Gastroenterology 2016;151:351–363

BASIC AND TRANSLATIONAL—BILIARY

Four Susceptibility Loci for Gallstone Disease Identified in a Meta-analysis of Genome-Wide Association Studies

Amit D. Joshi,[1,2,3,*] **Charlotte Andersson**,[4,*] **Stephan Buch**,[5,*] **Stefan Stender**,[6,*] **Raymond Noordam**,[7,8,*] **Lu-Chen Weng**,[9,*] **Peter E. Weeke**,[10,11,*] Paul L. Auer,[12,13] Bernhard Boehm,[14] Constance Chen,[1] Hyon Choi,[15] Gary Curhan,[16,17] Joshua C. Denny,[10,18] Immaculata De Vivo,[1,16,19] John D. Eicher,[4,20] David Ellinghaus,[21] Aaron R. Folsom,[9] Charles Fuchs,[16,22] Manish Gala,[2] Jeffrey Haessler,[13] Albert Hofman,[8] Frank Hu,[19,23] David J. Hunter,[1,19] Harry L. A. Janssen,[24,25] Jae H. Kang,[16] Charles Kooperberg,[13] Peter Kraft,[1,19] Wolfgang Kratzer,[14] Wolfgang Lieb,[26] Pamela L. Lutsey,[9] Sarwa Darwish Murad,[24] Børge G. Nordestgaard,[27,28,29] Louis R. Pasquale,[16,30] Alex P. Reiner,[13] Paul M. Ridker,[31] Eric Rimm,[16,19,23] Lynda M. Rose,[31] Christian M. Shaffer,[10] Clemens Schafmayer,[32] Rulla M. Tamimi,[16,19] André G. Uitterlinden,[7,8] Uwe Völker,[33] Henry Völzke,[34,35,36] Yoshiyuki Wakabayashi,[37] Janey L. Wiggs,[30] Jun Zhu,[37] Dan M. Roden,[10,§] Bruno H. Stricker,[7,8,§] Weihong Tang,[9,§] Alexander Teumer,[34,§] Jochen Hampe,[5,§] Anne Tybjærg-Hansen,[6,28,§] Daniel I. Chasman,[31,§] Andrew T. Chan,[2,3,16,§] and Andrew D. Johnson[4,20,§]

[1]Program in Genetic Epidemiology and Statistical Genetics, [19]Department of Epidemiology, [23]Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts; [2]Division of Gastroenterology, [3]Clinical and Translational Epidemiology Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts; [4]The National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, Massachusetts; [5]Medical Department 1, University Hospital Dresden, Technische Universität Dresden, Dresden, Germany; [6]Department of Clinical Biochemistry, [11]Department of Cardiology, The Heart Centre, Rigshospitalet, Copenhagen, Denmark; [7]Department of Internal Medicine, [8]Department of Epidemiology, [24]Department of Gastroenterology and Hepatology, Erasmus Medical Center, Rotterdam, The Netherlands; [9]Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minnesota; [10]Department of Medicine, [18]Department of Biomedical Informatics, Vanderbilt University, Nashville, Tennessee; [12]Joseph J. Zilber School of Public Health, University of Wisconsin, Milwaukee, Wisconsin; [13]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington; [14]Department of Internal Medicine I, Ulm University Hospital, Ulm, Germany; [15]Division of Rheumatology, Allergy, and Immunology, Massachusetts General Hospital, Boston, Massachusetts; [16]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts; [17]Renal Division, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts; [20]Population Sciences Branch, National Heart, Lung, and Blood Institute, National Institutes of Health, Framingham, Massachusetts; [21]Institute of Clinical Molecular Biology, [26]Institute of Epidemiology, Christian-Albrechts-University of Kiel, Kiel, Germany; [22]Department of Medical Oncology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, Massachusetts; [25]Toronto Centre for Liver Disease, Toronto Western and General Hospital, University Health Network, Toronto, Canada; [27]The Copenhagen General Population Study, [28]Department of Clinical Biochemistry, Herlev Hospital, Herlev, Denmark; [29]Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; [30]Department of Ophthalmology, Harvard Medical School, Massachusetts Eye and Ear Infirmary, Boston, Massachusetts; [31]Division of Preventive Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts; [32]Department of General, Abdominal, Thoracic and Transplantation Surgery, University of Kiel, Kiel, Germany; [33]Department of Functional Genomics, Interfaculty Institute for Genetics and Functional Genomics, [34]Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany; [35]German Center for Cardiovascular Research, Greifswald, Germany; [36]German Center for Diabetes Research, Greifswald, Germany; [37]National Heart, Lung, and Blood Institute, DNA Sequencing Core Laboratory, Bethesda, Maryland

BACKGROUND & AIMS: A genome-wide association study (GWAS) of 280 cases identified the hepatic cholesterol transporter *ABCG8* as a locus associated with risk for gallstone disease, but findings have not been reported from any other GWAS of this phenotype. We performed a large-scale, meta-analysis of GWASs of individuals of European ancestry with available prior genotype data, to identify additional genetic risk factors for gallstone disease. **METHODS:** We obtained per-allele odds ratio (OR) and standard error estimates using age- and sex-adjusted logistic regression models within each of the 10 discovery studies (8720 cases and 55,152 controls). We performed an inverse variance weighted, fixed-effects meta-analysis of study-specific estimates to identify single-nucleotide polymorphisms that were associated independently with gallstone disease. Associations were replicated in 6489 cases and 62,797 controls. **RESULTS:** We observed independent associations for 2 single-nucleotide polymorphisms at the *ABCG8* locus: rs11887534 (OR, 1.69; 95% confidence interval [CI], 1.54–1.86; $P = 2.44 \times 10^{-60}$) and rs4245791 (OR, 1.27; $P = 1.90 \times 10^{-34}$). We also identified and/or replicated associations for

rs9843304 in *TM4SF4* (OR, 1.12; 95% CI, 1.08–1.16; $P = 6.09 \times 10^{-11}$), rs2547231 in *SULT2A1* (encodes a sulfoconjugation enzyme that acts on hydroxysteroids and cholesterol-derived sterol bile acids) (OR, 1.17; 95% CI, 1.12–1.21; $P = 2.24 \times 10^{-10}$), rs1260326 in glucokinase regulatory protein (OR, 1.12; 95% CI, 1.07–1.17; $P = 2.55 \times 10^{-10}$), and rs6471717 near *CYP7A1* (encodes an enzyme that catalyzes conversion of cholesterol to primary bile acids) (OR, 1.11; 95% CI, 1.08–1.15; $P = 8.84 \times 10^{-9}$). Among individuals of African American and Hispanic American ancestry, rs11887534 and rs4245791 were associated positively with gallstone disease risk, whereas the association for the rs1260326 variant was inverse. **CONCLUSIONS:** In this large-scale GWAS of gallstone disease, we identified 4 loci in genes that have putative functions in cholesterol metabolism and transport, and sulfonylation of bile acids or hydroxysteroids.

Keywords: Genetics; Risk Factors; SNP; GWAS.

Accounting for a substantial clinical burden in the United States, gallstone disease afflicts 6.3 million men and 14.2 million women between the ages of 20 and 74 years, leading annually to 700,000 cholecystectomies and an economic burden of 6.5 billion dollars.[1] It was hypothesized as early as the 1960s that the composition of bile may play an important role in gallstone formation.[2] Bile is formed by the transportation of cholesterol, bile acids, and other organic molecules such as bilirubin from within the hepatocytes to the biliary canaliculi, and serves as a medium for excretion of lipid-soluble products of metabolism. Precipitation of biliary constituents from their soluble state into their insoluble form initiates the process of gallstone formation. Clinical conditions with chronic hemolytic states such as sickle cell disease frequently have been associated with pigmented gallstones[3] as a result of the increased delivery of unconjugated bilirubin into the bile via hepatocytes.[4] However, the most common (80%–90%) constituent of gallstones retrieved during cholecystectomy surgery or autopsy is biliary cholesterol. Studies that compared the constituents of lithogenic bile and normal bile observed that higher concentrations of cholesterol, or the alterations in relative proportions of other bile components such as bile salts and phospholipids, can result in supersaturation of cholesterol.[2,5] Redinger and Small[6] further showed a correlation between the percentage saturation of biliary cholesterol in various ethnic groups and estimated gallstone prevalence rates in the same population in an ecologic study. Consequently, several lifestyle determinants such as female sex, greater parity, postmenopausal hormone therapy, Native American ancestry, high body mass index (BMI), and dyslipidemia are among the most important risk factors for gallstone disease, primarily because of their influence on cholesterol concentration in the bile.[5,7]

Based on familial clustering of gallstone disease, a 2- to 3-fold increased risk among first-degree relatives,[8–10] and heritability estimates of 25%–29% from twin studies,[10,11] it has been suggested that genetic factors may play an important contributory role in cholelithiasis. More evidence to support this hypothesis was established using experimental crosses of inbred mice strains with varying prevalence of gallstones.[12,13] Quantitative trait loci-based approaches were used to generate a murine gallstone genetic map of several candidate lithogenic (*lith*) loci,[12,14] with the idea that orthologous human *LITH* genes may be predicted owing to homology between human and mouse genomes. These murine *lith* loci co-localized with approximately 7 "likely," and approximately 20 "plausible" candidate genes for gallstone disease, many of which are involved in cholesterol (eg, *ABCG5/ABCG8*) and bile acid (eg, *ABCB11*) synthesis, transport, or metabolism.[13]

The identification of genetic risk factors of gallstone disease in human beings was undertaken in 2007 in a discovery-based genome-wide association study (GWAS) of 280 cases and 360 controls.[15] This study identified and replicated an approximately 2-fold increased risk for carriers of the H-allele of D19H in the hepatic cholesterol transporter gene *ABCG8* (rs11887534; risk allele frequency, ~7%).[15,16] Other studies that examined genetic associations with gallstone disease were based on biological insights of candidate loci or pathways. Buch et al[17] investigated the association of known bilirubin loci[18] with the incidence of gallstone disease, and observed a recessive mode of inheritance at the *UGT1A1* SNP locus rs6742078, finding that carriers of the T/T genotype were predisposed to an increased risk of gallstone disease among men, but not among women.[17] Moreover, a recent study in women, examining associations of approximately 2000 gene-centric loci in known lipid metabolism and obesity pathways,[19] reported additional associations for the glucokinase regulatory protein (*GCKR*) SNP rs1260326 and the *TTC39B* SNP rs686030 with gallstone disease; however, these associations were not replicated.

Although there is strong evidence for genetic contribution toward the risk of gallstone disease, there are few replicated susceptibility loci identified from genome-wide, discovery-based approaches because of the limited size and scope of prior studies. In this study, we therefore conducted a large-scale GWAS meta-analysis in individuals with pre-existing genetic data on more than 2 million genetic variants, to discover additional loci associated with the risk

*Authors share co-first authorship; § Authors share co-senior authorship.

Abbreviations used in this paper: **ARIC, Atherosclerosis Risk in Communities Study; BioVU, Vanderbilt DNA Biobank; BMI, body mass index; CI, confidence intervals; eSNP, expression single-nucleotide polymorphism; eQTL, expression quantitative trait loci; FHS, Framingham Heart Study; GCKR, glucokinase regulatory protein; GCTA, genome-wide complex trait analysis; GRS, genetic risk score; GWAS, genome-wide association studies; HPFS, Health Professionals Follow-up Study; ICD, International Classification of Diseases; lith, lithogenic; MAF, minor allele frequency; NHS, Nurses' Health Study; OR, odds ratio; RPKM, reads per kilobase per million; SHIP, Study of Health in Pomerania; SNP, single-nucleotide polymorphism; WGHS, Women's Genome Health Study; WHI, Women's Health Initiative.**

Most current article

0016-5085/$36.00
http://dx.doi.org/10.1053/j.gastro.2016.04.007

of gallstone disease in individuals of European ancestry. We replicated the SNPs within each of the newly discovered loci in independent samples, and queried transcriptomic and metabolomic databases to derive clues about potential causal variants near the SNPs with highest evidence for association with gallstone disease.

Materials and Methods

Study Participants

The study population for the discovery set consisted of individuals with extant genome-wide genotyping data available from previous studies, among whom we identified 8720 cases and 55,152 controls within the following 10 cohorts: the Study of Health in Pomerania (SHIP) and SHIP-TREND,[20] the Nurses' Health Study (NHS) I and II,[21] the Health Professionals Follow-up Study (HPFS), Women's Genome Health Study,[22] Atherosclerosis Risk in Communities Study (ARIC),[23] the Framingham Heart Study (FHS) original and offspring cohorts,[24] the Rotterdam study,[25,26] community-based cases and controls from the PopGen biobank,[27,28] and a case-control cohort from the Vanderbilt DNA Biobank[29] (Table 1). The validation set comprised an additional 6489 cases and 62,797 controls from the Copenhagen General Population Study and the Copenhagen City Heart Study, the Kiel Study (Germany), and from a subset of the samples from the NHS1/NHSII and HPFS that did not overlap with the discovery set (Table 1). Details of study population, genotyping, quality control, and imputation in each study are described in detail in the Supplementary Materials and Methods section and in Supplementary Figure 1. The definition and assessment of gallstone disease in each cohort is detailed in Supplementary Table 1. Briefly, gallstone disease cases were defined either by self-report in a questionnaire asking directly about gallstone disease or prior cholestectomy (Women's Genome Health Study, NHS, HPFS, FHS, ARIC, FHS, Women's Health Initiative [WHI]) or International Classification of Diseases (ICD) codes (Rotterdam study, Vanderbilt DNA Biobank, Copenhagen General Study Population, and Copenhagen City Heart Study), or abdominal ultrasonography (SHIP, SHIP-TREND, PopGen, and the study from Kiel).

