ernst, J. & Kellis, M. ChromHMM: automatting chromatin-state discovery and characterization. Nature Methods 9, 215–216 (2012).

40. Kent, W. J. et al. The human genome browser at UCSC. Genome Res. 12, 996–1006 (2002).