Statistical Analysis

Within each discovery study, we estimated the association between genotyped or imputed single-nucleotide polymorphisms (SNPs) and the risk of gallstone disease by calculating β coefficients and their standard errors using logistic regression models adjusted for age, sex, and additional study-specific covariates, assuming log-additive genetic effects. Before meta-analyses, we excluded imputed SNPs with imputation quality score and/or imputation $R^2 < 0.3$. We also used a minor allele frequency (MAF) filter, excluding SNPs with a MAF less than 0.01 for cohorts with more than 500 cases. For cohorts with fewer than 500 cases, we used a more stringent MAF threshold of 5, divided by the number of cases, thereby limiting analysis to SNPs expected to have 10 or more minor alleles within cases, to obtain robust estimates. Inverse variance weighted, fixed-effects, meta-analysis[30] of study-specific estimates was performed to identify SNPs associated with gallstone disease, using METAL (http://genome.sph.umich.edu/wiki/METAL_documentation). We selected the strongest independent markers at each locus to attempt replication as well as to aid in functional/molecular interpretation, by performing conditional analyses in genomic regions (10 megabase windows using a less-stringent nominal significance threshold for SNPs [discovery $P < 5 \times 10^{-6}$]), using genome-wide complex trait analysis software[31] (http://www.complextraitgenomics.com/software/gcta/). Conditional analysis is a mechanism to try to reduce the number of significant associations to the top most independent associations. We used 1753 healthy controls of European ancestry from the type 2 diabetes data set within the NHS as the reference population. Replication was performed for SNPs that were observed to be associated with gallstone disease risk at a genome-wide significance threshold of P less than 5×10^{-8} after conditional analysis. We genotyped newly identified SNPs using the TaqMan (Applied Biosystems Inc, Foster City, CA) or KASPar (KBiosciences, Hoddesdon, Hertfordshire, UK) assay in the replication data sets, except the NHS and HPFS studies, in which we had pre-existing genotype/imputation data. We reported fixed-effects, meta-analytic odds ratios (ORs) and 95% confidence intervals (CIs) for combined associations from discovery and replication studies for all of the replicated SNPs. The heterogeneity of effect sizes between studies was determined using Cochran's Q-test for heterogeneity[32] as implemented in METAL,[30] and also by determining the I^2 statistics[33] that compute the proportion of overall variance that can be attributed as a result of differences in effect sizes between studies. For these SNPs, if discovery studies showed an evidence of heterogeneity ($P < .05$), we reported association results using random-effects meta-analysis in the combined discovery and replication studies.

In the replication studies, we additionally determined the strength of association for unit of SD increase in the weighted genetic risk score with gallstone disease risk. For the purpose of developing a genetic risk score, SNPs with missing information within the replication data sets were imputed by random sampling with replacement, from individuals with the SNP information available, and conditional on case-control status. We derived a genetic risk score for each study participant by assigning weights to each risk allele proportional to the logarithm of per-allele relative risk estimate in the meta-analysis of discovery studies. The weighted genetic risk score (GRS) was standardized to have a zero mean and unit SD.

We performed a sensitivity analysis to exclude possible genetic associations mediated by BMI. Logistic regression models in each of the discovery studies were used to obtain β coefficients and standard errors, after adjusting for BMI in addition to age and sex, followed by a meta-analysis of study-specific effect size estimates.

Post Hoc Analysis

We performed ancestry-specific analyses to determine whether any of the variants with P less than 5×10^{-8} in the discovery and replication data sets showed an association in African American or Hispanic American individuals, and whether they showed differences in allelic frequencies across populations. Analysis was performed in individuals of African American ancestry for 115 prevalent gallstone disease cases and 2484 controls in the ARIC cohort, and for 1384 incident and prevalent cases and 6661 controls in the WHI cohort. Effect size estimates for Hispanic American ethnicity was performed

Table 1. Characteristics of Gallstone Disease GWAS Meta-Analysis Discovery and Replication Studies

Discovery Study	Study design	Cases	Controls	Female, %		Age, *y*, mean ± SD		Genotyping platform	Imputation platform
				Cases	Controls	Cases	Controls		
WGHS	Nested case-control	2853	20,436	100.0	100.0	55.6 ± 11.3	64.0 ± 7.1	Illumina Duo	HapMap
NHS1/2/HPFS Affymetrix	Nested case-control	1562	6211	72.2	53.2	60.5 ± 7.9	60.3 ± 8.1	Affymetrix SNP 6.0,	1000G
NHS1/2/HPFS Illumina		1019	4400	85.5	75.7	57.4 ± 8.4	56.3 ± 9.0	Illumina 550K, 660K	
SHIP	Nested case-control	843	3134	65.6	47.1	60.3 ± 13.2	46.6 ± 15.8	Affymetrix SNP 6.0	1000G
ARIC	Case-control (prevalent)	832	8032	76.3	51.1	55.0 ± 5.7	54.1± 5.7	Affymetrix 6.0	HapMap
Rotterdam	Nested case-control	705	5269	73.0	54.2	71.0 ± 8.8	68.7 ± 9.1	Illumina 550K	HapMap
FHS	Nested case-control	515	3783	71.3	53.2	67.2 ± 9.0	62.9 ± 9.6	Affymetrix 550K	HapMap
BioVU	Hospital-based case-control	202	2542	58.4	50.4	64.6 ± 16.1	62.4 ± 16.3	Human660W-Quad BeadChip	1000G
SPC (PopGen)	Nested case-control	122	527	59.0	43.2	57.9 ± 12.7	62.5 ± 8.4	Affymetrix 6.0	1000G
SHIP-TREND	Nested case-control	67	818	64.2	53.6	56.6 ± 12.9	48.4 ± 13.4	Illumina Omni 2.5	1000G
All discovery studies		**8720**	**55,152**						
Replication studies									
CCHS and CGPS	Prospective cohort study	3599	57,389	70.6	54.1	61.1 ± 13.0	56.8 ± 13.9	TaqMan/KASPar genotyping	
Kiel University	Hospital-based case-control	2104	2225	70.6	51.7	52.9 ± 11.2	39.7 ± 14.9	TaqMan genotyping	
NHS1/HPFS replication	Nested case-control	786	3183	82.7	69.90	60.6 ± 7.4	59.5 ± 7.8	Illumina OmniExpress	1000G
All replication studies		**6489**	**62,797**						
Combined discovery + replication		**15,209**	**117,949**						
Replication in non-European ancestry individuals									
WHI (African American)	Nested case-control	1384	6661	100.0	100.0	61.8 ± 6.9	61.5 ± 7.0	Affymetrix 6.0	1000G
ARIC (African American)	Case-control (prevalent)	115	2484					Affymetrix 6.0	HapMap
WHI (Hispanic American)	Nested case-control	1056	2403	100.0	100.0	60.9 ± 6.6	59.9 ± 6.7	Affymetrix 6.0	1000G

NOTE. Bolded entries provide the total number of cases and controls for discovery studies, replication studies and both combined.
Illumina Duo, 550K, 660K, Human660W-Quad BeadChip, Omni 2.5, OmniExpress (Illumina Inc, San Diego, CA), Affymetrix SNP 6.0 (Affymetrix Inc, Santa Clara, CA), TaqMan (Applied Biosystems Inc, Foster City, CA), KASPar (KBiosciences, Hoddesdon, Hertfordshire, UK)
BioVU, Vanderbilt DNA Biobank; CCHS, Cophenhagen City Heart Study; CGPS, Copenhagen General Study Population; WGHS, Women's Genome Health Study.

in 1056 cases of incident or prevalent gallbladder disease and 2403 controls within the WHI.

From the discovery GWAS meta-analyses summary statistics we determined the following associations: (1) known nonalcoholic fatty liver disease variants, (2) previously reported variants associated with gallstone disease that did not reach genome-wide significance in our data sets (*UGT1A1* rs6742078 and *TTC39B* SNP rs686030), and (3) overlap with *lith* genes described from murine models.[12–14]

In post hoc analysis within the NHS and HPFS cohorts, for SNPs with P less than 5×10^{-8}, we computed genotype-specific associations with gallstone disease, and percentage population attributable risk for each genotype, as described previously.[17] In addition, we tested for associations for these SNPs assuming different modes of inheritance (recessive and dominance effects), and for gene–gene interactions between these SNPs. For multiple independent associations at the same genetic locus (*ABCG8* SNPs), we tested for associations of each haplotype combination with gallstone disease risk. We also evaluated for confounding effects of history of self reported hypercholesterolemia, use of cholesterol-lowering drugs (ever/ never), and postmenopausal hormone use (ever/ never).

RNA Sequencing of Human Gallbladder

We performed RNA sequencing from 4 human gallbladders (3 healthy controls and 1 patient with chronic gallstones) and 1 liver sample from the gallstone patient. RNA was obtained from gallbladder and liver of 1 woman, age 71 years, with chronic cholecystitis and metastatic adenocarcinoma consistent with primary colon cancer (OriGene, Technologies, Inc, Rockville, MD CU0000000466). RNA also was obtained from 3 normal gallbladder samples, all women (ages 34, 46, and 64 years) (BioChain Institute, Inc, Hayward, CA, lot numbers A509245, A607331 and A509248 respectively).

RNA Seq libraries were prepared using Ovation RNAseq v2 (NuGEN Technologies, Inc, San Carlos, CA), following the guidelines for the Ovation SP Ultralow DR Multiplex System (NuGEN Technologies, Inc, San Carlos, CA). Library quality was verified for each sample using MiSeq (Illumina, Inc, San Diego, CA), sequencing with 75-bp paired-end reads. Samples next were sequenced using an Illumina HiSeq 2000 instrument (Illumina, Inc) with 75-bp paired-end reads. The raw reads in fastq format were mapped to human genome hg19 using Tophat (v2.0.9) with the following parameter setting: -g 1, -N 2, -r 200. RefSeq transcripts read count and reads per kilobase of transcript per million were calculated using RSeQC (v2.3.6). The runs generated an average of 4,063,889 uniquely mapped reads per sample, with good mapping rates: cholecystitis gallbladder (89.5% uniquely mapped), cholecystitis liver (83.8%), and normal gallbladder samples (96.0%, 96.1%, and 84.9%, respectively). These data are available through GEO accession number GSE66430 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE66430).

Expression Quantitative Trait Loci and Encyclopedia of DNA Elements (ENCODE) Regulatory Analyses

Proxy SNPs in linkage disequilibrium ($r^2 > 0.8$) in populations of European ancestry were identified for gallstone index and replication SNPs using SNAP (http://www.broadinstitute.org/mpg/snap/).[34] Index SNPs and proxies were queried against a collected database of expression SNP (eSNP) results. The collected eSNP results met the criteria for statistical thresholds for association with gene transcript levels as described in the original article.[35] A general overview of a subset of more than 50 expression quantitative trait loci (eQTL) studies has been published.[35] We assessed the concordance of the gallstone-identified eSNPs with the strongest eSNPs for each individual gene and data set using linkage disequilibrium metrics (R^2) and report the results for either the index SNP or SNPs in linkage disequilibrium with R^2 greater than 0.8. The resulting eQTL SNPs with gene expression associations with P less than 5×10^{-6} were queried for overlap with ENCODE regulatory features using HaploReg v3 (http://www.broadinstitute.org/mammals/haploreg/haploreg_v3.php).[36] More details on eQTL and ENCODE regulatory analyses methods are available in the Supplementary Materials and Methods section.

Prior GWAS Phenotype Analysis

Gallstone index and replication SNPs and their proxies (as defined earlier) were queried against the National Heart, Lung, and Blood Institute Genome-wide Repository of Associations between SNPs and Phenotypes, version 2.0.0.0 (http://apps.nhlbi.nih.gov/grasp). Only results with a P value less than 5×10^{-8} were retained. The strongest SNP per GWAS phenotype per gallstone locus was reported.

Results

Meta-analysis

Fixed-effects meta-analysis,[30] followed by conditional analyses within nominally significant regions[31] (10 Mb windows around SNPs with $P < 5 \times 10^{-6}$), yielded 7 SNPs from 5 genome-wide significant regions: *ABCG5/8*, *TM4SF4*, *SULT2A1*, *UBXN2B/CYP7A1*, and *GCKR*, independently associated with gallstone disease ($P < 5 \times 10^{-8}$) (Table 2, Figure 1, and Supplementary Table 2). There was no evidence of inflation of test statistics in the genome-wide meta-analysis ($\lambda = 1.037$; Q-Q plot in Supplementary Figure 2). The newly discovered SNPs had high imputation quality scores (>0.80) in each of the discovery studies (Supplementary Table 3). A sensitivity analysis adjusting for BMI before meta-analyses (to exclude genetic associations potentially mediated by BMI) yielded results that did not differ materially from those presented in Table 2 (Supplementary Table 4). Regional association plots for the 5 independent loci are shown in Supplementary Figure 3. Except for the *ABCG5* and *ABCG8* loci, SNPs with a P value of less than 1×10^{-4} in our discovery samples did not map to human orthologs of the candidate *lith* genes proposed in murine models. Although we did not observe a genome-wide significance for previously reported *TTC39B* SNP rs686030,[19] the A allele at the locus showed some evidence for an increased risk of gallstone disease (OR, 1.09; $P = .000438$).

Replication

We selected 6 SNPs (rs11887534 and rs4245791 [*ABCG8*], rs6471717 [*CYP7A1*], rs9843304 [*TM4SF4*], rs2547231 [*SULT2A1*], and rs1260326 [*GCKR*]) for replication (Table 2) and for subsequent functional analyses

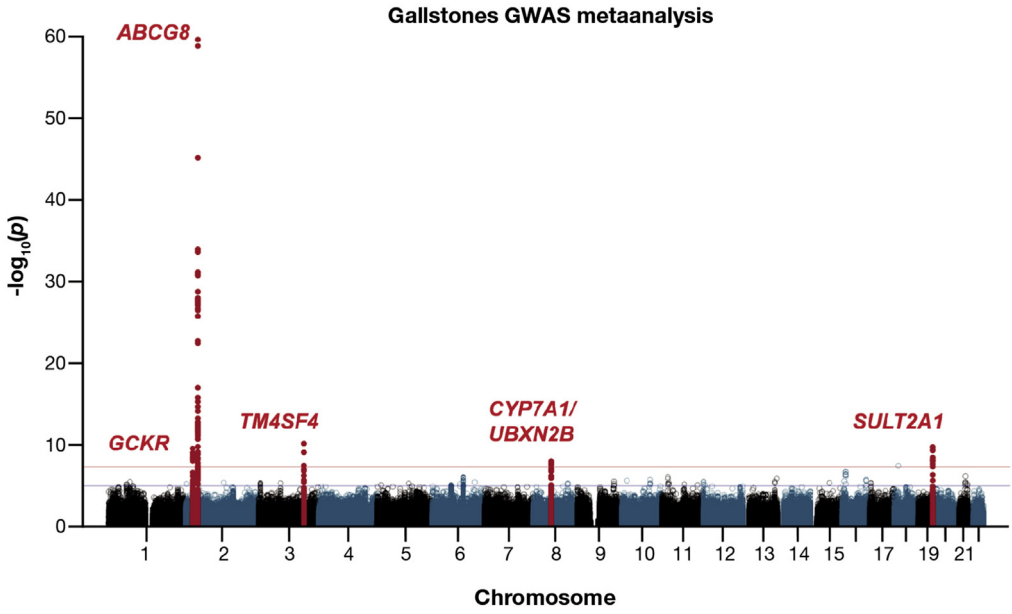


Figure 1. Manhattan plot of the results of genome-wide meta-analysis of gallstone disease in 10 studies. The plot shows $-\log_{10}$-transformed *P* values for all SNPs. The *red horizontal line* represents a *P* value of 5×10^{-8}. The *blue horizontal line* represents a *P* value of 1×10^{-5}.

(Table 3). For replication, we utlized an independent sample of 6489 cases and 62,797 controls from 3 population-based studies and a case-control study (Table 1). The *ABCG8* SNP rs4245791 (*P*-discovery = 1.90×10^{-34}, $R^2 = 1.0$ with rs4299376), and *SULT2A1* SNP rs2547231 (*P*-discovery = 2.24×10^{-10}, $R^2 = 0.90$ with rs296391), have been shown previously to be associated strongly with hepatic ABCG8[37] and SULT2A1[38] expressions, respectively, and therefore were selected for replication instead of the index SNPs. All of the selected SNPs were associated significantly with gallstone disease in meta-analysis from replication data sets. To account for heterogeneity of effect estimates for the *ABCG8* locus SNP rs11887534 and for the *UBXN2B/CYP7A1* SNP rs6471717 in the discovery meta-analysis, we report their effect sizes using both fixed- and random-effects meta-analysis in the combined discovery and replication analyses (Table 2 and Figure 2). The fixed- and random-effects *P* value for rs6471717 in combined discovery and replication analyses were 1.41×10^{-13} and 1.59×10^{-7}, respectively. It is likely that evidence of heterogeneity reflects differences in magnitude of effect sizes of the susceptibility locus, possibly owing to differences in study design or participant characteristics. However, the direction of effect was consistent for all replication SNPs across the studies (Figure 2). GRSs based on the 6 replicated SNPs and weighted on discovery stage β estimates were associated with an approximately 35% increased risk of gallstone disease for unit of SD increase in GRS in all replication studies and provided modest improvement in the area under the receiver operator characteristic curve (Supplementary Table 5 and Supplementary Figure 4).

SNP Associations in African American and Hispanic American Populations

We observed that 3 SNPs from 2 loci—rs1260326, rs11887534, and rs4245791—were associated significantly ($P < .05$) with gallstone disease among African American and Hispanic American individuals (Table 4). However, the direction of association was opposite to what we observed in the European population for rs1260326. We did not observe an association in these ethnicities for rs9843304, rs6471717, or rs2547231 SNPs. Moreover, we also observed marked differences in allele frequencies, for example, the T allele at rs1260326 is the major allele in individuals of European ancestry (frequency, 0.59), but the minor allele in African American individuals (frequency, 0.14) and individuals of Hispanic American ancestry (frequency, 0.22). Similarly, the C allele at rs9843304 has a frequency of 0.45 in individuals of European ancestry, but has a frequency of approximately 0.8 in African Americans and 0.42 in Hispanic Americans.

Post Hoc Analyses

Supplementary Table 6 shows the associations for dominant and recessive models and population attributable risks for each genotype of the 6 GWAS-significant variants within the NHS and HPFS cohort samples. We did not observe substantially stronger dominance/recessive effects for any of the SNPs compared with the log-additive models that we used for our discovery analyses. We conducted a haplotype analysis for the 2 independent associations in the *ABCG8* locus. In Supplementary Table 7, we show the associations of 6 different haplotype combinations at rs11887534 (C/G) and rs4245791 (T/C). We observed that the presence of at least 1 C-T haplotype at this locus (ie, the C allele at rs11887534 and the T allele at rs4247591) was associated with a substantial increase in the risk of gallstone disease in both males and females, compared with individuals without the CT haplotype. We confirmed using the haplotype analysis that rs11887534 is likely to be the main driver of the *ABCG8* association with gallstone disease risk. We did not observe any evidence for gene–gene interactions (Supplementary Table 8), after correcting for multiple

Table 2. Results of SNPs Associated With Gallstone Disease in Discovery and Replication Data Sets

				Discovery stage						Combined: discovery and replication		
SNP	Hg38/dbSNP 142 location	Gene, variant	Risk allele	RAF[a]	OR[b]	*P* value	Het[c] I^2	Het[c] *P*	RAF[a]	OR[b]	*P* value	OR[b] (95% CI)
rs1260326	chr2:27508073	*GCKR*, P446L	C	0.59	1.12	2.55×10^{-10}	<0.01	.550	0.61	1.12	7.74×10^{-8}	1.12 (1.09–1.15)
rs1025447[d]	chr2:43795831	*DYNC2LI1*, intron	T	0.83	1.18	4.21×10^{-12}	<0.01	.519				1.18 (1.13–1.24)
rs11887534[d]	chr2:43839108	*ABCG8,* D19H	C	0.07	1.69	2.44×10^{-60}	0.728	2.69×10^{-4}	0.07	1.88	1.99×10^{-75}	1.78 (1.70–1.86) 1.80[e] (1.65–1.96)
rs4245791[d,f]	chr2:43847292	*ABCG8*, intron	T	0.69	1.27	1.90×10^{-34}	0.368	.114	0.70	1.31	5.29×10^{-31}	1.28 (1.25–1.32)
rs9843304	chr3:149493600	*TM4SF4*, intron	C	0.45	1.12	6.09×10^{-11}	<0.01	.652	0.45	1.10	3.00×10^{-6}	1.11 (1.08–1.14)
rs6471717	chr8:58464798	*CYP7A1/ UBXN2B*, intergenic	G	0.35	1.11	8.84×10^{-9}	0.573	.016	0.34	1.10	3.16×10^{-6}	1.11 (1.08–1.14)
rs2547231[g]	chr19:47881800	*SULT2A1*, intron	A	0.84	1.17	2.24×10^{-10}	<0.01	.537	0.84	1.17	1.09×10^{-7}	1.12[e] (1.08–1.18) 1.17 (1.13–1.22)

[a]Risk allele frequency (RAF) was calculated using cases and controls.
[b]Odds ratio were obtained from fixed-effect meta-analysis of study-specific effect size estimates adjusted for age and sex in each discovery and replication study.
[c]Heterogeneity (het) I^2 and *P* values from fixed-effects meta-analysis.
[e]Calculated using random-effects meta-analysis (if discovery *P* heterogeneity < .05).
[d]Conditioned on each other, discovery *P* values for rs11887534, rs4245791, and rs1025447 were 2.01×10^{-47}, 3.39×10^{-21}, and 6.14×10^{-10}, respectively.
[f]Proxy SNP for rs4299376 (*P* discovery stage = 1.18×10^{-34}, $R^2 = 0.995$, and D' = 0.999 among 1753 Nurses' Health Study participants).
[g]Proxy SNP for rs296391 (*P* discovery stage = 1.59×10^{-10}, $R^2 = 0.904$, and D' = 0.969 among 1753 Nurses' Health Study participants).

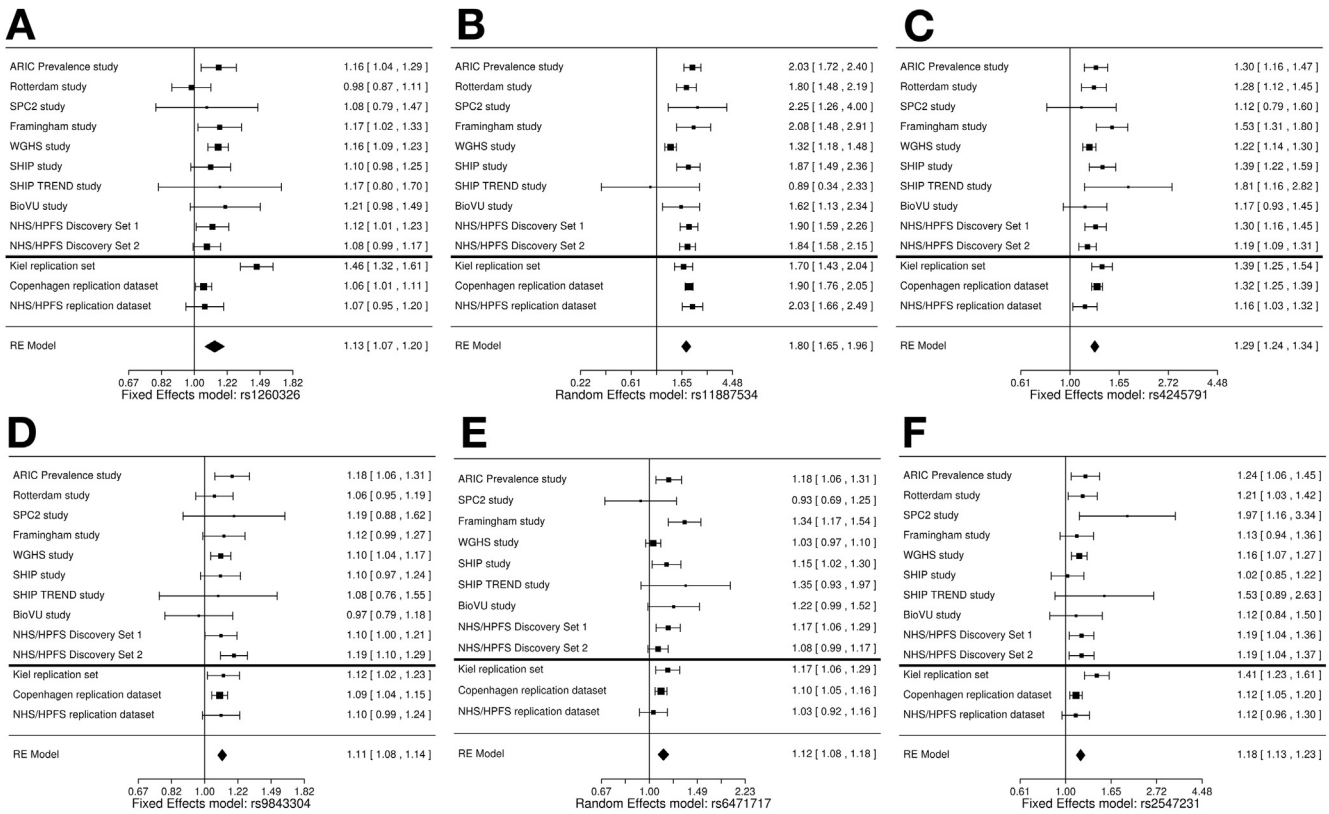


Figure 2. Forest plots of meta-analyses of genome-wide significant SNPs in each of the discovery and replication data sets. (*A*) Fixed-effects meta-analysis: rs1260326. (*B*) Random-effects meta-analysis: rs11887534. (*C*) Fixed-effects meta-analysis: rs4245791. (*D*) Fixed-effects meta-analysis: rs9843304. (*E*) Random-effects meta-analysis: rs6471717. (*F*) Fixed-effects meta-analysis: rs2547231.

comparisons. There was no evidence of confounding of genetic associations after adjusting for self-reported hypercholesterolemia, intake of cholesterol-lowering drugs (ever/never) in the NHS and HPFS cohorts, or for postmenopausal hormone therapy in the NHS cohort (Supplementary Table 9).

The *UGT1A1* SNP rs6742078 did not show an overall association with gallstone disease in log-additive models of our discovery data set ($P < .114$). However, in the NHS and the HPFS cohorts, we replicated the previously reported recessive mode of effect for rs6742078 TT genotype carriers with stronger evidence for association among size among males (OR, 1.45; 95% CI, 1.14–1.85; $P = .00284$), compared with females (OR, 1.16; 95% CI, 1.00–1.34; $P = .0498$)[17,39] (Supplementary Table 10).

After multiple comparisons correction, genetic variants associated with nonalcoholic fatty liver disease were not observed to be associated with overall gallstone disease in our GWAS meta-analysis (data not shown).

Expression QTL and ENCODE Regulatory Analyses of Discovered Loci

Queries of gallstone index and proxy ($R^2 > 0.8$ and $P < 5 \times 10^{-6}$) SNPs showed that several are strong eQTLs (Supplementary Table 11), with some of these located within ENCODE regulatory elements (Supplementary Table 12). Few gene expression studies, and no eQTL studies, have been conducted in gallbladder tissues. Gallstone index SNPs or proxies were the strongest eQTL for *TM4SF4* (in liver), *ABCG8* (in adipose), *SULT2A1* (in liver, brain, and lung), *C2orf16* (in liver), and *LITAF* (in liver, brain, and adipose) (Supplementary Table 13). Studies that have examined associations between SNPs and metabolite levels or ratios in blood have shown that rs2547231 and rs1260326 are highly significantly associated with ratios of metabolites in the cholesterol metabolism pathway (Supplementary Table 14).[40] Results of RNA sequencing from 4 human gallbladders (3 healthy controls and 1 patient with chronic gallstones) and 1 liver sample from the gallstone patient are reported in Table 3. The top GWAS loci *ABCG5/8*, *SULT2A1*, *GCKR*, and *CYP7A1* had higher expression in liver, compared with the gallbladder, suggesting that they may influence the composition of bile. In contrast, *TM4SF4* showed higher expression in gallbladder than the liver, with expression nearly twice as high in the chronic gallstones gallbladder as in the 3 normal samples (Table 3 and Supplementary Figure 5), suggesting a local mechanism of action for this gene in gallbladder.

Discussion

In this large-scale, genome-wide association meta-analysis, we discovered 4 novel susceptibility loci (*SULT2A1*, *TM4SF4*, *GCKR*, and *CYP7A1*) and confirmed 1 known locus (*ABCG8*). The only previous GWAS of gallstone disease, comprising 280

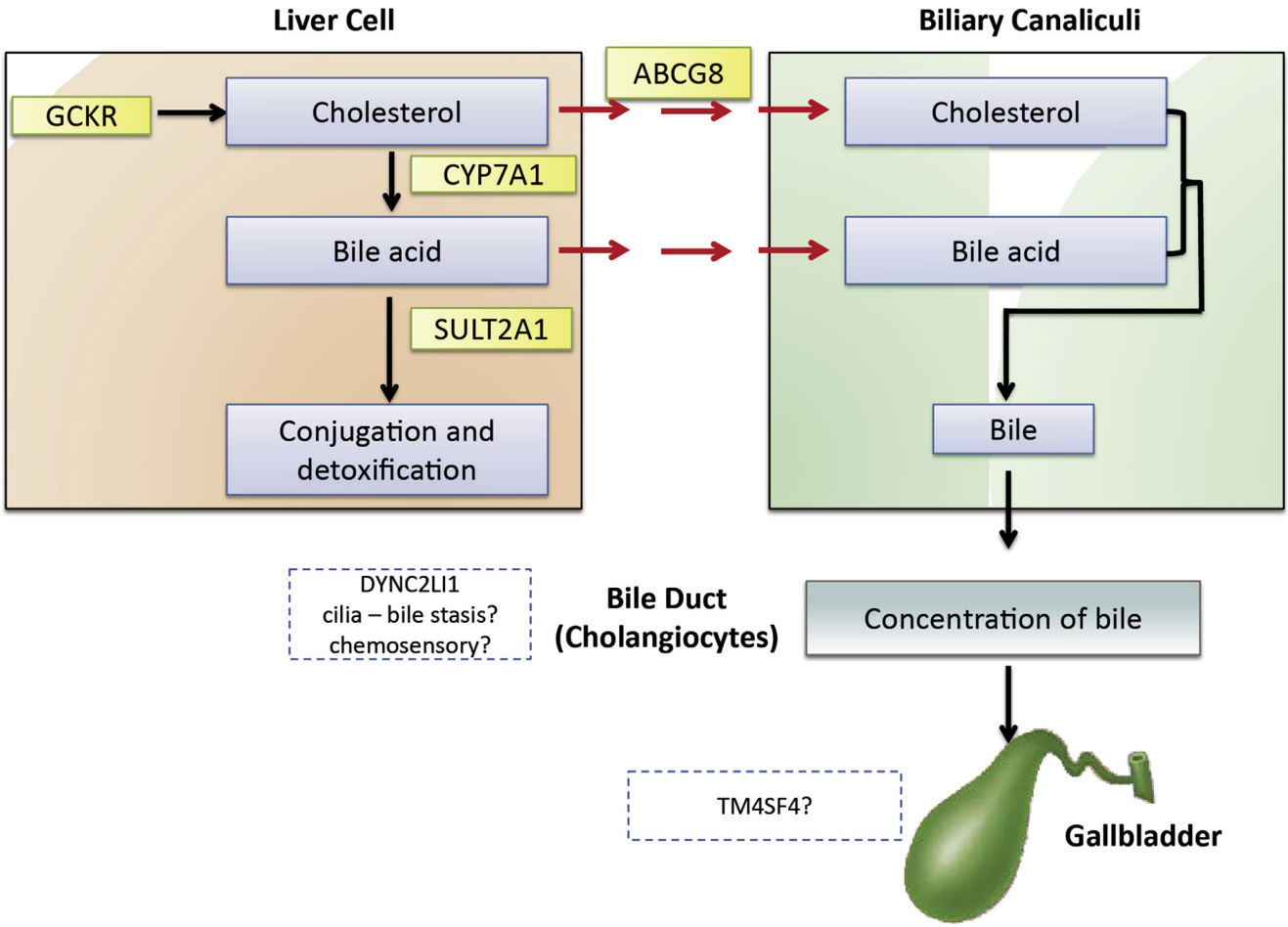


Figure 3. The possible role of novel susceptibility loci in gallstone formation.

cases and 360 controls in the discovery cohort, identified rs11887534 in *ABCG8* as associated with gallstone disease.[15] In addition to confirming this association, we observed an independent association of rs4245791, an intronic variant in *ABCG8*, consistent with results from previous fine-mapping efforts.[41] Thus, there are at least 2 independent gallstone risk variants at the *ABCG8* locus. The biological role of ABCG5/8 is to facilitate efflux of cholesterol from enterocytes and hepatocytes into the intestine and bile, respectively.[42] Therefore, genetic variants in *ABCG5/8* that increase the risk of gallstone disease would be expected to confer a gain-of-function because high bile cholesterol concentration promotes the formation of cholesterol gallstones.[7] Indeed, the gallstone-associated H-allele of D19H has been shown to increase cholesterol efflux approximately 3-fold in vitro, and the gallstone-associated allele of rs4245791 has been associated with increased messenger RNA levels (ie, a gain-of-function effect).[37,43] A third independent association within 5 Mb of rs11887534, mapped to *DYNC2LI1*, was identified, but was not carried forward to replication owing to limited capacity. DYNC2LI1 is a component of cilia structure, and potentially relevant because primary cilia of cholangiocytes regulate osmolarity and flow of bile.[44]

Several of the newly discovered loci are in or near genes known to play a role in cholesterol or bile acid metabolism (Supplementary Table 9 and Figure 3). Association of the

Table 3. RNA Sequencing Reads per Kilobase of Transcript per Million Mapped Reads (RPKM) Values Observed for Genes Near Regions of Discovered SNPs

Locus/gene	Normal gallbladder, (n = 3)[a]	Cholelithiasis gallbladder (n = 1)	Cholelithiasis liver (n = 1)
ABCG5/8	<10	<10	47.3 (*ABCG5*)
TM4SF4	348.07	634	107.7
GCKR	<10	<10	143
SULT2A1	<10	<10	217
CYP7A1	<10	<10	20.6

[a]For normal gallbladder samples the values reflect the mean reads per kilobase of transcript per million across samples.

Table 4. Results of SNPs Associated With Gallstone Disease in African American and Hispanic American Ethnicities

SNP	Risk/other allele	European ancestry meta-analysis		African American ancestry (ARIC) (cases, 115; controls, 2484)			African American ancestry (WHI) (cases, 1384; controls, 6661)			Hispanic American ancestry (WHI) (cases, 1056; controls, 2403)		
		RAF[a]	OR[b] (95% CI)	RAF[a]	OR (95% CI)	*P*	RAF[a]	OR[b] (95% CI)	*P*	RAF[a]	OR[b] (95% CI)	*P*
rs1260326	T/C	0.59	1.12 (1.09–1.15)	0.16	0.90 (0.61–1.34)	.616	0.15	0.86 (0.76–0.97)	.018	0.35	0.85 (0.76–0.95)	.004
rs11887534	C/G	0.07	1.78 (1.70–1.86)	0.06	0.58 (0.26–1.32)	.196	0.09	1.22 (1.08–1.38)	.002	0.20	1.13 (1.02–1.24)	.017
rs4245791	T/C	0.69	1.28 (1.25–1.32)	0.86	1.03 (0.69–1.54)	.877	0.86	1.30 (1.15–1.47)	4.52×10^{-5}	0.78	1.35 (1.19–1.54)	6.82×10^{-6}
rs9843304	C/T	0.45	1.11 (1.08–1.14)	0.85	0.93 (0.61–1.42)	.737	0.78	1.08 (0.98–1.18)	.104	0.42	1.06 (0.96–1.18)	.253
rs6471717	G/A	0.35	1.11 (1.08–1.14)	0.21	1.05 (0.74–1.47)	.801	0.22	0.93 (0.84–1.04)	.196	0.23	1.04 (0.92–1.18)	.513
rs2547231	A/C	0.84	1.17 (1.13–1.22)	0.90	0.77 (0.47–1.26)	.300	0.90	0.92 (0.81–1.06)	.239	0.90	1.12 (0.94–1.33)	.205

[a]Risk allele frequency (RAF) was calculated using cases and controls.
[b]Odds ratio were adjusted for age and sex.

discovered SNPs with the genes was made on the basis of the following: (1) missense mutations as a result of the variant such as D19H in *ABCG8* and P446L in *GCKR*, or (2) mapping of the SNP in the intron of the gene, coupled with strong evidence of association from eQTL (*TM4SF4* and *SULT2A1*) and mQTL data (*GCKR* and *SULT2A1*), or (3) genomic proximity to genes with strong evidence of relevance in cholesterol/bile acid metabolism pathways (eg, CYP7A1). GCKR regulates the conversion of glucose to glucose-6-phosphate in the liver. The *GCKR* P446L variant associated with gallstone disease, even after adjustment for BMI, has been associated with other phenotypes/traits, including lipid levels, glycemic traits, and type 2 diabetes. We postulate that P446L may influence the risk of gallstone disease by increasing the availability of cholesterol to the liver (via high endogenous synthesis), thereby increasing cholesterol concentration in the bile.[45–47] We also identified rs6471717 near *CYP7A1*, associated with gallstone disease. Inside the liver, the rate-limiting step in the conversion of cholesterol to primary bile acids is catalyzed by the enzyme CYP7A1.[48] Thus, genetic variation influencing CYP7A1 activity may influence gallstone disease both via increased cholesterol and by decreased bile acid levels. In support of this, individuals homozygous for deleterious mutations in *CYP7A1* suffer from premature gallstone disease.[49] SULT2A1 catalyzes the conjugation of sulfates to a wide range of steroids and bile acids before biliary excretion.[50] Bile acids help to solubilize biliary cholesterol, and thus prevent gallstone formation. Altered hepatic sulfation of bile acids caused by genetic variation in *SULT2A1* may influence bile acid metabolism and, in turn, biliary levels of bile acids, and, ultimately, the risk of gallstone formation. The rs2547231 variant near SULT2A1 has been associated with SULT2A1 expression,[38] and with the ratio of 2 products of SULT2A1 (X-11440 and androsten-3β,17β-diol disulfate 2).[40] Finally, we found that an intronic variant in *TM4SF4* was associated significantly with gallstone disease. *TM4SF4* encodes transmembrane 4 L 6 family member 4, which has been implicated in liver regeneration as well as pancreas development.[51] The role of *TM4SF4* in gallstone disease has yet to be examined. *TM4SF4* was identified as expressed in liver via eQTL results, with evidence for binding of liver-regulatory elements in ENCODE project data. Furthermore, our RNA sequencing data showed that *TM4SF4* is highly expressed in gallbladder tissue, particularly in the chronic gallstone disease sample. Queries of the Protein Atlas also confirmed the TM4SF4 RNA and protein is expressed most highly in glandular cells of the gallbladder, duodenum, and small intestine, as well as liver bile duct and hepatocytes.[52]

The major strength of this study was the large discovery and replication data sets compared with the only prior gallstone GWAS. However, several limitations are noteworthy. First, we did not have information on gallstone composition (cholesterol/pigment/mixed), and could not discern between stone types. Second, gallstone case definitions varied across cohort settings. However, this concern is minimized by the observation that *ABCG8* D19H, a known susceptibility locus, showed similar risk associations in most subcohorts. Third, the majority of studies defined

gallstones as a history of gallstones or prior cholecystectomy. We expect this led to under-representation of asymptomatic gallstones (~80% of all gallstones are asymptomatic) and would bias toward the null hypothesis. However, because symptomatic gallstone cases require medical interventions, their over-representation may lead to the discovery of markers that have more clinical relevance. Fourth, in ethnicity-specific analyses, we observed an opposite direction of association among European vs African/Hispanic ancestry individuals for rs1260326, which suggests that this variant may not be truly causal, but may be tagging the true causal SNPs, and owing to differences in linkage disequilibrium patterns or haplotype structures across populations, this correlation may be direct in one population and inverse in the other. Nevertheless, the replication of these loci in diverse populations reinforces the importance of these loci in gallstone disease owing to marginal consistent associations across ethnicities. Fifth, another limitation of this study was the relatively small sample size of available RNA sequencing data, which limited our ability to determine whether *cis* genes were expressed in our tissues of interest. However, to our knowledge, there is no database that reports eQTL results for gallbladder tissue and, with this small sample, we could not derive conclusive evidence of comparative expression levels in gallbladder vs liver. Sixth, in the absence of functional studies, the hypothesized associations between SNPs and the genes based on bioinformatics/eQTL data may be speculative, and the true mechanisms by which these SNPs may impact gallstone disease may have been missed. Seventh, we used log-additive models to assess associations with gallstone disease. This may have reduced our ability to detect genetic associations that follow other modes of inheritance. Finally, we may not have been able to detect rare causal alleles in linkage disequilibrium with the most significant GWAS SNPs because conditional analysis using genome-wide complex trait analysis requires a large reference sample to estimate linkage disequilibrium.

In summary, this GWAS meta-analysis of previously genotyped cohorts discovered novel SNPs associated with gallstone disease in European ancestry individuals from 4 distinct and biologically plausible loci. These genetic variants were replicated in independent samples, bringing the total number of GWAS-identified lithogenic loci to 5. Further studies addressing the functionality of these novel candidate genes are warranted to establish their causal role in gallstone development.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at www.gastrojournal.org, and at http://dx.doi.org/10.1053/j.gastro.2016.04.007.

References

1. Shaffer EA. Gallstone disease: epidemiology of gallbladder stone disease. Best Pract Res Clin Gastroenterol 2006;20:981–996.
2. Admirand WH, Small DM. The physicochemical basis of cholesterol gallstone formation in man. J Clin Invest 1968;47:1043–1052.
3. Sarnaik S, Slovis TL, Corbett DP, et al. Incidence of cholelithiasis in sickle cell anemia using the ultrasonic gray-scale technique. J Pediatr 1980;96:1005–1008.
4. Cahalane MJ, Neubrand MW, Carey MC. Physical-chemical pathogenesis of pigment gallstones. Semin Liver Dis 1988;8:317–328.
5. Wang DQ, Carey MC. Characterization of crystallization pathways during cholesterol precipitation from human gallbladder biles: identical pathways to corresponding model biles with three predominating sequences. J Lipid Res 1996;37:2539–2549.
6. Redinger RN, Small DM. Bile composition, bile salt metabolism and gallstones. Arch Intern Med 1972;130:618–630.
7. Portincasa P, Moschetta A, Palasciano G. Cholesterol gallstone disease. Lancet 2006;368:230–239.
8. Sarin SK, Negi VS, Dewan R, et al. High familial prevalence of gallstones in the first-degree relatives of gallstone patients. Hepatology 1995;22:138–141.
9. Hsing AW, Bai Y, Andreotti G, et al. Family history of gallstones and the risk of biliary tract cancer and gallstones: a population-based study in Shanghai, China. Int J Cancer 2007;121:832–838.
10. Nakeeb A, Comuzzie AG, Martin L, et al. Gallstones: genetics versus environment. Ann Surg 2002;235:842–849.
11. Katsika D, Grjibovski A, Einarsson C, et al. Genetic and environmental influences on symptomatic gallstone disease: a Swedish study of 43,141 twin pairs. Hepatology 2005;41:1138–1143.
12. Lammert F, Carey MC, Paigen B. Chromosomal organization of candidate genes involved in cholesterol gallstone formation: a murine gallstone map. Gastroenterology 2001;120:221–238.
13. Lyons MA, Wittenburg H. Cholesterol gallstone susceptibility loci: a mouse map, candidate gene evaluation, and guide to human LITH genes. Gastroenterology 2006;131:1943–1970.
14. Paigen B, Schork NJ, Svenson KL, et al. Quantitative trait loci mapping for cholesterol gallstones in AKR/J and C57L/J strains of mice. Physiol Genomics 2000;4:59–65.
15. **Buch S, Schafmayer C**, Volzke H, et al. A genome-wide association scan identifies the hepatic cholesterol transporter ABCG8 as a susceptibility factor for human gallstone disease. Nat Genet 2007;39:995–999.
16. Goodloe R, Brown-Gentry K, Gillani NB, et al. Lipid trait-associated genetic variation is associated with gallstone disease in the diverse Third National Health and Nutrition Examination Survey (NHANES III). BMC Med Genet 2013;14:120.
17. Buch S, Schafmayer C, Völzke H, et al. Loci from a genome-wide analysis of bilirubin levels are associated with gallstone risk and composition. Gastroenterology 2010;139:1942–1951.e2.
18. **Johnson AD, Kavousi M, Smith AV**, et al. Genome-wide association meta-analysis for total serum bilirubin levels. Hum Mol Genet 2009;18:2700–2710.
19. **Rodriguez S, Gaunt TR, Guo Y**, et al. Lipids, obesity and gallbladder disease in women: insights from genetic

studies using the cardiovascular gene-centric 50K SNP array. Eur J Hum Genet 2016;24:106–112.
20. Volzke H, Alte D, Schmidt CO, et al. Cohort profile: the study of health in Pomerania. Int J Epidemiol 2011; 40:294–307.
21. Colditz GA, Manson JE, Hankinson SE. The Nurses' Health Study: 20-year contribution to the understanding of health among women. J Womens Health 1997;6:49–62.
22. Ridker PM, Chasman DI, Zee RY, et al. Rationale, design, and methodology of the Women's Genome Health Study: a genome-wide association study of more than 25,000 initially healthy American women. Clin Chem 2008;54:249–255.
23. The Atherosclerosis Risk in Communities (ARIC) study: design and objectives. The ARIC investigators. Am J Epidemiol 1989;129:687–702.
24. Kannel WB, Feinleib M, McNamara PM, et al. An investigation of coronary heart disease in families. The Framingham offspring study. Am J Epidemiol 1979;110:281–290.
25. Hofman A, Grobbee DE, de Jong PT, et al. Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. Eur J Epidemiol 1991;7:403–422.
26. Hofman A, Brusselle GG, Darwish Murad S, et al. The Rotterdam Study: 2016 objectives and design update. Eur J Epidemiol 2015;30:661–708.
27. Nothlings U, Krawczak M. [PopGen. A population-based biobank with prospective follow-up of a control group]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2012;55:831–835.
28. Krawczak M, Nikolaus S, von Eberstein H, et al. PopGen: population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. Community Genet 2006;9:55–61.
29. Ritchie MD, Denny JC, Crawford DC, et al. Robust replication of genotype-phenotype associations across multiple diseases in an electronic medical record. Am J Hum Genet 2010;86:560–572.
30. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010;26:2190–2191.
31. Yang J, Lee SH, Goddard ME, et al. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 2011;88:76–82.
32. Cochran WG. The combination of estimates from different experiments. Biometrics 1954;10:101–129.
33. Higgins JPT, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. BMJ 2003;327:557–560.
34. **Johnson AD, Handsaker RE**, Pulit SL, et al. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics 2008;24:2938–2939.
35. Zhang X, Gierman HJ, Levy D, et al. Synthesis of 53 tissue and cell line expression QTL datasets reveals master eQTLs. BMC Genomics 2014;15:532.
36. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res 2012;40:D930–D934.
37. Teupser D, Baber R, Ceglarek U, et al. Genetic regulation of serum phytosterol levels and risk of coronary artery disease. Circ Cardiovasc Genet 2010;3:331–339.
38. **Zhai G, Teumer A, Stolk L**, et al. Eight common genetic variants associated with serum DHEAS levels suggest a key role in ageing mechanisms. PLoS Genet 2011; 7:e1002025.
39. Stender S, Frikke-Schmidt R, Nordestgaard BG, et al. Extreme bilirubin levels as a causal risk factor for symptomatic gallstone disease. JAMA Intern Med 2013; 173:1222–1228.
40. **Shin SY, Fauman EB, Petersen AK**, et al. An atlas of genetic influences on human blood metabolites. Nat Genet 2014;46:543–550.
41. Stender S, Frikke-Schmidt R, Nordestgaard BG, et al. The ABCG5/8 cholesterol transporter and myocardial infarction versus gallstone disease. J Am Coll Cardiol 2014;63:2121–2128.
42. Graf GA, Yu L, Li WP, et al. ABCG5 and ABCG8 are obligate heterodimers for protein trafficking and biliary cholesterol excretion. J Biol Chem 2003;278: 48275–48282.
43. von Kampen O, Buch S, Nothnagel M, et al. Genetic and functional identification of the likely causative variant for cholesterol gallstone disease at the ABCG5/8 lithogenic locus. Hepatology 2013;57:2407–2417.
44. Larusso NF, Masyuk TV. The role of cilia in the regulation of bile flow. Dig Dis 2011;29:6–12.
45. **Teslovich TM, Musunuru K**, Smith AV, et al. Biological, clinical and population relevance of 95 loci for blood lipids. Nature 2010;466:707–713.
46. **Scott RA, Lagou V, Welch RP**, et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. Nat Genet 2012;44:991–1005.
47. **Kottgen A, Albrecht E, Teumer A**, et al. Genome-wide association analyses identify 18 new loci associated with serum urate concentrations. Nat Genet 2013; 45:145–154.
48. Paumgartner G, Sauerbruch T. Gallstones: pathogenesis. Lancet 1991;338:1117–1121.
49. Pullinger CR, Eng C, Salen G, et al. Human cholesterol 7alpha-hydroxylase (CYP7A1) deficiency has a hypercholesterolemic phenotype. J Clin Invest 2002; 110:109–117.
50. Weinshilboum RM, Otterness DM, Aksoy IA, et al. Sulfation and sulfotransferases 1: sulfotransferase molecular biology: cDNAs and genes. FASEB J 1997;11:3–14.
51. Anderson KR, Singer RA, Balderes DA, et al. The L6 domain tetraspanin Tm4sf4 regulates endocrine pancreas differentiation and directed cell migration. Development 2011;138:3213–3224.
52. Uhlen M, Fagerberg L, Hallstrom BM, et al. Proteomics. Tissue-based map of the human proteome. Science 2015;347:1260419.

Author names in bold designate shared co-first authors.

Received July 7, 2015. Accepted April 7, 2016.

Reprint requests
Address requests for reprints to: Amit D. Joshi, MBBS, PhD, Clinical and Translational Epidemiology Unit, Division of Gastroenterology, Massachusetts General Hospital, 55 Fruit Street, Boston, Massachusetts

02114. e-mail: ajoshi@hsph.harvard.edu; fax: (617) 426-0040; or Charlotte Andersson, MD, PhD, The Framingham Heart Study, 73 Mount Wayte Avenue, Framingham, Massachusetts 01702. e-mail: ca@heart.dk; fax: (508) 626-1262; or Andrew T. Chan, MD, MPH, Massachusetts General Hospital and Harvard Medical School, Clinical and Translational Epidemiology Unit, Division of Gastroenterology, GRJ-825C, Boston, Massachusetts 02114. e-mail: achan@mgh.harvard.edu; fax: (617) 726-3673; or Andrew D. Johnson, PhD, Division of Intramural Research, National Heart, Lung and Blood Institute, Cardiovascular Epidemiology and Human Genomics Branch, The Framingham Heart Study, 73 Mount Wayte Avenue, Suite 2, Framingham, Massachusetts, 01702. e-mail: johnsonad2@nhlbi.nih.gov; fax: (508) 626-1262.

Acknowledgments
The authors wish to thank all study participants, researchers, clinicians, technicians, and administrative staff who contributed to this study. Acknowledgments for discovery GWAS studies and replication data sets are available in the Supplementary Materials and Methods section.

Accession number for publicly accessible data repository: RNA sequencing data are available through GEO accession number GSE66430 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE66430).

Conflicts of interest
The authors disclose no conflicts.

Funding
The meta-analysis was supported by National Institutes of Health grants K24DK098311 (A.T.C.). RNA sequencing, the Framingham Heart Study, and expression quantitative trait loci analyses were supported with National Heart, Lung, and Blood Institute Intramural funds.

In addition to the acknowledgments for the meta-analysis, replication studies, and RNA expression studies mentioned in the main article, the authors would like to acknowledge funding for individual GWAS studies.

This research was funded by American College of Gastroenterology (ACG) Junior Faculty Development Award and National Institutes of Health grants: K23DK103119 (M.G.), NHGRI U01HG004728 (L.R.P.), National Eye Institute R01EY015473 (L.R.P.), K24DK098311 (A.T.C.); the German Ministry of Education and Research, and the Deutsche Forschungsgemeinschaft Cluster of Excellence "Inflammation at Interfaces."

SHIP is part of the Community Medicine Research net of the University of Greifswald (Germany), which is funded by the Federal Ministry of Education and Research (grants 01ZZ9603, 01ZZ0103, and 01ZZ0403), the Ministry of Cultural Affairs, as well as the Social Ministry of the Federal State of Mecklenburg–West Pomerania, and the network "Greifswald Approach to Individualized Medicine" funded by the Federal Ministry of Education and Research (grant 03IS2061A). SHIP genome-wide data have been supported by the Federal Ministry of Education and Research (grant 03ZIK012) and a joint grant from Siemens Healthcare (Erlangen, Germany) and the Federal State of Mecklenburg–West Pomerania. The University of Greifswald is a member of the Center of Knowledge Interchange program of the Siemens AG and the Caché Campus program of the InterSystems GmbH.

The generation and management of GWAS genotype data for the Rotterdam Study is supported by The Netherlands Organisation of Scientific Research (NWO) Investments (no. 175.010.2005.011, 911-03-012). This study is funded by the Research Institute for Diseases in the Elderly (014-93-015; Research Institute for Diseases in the Elderly 2), The Netherlands Genomics Initiative/Netherlands Organisation for Scientific Research (NWO) project 050-060-810. The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University (Rotterdam, The Netherlands) Organization for Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly, the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam.

The Atherosclerosis Risk in Communities Study is performed as a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C), R01HL087641, R01HL59367, and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C.

Infrastructure was partly supported by UL1RR025005, a component of the National Institutes of Health and the National Institutes of Health Roadmap for Medical Research.

The data set used from the Vanderbilt University Medical Center's Vanderbilt DNA Biobank was supported by institutional funding and by Vanderbilt Clinical and Translational Science Award (CTSA) grant UL1 TR000445 from National Center for Advancing Translational Sciences (NCATS)/National Institutes of Health, with additional support from grants U01 HG004603 and RC2 GM092618.

The NHS is supported by the National Cancer Institute (P01CA087969, UM1 CA186107, and R01CA137178), the National Institute of Diabetes and Digestive and Kidney Diseases (R01DK058845), and the National Heart, Lung, and Blood Institute (R01HL35464), the National Eye Institute (R01EY015473), with additional support for the collection and management of genetic data. The NHS Breast Cancer GWAS (database of Genotypes and Phenotypes (dbGaP):phs000147.v1.p1) was performed as part of the Cancer Genetic Markers of Susceptibility initiative of the National Cancer Institute (R01CA40356 and U01CA98233). The NHS type 2 diabetes (dbGaP:phs000091.v2.p1) and primary open-angle glaucoma (dbGaP:phs000308.v1.p1) GWAS were funded as part of the Gene Environment-Association Studies project under the National Institutes of Health Genes, Environment, and Health Initiative (type 2 diabetes: U01HG004399, glaucoma U01HG004728). Genotyping for the NHS coronary heart disease GWAS was supported by Merck/Rosetta Research Laboratories (North Wales, PA). The NHS kidney stone GWAS (dbGaP:phs000460.v1.p1) was supported by the National Institute of Diabetes and Digestive and Kidney Diseases (5P01DK070756). The NHS colon cancer GWAS (dbGAP: in progress) was funded as part of the Colorectal Cancer GWAS Consortium funded by the National Cancer Institute (U01 CA137088).

The HPFS is supported by the National Cancer Institute (P01CA055075, UM1 CA167552, and R01CA137178), the National Institute of Diabetes and Digestive and Kidney Diseases (R01DK058845), and the National Heart, Lung, and Blood Institute (R01HL35464), with additional support for the collection and management of genetic data. The HPFS type 2 diabetes (dbGaP:phs000091.v2.p1) and primary open-angle glaucoma (dbGaP:phs000308.v1.p1) GWAS were funded as part of the Gene Environment-Association Studies project under the National Institutes of Health Genes, Environment, and Health Initiative (type 2 diabetes, U01HG004399; primary open-angle glaucoma, U01HG004728). Genotyping for the HPFS coronary heart disease GWAS was supported by Merck/Rosetta Research Laboratories. The HPFS kidney stone GWAS (dbGaP:phs000460.v1.p1) was supported by the National Institute of Diabetes and Digestive and Kidney Diseases (5P01DK070756). The HPFS colon cancer GWAS (dbGAP: in progress) was funded as part of the Colorectal Cancer GWAS Consortium supported by the National Cancer Institute (U01 CA137088, R01CA059045). Also supported by a Harvard Medical School Distinguished Ophthalmology Scholar Award and by the Harvard Glaucoma Center of Excellence (L.P.).

The PopGen 2.0 network is supported by a grant from the German Ministry for Education and Research (01EY1103).

This work was supported by the National Heart, Lung, and Blood Institute's Framingham Heart Study (contract N01-HC-25195) and its contract with Affymetrix, Inc, for genotyping services (contract N02-HL-6-4278). A portion of this research used the Linux Cluster for Genetic Analysis II, funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center.

The data set from Vanderbilt University Medical Center used for the analyses described were obtained from Vanderbilt's DNA Biobank, which is supported by institutional funding and by the Vanderbilt CTSA grant ULTR000445 from NCATS/National Institutes of Health. Genotyping of the Vanderbilt DNA Biobank samples was performed by Vanderbilt Technologies for Advanced Genomics (VANTAGE) which is supported by the Vanderbilt Ingram Cancer Center (P30 CA68485), the Vanderbilt Vision Center (P30 EY08126), and the National Institutes of Health/National Center for Research Resources (NCRR) (G20 RR030956). In addition, funding also was received through the National Human Genome Research Institute (NHGRI) (U01HG006378).

The Women's Genome Health Study is supported by HL043851 and HL080467 from the National Heart, Lung, and Blood Institute, by CA047988 from the National Cancer Institute, the Donald W. Reynolds Foundation, and the Fondation Leducq, with collaborative scientific support and funding for genotyping provided by Amgen, Inc, Thousand Oaks, CA.

The Copenhagen General Study Population and Copenhagen City Heart Study were supported by the Danish Medical Research Council (10-083788); the Research Fund at Rigshospitalet, Copenhagen University Hospital; Chief Physician Johan Boserup and Lise Boserup's Fund; Ingeborg and Leo Dannin's Grant; Henry Hansen and Wife's Grant; and a grant from the Odd Fellow Order.

The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, and the US Department of Health and Human Services through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C.

Supplementary Materials and Methods

Study Sample, Phenotypes, Genotyping, and Imputation

Women's Genome Health Study. The Women's Genome Health Study is a prospective cohort of initially healthy, female North American health care professionals at least 45 years old at baseline representing participants in the Women's Health Study who provided a blood sample at baseline and consent for blood-based analyses. The Women's Health Study was a 2 × 2 trial beginning in 1992–1994 of vitamin E and low-dose aspirin in the prevention of cancer and cardiovascular disease with approximately 10 years of follow-up evaluation. Since the end of the trial, follow-up evaluation has continued in observational mode. Additional information related to health and lifestyle were collected by questionnaire throughout the Women's Health Study trial and continuing observational follow-up evaluation.

Genotyping in the Women's Genome Health Study sample was performed using the HumanHap300 Duo "+" chips or the combination of the HumanHap300 Duo and iSelect chips (Illumina, San Diego, CA) with the Infinium II protocol.[1] In either case, the custom SNP content was the same; these custom SNPs were chosen without regard to MAF to saturate candidate genes for cardiovascular disease as well as to increase coverage of SNPs with known or suspected biological function (eg, disease association, nonsynonymous changes, substitutions at splice sites, and so forth). For quality control, all samples were required to have successful genotyping using BeadStudio v. 3.3 software (Illumina) for at least 98% of the SNPs. A subset of 23,294 individuals were identified with self-reported European ancestry that could be verified on the basis of multidimensional scaling analysis of identity by state using 1443 ancestry informative markers in PLINK version 1.06. In the final data set of these individuals, a total of 339,596 SNPs were retained with a MAF greater than 1%, successful genotyping in 90% of the subjects, and deviations from Hardy–Weinberg equilibrium not exceeding a P value of 10^{-6} in significance. Among the final 23,294 individuals of verified European ancestry, genotypes for a total of 2,608,509 SNPs were imputed from the experimental genotypes for 340,349 SNPs and linkage disequilibrium relationships implicit in the HapMap r. 22 CEU samples. Imputation was performed with MaCH 1.0.16.

Nurses' Health Study I and II and Health Professional's Follow Up Studies. The Nurses' Health Studies comprise female registered nurses in the United States. In 1976, 121,700 women between 30 and 55 years of age were included in the NHS I cohort. In 1989, 116,430 female registered nurses between 25 and 42 years of age were enrolled in NHS II. All individuals completed a baseline mailed questionnaire on their medical history and lifestyle characteristics. Every other year, follow-up questionnaires are sent to both cohorts to update newly diagnosed medical conditions. The response rates consistently have exceeded 90%. The NHS I and II were approved by the institutional review board on the use of human subjects in research of the Brigham and Women's Hospital and the Harvard School of Public Health in Boston.

The Health Professionals Follow-up Study comprises 51,529 men aged 40–75 years in 1986 (29,683 dentists, 10,098 veterinary surgeons, 4185 pharmacists, 3745 optometrists, 2218 osteopathic physicians, and 1600 podiatrists). The study was approved by the institutional review board on the use of human subjects in research of the Harvard School of Public Health in Boston.

In accordance with previous work, the presence of cholecystectomy or self-reported gallstones in NHS, NHS II, and HPFS were used to define cases for the present study.[2] These measures have been validated with high precision previously.[2] Gallstone cases and noncases for whom genotyping data were available from 12 studies for different primary traits within these Harvard cohorts were included in analysis for the present study. The primary traits were as follows: breast cancer,[3] pancreatic cancer,[4] glaucoma,[5] endometrial cancer,[6] colon cancer,[7] ovarian cancer, glioma,[8] prostate cancer,[9] type 2 diabetes,[10] coronary heart disease,[11] kidney stone, gout, and mammographic density.[12] Study participants from 3 broad platform categories—the earlier generation of Illumina arrays (HumanHap), the Illumina OmniExpress array, and Affymetrix 6.0 array, were grouped into 3 nonoverlapping data sets: HumanHap comprising 6 GWAS data sets, OmniExpress comprising 4 GWAS data sets, and Affymetrix 6.0 comprising 2 GWAS data sets. Imputation was performed separately for the 3 data sets using the 1000 Genomes Project ALL Phase I Integrated Release Version 3 Haplotypes excluding monomorphic and singleton sites as reference panel. We obtained data set–specific effect size estimates for the risk of gallstone disease by logistic regression analysis assuming log-additive genetic effects, adjusting for age, cohort (includes sex), primary trait, and top for eigenvectors. We further adjusted for BMI in the sensitivity analysis. All analyses were performed using ProbABEL.[13]

Framingham Heart Study. The Framingham Heart Study is a prospective community-based observational study that aims to investigate risk factors for cardiovascular disease initiated in 1948 by enrollment of the original cohort (n = 5209).[14] In 1971, the children of the original cohort and their spouses were enrolled in the offspring cohort (n = 5124).[15] For the present study we used data from both the original and offspring cohorts. Cases were identified as having a history of gallstones based on questionnaires asking direct questions about prior gallstones, gallbladder disease, or gallbladder surgery. Such questionnaires were available at examination 12 (1971–1974; mean age, 64 years), 13, 17, and 18 (1983–1985; mean age, 74 years) for the original cohort, and for examination 6 (1995–1998; mean age, 59 years) and 7 (1998–2001; mean age, 62 years) for the offspring cohort. Cases were defined as cases from the day when they first replied "yes" to any of the questions, and controls were defined as controls after the last examination during which they had been consecutively free of gallbladder disease. DNA was extracted and genotyped for consenting FHS participants with Affymetrix

500K arrays and additional gene-focused 50K arrays in the SNP Health Association Resource project. FHS used MACH 1.0 to impute approximately 2.54 million SNPs based on the HapMap CEU (Utah residents with Northern and Western European ancestry from the CEPH (Centre d'Etude du Polymorphisme Humain) collection) phased haplotypes (build 22). SNPs used in the imputation process for FHS met the following criteria: MAF 1% or greater, Hardy–Weinberg equilibrium *P* value greater than 1.0×10^{-6}, SNP call rate greater than 97.0%, MISHAP test *P* value greater than 1.0×10^{-9}, and Mendelian errors less than 100.

Rotterdam study. The Rotterdam Study is a prospective cohort study in a suburb (Ommoord) of Rotterdam, The Netherlands.[16] Between 1990 and 1991, all inhabitants aged 55 years and older were invited to participate. In total, 7983 inhabitants agreed to participate (response rate, 78%). At baseline, participants were asked about a history of gallstone disease. Furthermore, they were linked to a hospital admission registry in the region for cases of cholelithiasis, gallbladder disease, cholecystitis, cholecystectomy, or biliary obstruction (ICD codes, 574–576). A total of 5974 Caucasian participants were genotyped successfully (Illumina 550K). Genotyped data were imputed with the HapMap reference panel. The Rotterdam Study has been approved by the medical ethics committee according to the Wet Bevolkingsonderzoek: ERGO (Population Study Act Rotterdam Study), executed by the Ministry of Health, Welfare, and Sports of The Netherlands, and written informed consent was obtained from all study participants.

Atherosclerosis Risk in Communities Study. The ARIC study is a population-based, prospective, cohort study of cardiovascular disease. ARIC included 15,792 individuals aged 45–64 years at baseline (1987–1989) from 4 US communities. Participants were examined 5 times (1987–1989, 1990–1992, 1993–1995, 1996–1998, and 2011–2013). For the present study, we analyzed prevalent, self-reported cases at the study's baseline examination (1987–1989). Information regarding prevalent gallbladder disease at baseline was ascertained retrospectively during the medical history telephone interview (1994–1996).[17] During the interview, participants were asked 2 questions: "Have you ever been diagnosed by a doctor as having gallstones or a gallbladder attack?" and "At what age were you first told you had a gallbladder problem?" Participants who responded "yes" to the first item and whose response to the second item was an age younger than their age at the baseline examination were defined as having prevalent gallbladder disease at baseline. A participant's baseline status was set to missing if he/she failed to complete the follow-up medical history interview. DNA was extracted at baseline or at the second visit. A genome-wide scan was conducted with the Affymetrix Genome-Wide Human SNP Array 6.0 (Affymetrix) in almost the whole ARIC cohort. Quality control (QC) at SNP level included exclusion of SNPs for not passing laboratory QC, no chromosome location, monomorphic, call rate less than 95%, and autosomal SNPs with a Hardy–Weinberg equilibrium *P* value less than 10^{-6}. Imputation to approximately 2.5 million autosomal SNPs identified in HapMap phase II CEU samples was performed using MACH (v1.0.16).[18] SNPs that met the following criteria were used in the following imputation: MAF of 1% or greater, call rate 95% or greater, and a Hardy–Weinberg equilibrium *P* value of 10^{-5} or greater. In the primary analysis, we used a logistic regression model with gallbladder disease as the outcome, assuming an additive genetic effect for SNP dosage and adjusted for age, sex, and field centers. We further adjusted for BMI in the sensitivity analysis. All analyses in ARIC were performed by ProbABEL.[13]

Vanderbilt University DNA Biobank case-control study. Cases and controls were identified from the Vanderbilt University DNA Biobank, which holds data on DNA extracted from blood remaining from routine clinical testing at Vanderbilt University hospital.[19] The Vanderbilt DNA Biobank is linked to the Vanderbilt electronic health record, which includes discharge diagnoses from all hospitalizations registered on the ICD-9.[20] For the present study, we identified cases as having 2 or more ICD-9 codes 574.X (calculus of gallbladder with acute cholecystitis) or a history of cholecystectomy (ICD-9 codes 51.22 [open cholecystectomy], 51.23 [laparoscopic cholecystectomy], or 51.24 [laparoscopic partial cholecystectomy]) that were not performed in conjunction with other intra-abdominal surgeries. Controls comprised an age- and sex-matched sample free from any prior gallstone diagnosis (ICD-9 codes 574.X) or related procedures. All cases and controls were reviewed manually; a positive predictive value greater than 95% was identified for both cases and controls. Relevant ethical committees approved the study.

SHIP and SHIP–TREND cohorts. SHIP and SHIP–TREND are 2 independent cohorts from the Study of Health in Pomerania. The SHIP cohort comprised 4308 randomly selected individuals aged 20–79 years from the general population in the Pomerania district in Germany.[21] The first examination of the SHIP cohort was performed between 1997 and 2001. Another sample of 4420 adults aged 20–79 years subsequently was included in the SHIP-TREND cohort (first examination in 2008–2012). A total of 4081 SHIP and 986 SHIP-TREND subjects with complete GWAS information underwent an abdominal ultrasound (prevalent gallstones: SHIP, n = 843; SHIP–TREND, n = 67) and a full physical examination (exclusions owing to missing ultrasound data or cholecystectomy scar: SHIP, n = 104; SHIP–TREND, n = 101). Before study participation, all individuals provided written, informed consent.

PopGen case-control study. A community-based sample was recruited via the local population registry between 2005 and 2007 and underwent an additional physical examination between 2010 and 2012 at the PopGen facilities that included an abdominal ultrasound by a trained physician. All cases with gallstone disease had undergone cholecystectomy (N = 60) or were diagnosed with cholecystolithiasis (N = 62) using B-mode ultrasonography. The gallstone-free controls were confirmed to be gallstone-free by ultrasonography. For both cases and controls, the study was restricted to probands of German ethnicity; in other words, only individuals whose parents were born in Germany were included. All cases and

controls provided written informed consent before the study, and the study protocol was approved by the institutional review and ethics committees of the Kiel Medical Faculty (Ethikkommission der Medizinischen Fakultät der Christian-Albrechts-Universität Kiel, #A156/03). Details about recruitment and clinical characterization have been reported previously[22,23] (http://www.popgen.de). PopGen participants were genotypes with Affymetrix 6.0 arrays. PopGen samples were imputed with IMPUTEv2 and ShapeITv1 using default parameters based on the 1000 Genomes phase I haplotypes (build 37). Original files were preprocessed using the following measures: variants with MAF less than 0.5% or INFO less than 0.1 were removed.

Kiel case-control replication study. German cases were recruited through clinical centers at Kiel University and all had undergone cholecystectomy for cholecystolithiasis. German controls all were confirmed to be gallstone-free by ultrasonography and were drawn from a randomly selected urban population sample. Details about recruitment and clinical characterization have been reported previously for cases[24] and controls.[25] Written informed consent was obtained from all study participants. The study was approved by the research Ethics Committee of Kiel University Hospital and the Baden-Württemberg General Medical Council (Landesärztekammer Baden-Württemberg).

Copenhagen General Study Population and Copenhagen City Heart Study. Participants in 2 prospective studies of the Danish general population, the Copenhagen General Study Population and Copenhagen City Heart Study, were combined, yielding a total of 60,988 participants, including 3599 with symptomatic gallstone disease. Studies were approved by institutional review boards and Danish ethical committees, and were conducted according to the Declaration of Helsinki. Written informed consent was obtained from participants. All participants were white and of Danish descent, as determined by the National Danish Person Registration System. There was no overlap of individuals between the studies.

The Copenhagen General Study Population[26,27] is a prospective study of the Danish general population initiated in 2003 with ongoing enrollment. Individuals are selected based on the National Danish Civil Registration System to reflect the adult Danish population aged 20–100 years. Data are obtained from a self-administered questionnaire reviewed together with an investigator at the day of attendance, a physical examination, and from blood samples including DNA extraction. We included 52,716 consecutive participants from this study in the present analysis. The Copenhagen City Heart Study[26,27] is a prospective study of the Danish general population initiated in 1976–1978 with follow-up examinations in 1981–1983, 1991–1994, and 2001–2003. Participants were recruited and examined exactly as in the Copenhagen General Study Population. Blood samples for DNA extraction were drawn at the 1991–1994 and 2001–2003 examinations. We included 8272 consecutive participants in the present analysis.

In both studies, diagnoses of symptomatic gallstone disease (ICD-8 codes, 574–575; ICD-10 codes, K80–K81) were collected from the National Danish Patient Registry and the National Danish Causes of Death Registry from January 1, 1977, to May 10, 2011. The National Danish Patient Registry has information on all patient contacts with all clinical hospital departments and outpatient clinics in Denmark, including emergency wards (from 1994). The National Danish Causes of Death Registry contains data on the causes of all deaths in Denmark, as reported by hospitals and general practitioners.

Women's Health Initiative. The WHI is a US study focusing on common health issues in postmenopausal women. A total of 161,808 postmenopausal women aged 50–79 years old were recruited between 1993 and 1998, including 12,151 self-identified African Americans and 5469 self-identified Hispanic Americans. Details of the study design and cohort characteristics have been described previously.[28] Clinical information was collected by self-report and physical examination. All participants provided written informed consent as approved by local Human Subjects Committees. A cohort of 8515 self-identified African American and 3642 self-identified Hispanic American participants from WHI, who had consented to genetic research, were selected for the WHI SNP Health Association Resource project (n = 12,157) and genotyped on the Affymetrix 6.0 array. Genotype quality control criteria included call rate, concordance rates for blinded and unblinded duplicates, and sex discrepancy. Furthermore, individuals whose genetic ancestries differed from self-reported ethnicities and 1 individual from each close relative pair were excluded. In total, 11,740 individuals passed all genotype and sample QC criteria (8153 African American 3587 Hispanic American). Details of the QC procedures have been described in previous WHI–SNP Health Association Resource studies.[29,30] The sample analyzed in the current study included African American and Hispanic American WHI women for whom both DNA samples were genotyped successfully, and for which information was available for gallbladder disease status as well as study covariates.

Expression QTL and ENCODE regulatory analyses. The eQTL SNPs with gene expression associations with a P value less than 5×10^{-6} were queried for overlap with ENCODE regulatory features using HaploReg v3 (http://www.broadinstitute.org/mammals/haploreg/haploreg_v3.php).[31] Blood cell–related eQTL studies included fresh lymphocytes,[32] fresh leukocytes,[33] leukocyte samples in individuals with celiac disease,[34] whole blood samples,[35–49] lymphoblastoid cell lines derived from asthmatic children,[50,51] HapMap lymphoblastoid cell lines from 3 populations,[52] a separate study on HapMap CEU lymphoblastoid cell lines,[53] additional lymphoblastoid cell line population samples,[54–59] CD19+ B cells,[60] primary PHA-stimulated T cells,[54,57] CD4+ T cells,[61] peripheral blood monocytes,[60,62,63] and CD14+ monocytes before and after stimulation with lipopolysaccharide or interferon-γ,[64] CD11+ dendritic cells before and after *Mycobacterium tuberculosis* infection,[65] and a separate study of dendritic cells before or after stimulation with lipopolysaccharide,

influenza, or interferon-β.[66] MicroRNA QTLs[67] and DNase-I QTLs also were queried for lymphoblastoid cell lines.[68]

Non–blood cell tissue eQTLs searched included omental and subcutaneous adipose,[35,43,56,69] stomach,[69] endometrial carcinomas,[70] estrogen receptor positive (ER+) and estrogen receptor negative (ER–) breast cancer tumor cells,[71] liver,[69,72–75] osteoblasts,[76] intestine,[77] normal and cancerous colon,[78] skeletal muscle,[79] breast tissue (normal and cancer),[80,81] lung,[43,81,82] skin,[43,56,83] primary fibroblasts,[54,57,84] sputum,[85] pancreatic islet cells,[86] and heart tissue from left ventricles and left and right atria.[43,87,88] MicroRNA QTLs also were queried for gluteal and abdominal adipose[89] and liver.[90] Further messenger RNA and microRNA QTLs were queried from ER+ invasive breast cancer samples, colon, kidney renal clear, lung, and prostate adenocarcinoma samples.[91]

Brain eQTL studies included brain cortex,[62,92,93] cerebellar cortex,[94] cerebellum,[93,95–98] frontal cortex,[94,95,97] gliomas,[99] hippocampus,[94,97] inferior olivary nucleus (from medulla),[94] intralobular white matter,[94] occiptal cortex,[94] parietal lobe,[96] pons,[95] prefrontal cortex,[97,98,100,101] putamen (at the level of anterior commissure),[94] substantia nigra,[94] temporal cortex,[93–95,97] thalamus,[97] and visual cortex.[98]

Additional eQTL data were integrated from online sources including ScanDB, the Broad Institute GTex browser, and the Pritchard Laboratory (eqtl.uchicago.edu). Cerebellum, parietal lobe, and liver eQTL data were downloaded from ScanDB and cis-eQTLs were limited to those with a P value less than 1.0×10^{-6} and trans-eQTLs with a P value less than 5.0×10^{-8}. The top 1000 eQTL results were downloaded from the GTex Browser at the Broad Institute for 9 tissues on November 26, 2013: thyroid, leg skin (sun exposed), tibial nerve, tibial artery, skeletal muscle, lung, heart (left ventricle), whole blood, and subcutaneous adipose.[43] All GTex results had associations with a P value less than 8.4×10^{-7}.

Genetic Risk Score and Discriminative Ability

In the Kiel data set, the weighted GRS ranged from -2.57 to +4.27, with a median of -0.047. After adjusting for age, sex, and BMI, an increase in 1 SD of weighted GRS was associated with an increased risk of gallstone disease with an OR of 1.50 and a 95% CI of 1.39–1.61. The addition of weighted GRS to a risk prediction model with age, sex, and BMI, showed modest improvements in the Nagelkerke's R^2 from 0.323 to 0.351 and the area under the curve for the receiver operating characteristic plot from 0.783 to 0.798 (Supplementary Figure 4). These improvements in risk prediction measures were similar among males and females in the Kiel cohort.

In the NHS/HPFS replication data set, the weighted GRS ranged from -2.71 to +4.69, with a median of -0.195. The relative risk associated with a SD increase in genetic risk score was 1.33 (1.23, 1.43), after adjusting for age, sex, and BMI at blood draw. The improvement in Nagelkerke's R^2 was from 0.085 to 0.103, and improvement in the area under the curve of the receiver operating characteristic plot was from 0.663 to 0.679. The addition of a GRS yielded a greater improvement in risk prediction in the NHS (women) compared with the HPFS (men) (Supplementary Figure 4).

References

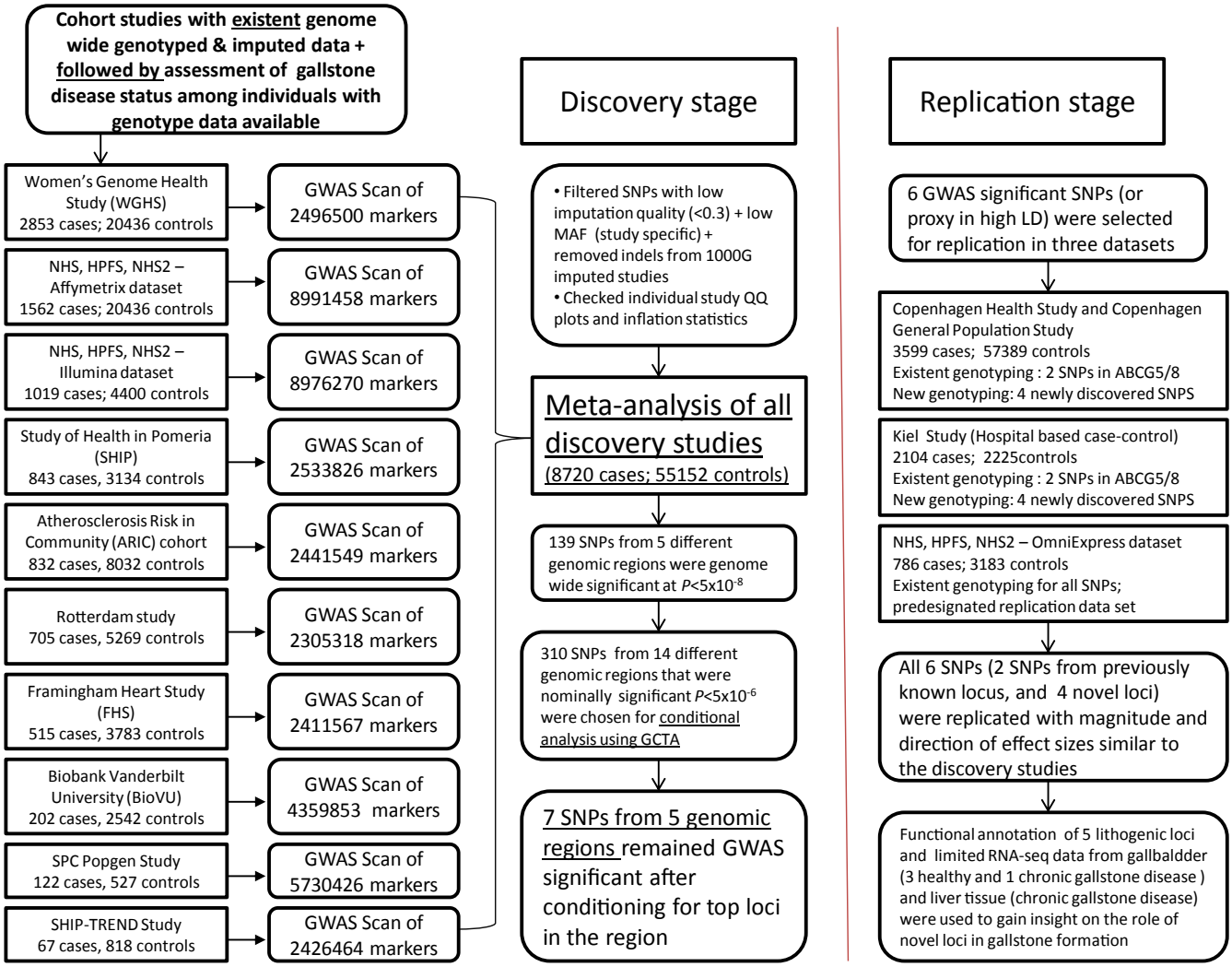
1. Ridker PM, Chasman DI, Zee RY, et al. Rationale, design, and methodology of the Women's Genome Health Study: a genome-wide association study of more than 25,000 initially healthy American women. Clin Chem 2008;54:249–255.
2. Tsai CJ, Leitzmann MF, Willett WC, et al. Statin use and the risk of cholecystectomy in women. Gastroenterology 2009;136:1593–1600.
3. Hunter DJ, Kraft P, Jacobs KB, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. Nat Genet 2007;39:870–874.
4. Amundadottir L, Kraft P, Stolzenberg-Solomon RZ, et al. Genome-wide association study identifies variants in the ABO locus associated with susceptibility to pancreatic cancer. Nat Genet 2009;41:986–990.
5. Wiggs JL, Kang JH, Yaspan BL, et al. Common variants near CAV1 and CAV2 are associated with primary open-angle glaucoma in Caucasians from the USA. Hum Mol Genet 2011;20:4707–4713.
6. De Vivo I, Prescott J, Setiawan VW, et al. Genome-wide association study of endometrial cancer in E2C2. Hum Genet 2014;133:211–224.
7. Peters U, Jiao S, Schumacher FR, et al. Identification of genetic susceptibility loci for colorectal tumors in a genome-wide meta-analysis. Gastroenterology 2013;144:799–807 e24.
8. Rajaraman P, Melin BS, Wang Z, et al. Genome-wide association study of glioma and meta-analysis. Hum Genet 2012;131:1877–1888.
9. Schumacher FR, Berndt SI, Siddiq A, et al. Genome-wide association study identifies new prostate cancer susceptibility loci. Hum Mol Genet 2011;20:3867–3875.
10. Qi L, Cornelis MC, Kraft P, et al. Genetic variants at 2q24 are associated with susceptibility to type 2 diabetes. Hum Mol Genet 2010;19:2706–2715.
11. Jensen MK, Pers TH, Dworzynski P, et al. Protein interaction-based genome-wide analysis of incident coronary heart disease. Circ Cardiovasc Genet 2011;4:549–556.
12. Stevens KN, Lindstrom S, Scott CG, et al. Identification of a novel percent mammographic density locus at 12q24. Hum Mol Genet 2012;21:3299–3305.
13. Aulchenko YS, Struchalin MV, van Duijn CM. ProbABEL package for genome-wide association analysis of imputed data. BMC Bioinformatics 2010;11:134.
14. Dawber TR, Meadors GF, Moore FE Jr. Epidemiological approaches to heart disease: the Framingham Study. Am J Public Health Nations Health 1951;41:279–281.
15. Kannel WB, Feinleib M, McNamara PM, et al. An investigation of coronary heart disease in families. The Framingham offspring study. Am J Epidemiol 1979;110:281–290.
16. Hofman A, Grobbee DE, de Jong PT, et al. Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. Eur J Epidemiol 1991;7:403–422.
17. Boland LL, Folsom AR, Rosamond WD, et al. Hyperinsulinemia, dyslipidemia, and obesity as risk factors for

hospitalized gallbladder disease. A prospective study. Ann Epidemiol 2002;12:131–140.
18. Li Y, Willer CJ, Ding J, et al. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet Epidemiol 2010;34:816–834.
19. Ritchie MD, Denny JC, Crawford DC, et al. Robust replication of genotype-phenotype associations across multiple diseases in an electronic medical record. Am J Hum Genet 2010;86:560–572.
20. Pendergrass S, Dudek SM, Roden DM, et al. Visual integration of results from a large DNA biobank (BioVU) using synthesis-view. Pac Symp Biocomput 2011:265–275.
21. Volzke H, Alte D, Schmidt CO, et al. Cohort profile: the study of health in Pomerania. Int J Epidemiol 2011; 40:294–307.
22. Krawczak M, Nikolaus S, von Eberstein H, et al. PopGen: population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. Community Genet 2006;9:55–61.
23. Nothlings U, Krawczak M. [PopGen. A population-based biobank with prospective follow-up of a control group]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2012;55:831–835.
24. **Buch S, Schafmayer C**, Volzke H, et al. A genome-wide association scan identifies the hepatic cholesterol transporter ABCG8 as a susceptibility factor for human gallstone disease. Nat Genet 2007;39:995–999.
25. Walcher T, Haenle MM, Kron M, et al. Pregnancy is not a risk factor for gallstone disease: results of a randomly selected population sample. World J Gastroenterol 2005; 11:6800–6806.
26. Stender S, Frikke-Schmidt R, Nordestgaard BG, et al. The ABCG5/8 cholesterol transporter and myocardial infarction versus gallstone disease. J Am Coll Cardiol 2014;63:2121–2128.
27. Jorgensen AB, Frikke-Schmidt R, Nordestgaard BG, et al. Loss-of-function mutations in APOC3 and risk of ischemic vascular disease. N Engl J Med 2014; 371:32–41.
28. Hays J, Hunt JR, Hubbell FA, et al. The Women's Health Initiative recruitment methods and results. Ann Epidemiol 2003;13(Suppl):S18–S77.
29. Reiner AP, Beleza S, Franceschini N, et al. Genome-wide association and population genetic analysis of C-reactive protein in African American and Hispanic American women. Am J Hum Genet 2012;91:502–512.
30. Carty CL, Johnson NA, Hutter CM, et al. Genome-wide association study of body height in African Americans: the Women's Health Initiative SNP Health Association Resource (SHARe). Hum Mol Genet 2012;21:711–720.
31. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res 2012;40:D930–D934.
32. Goring HH, Curran JE, Johnson MP, et al. Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. Nat Genet 2007;39:1208–1216.
33. Idaghdour Y, Czika W, Shianna KV, et al. Geographical genomics of human leukocyte gene expression variation in southern Morocco. Nat Genet 2010;42:62–67.
34. Heap GA, Trynka G, Jansen RC, et al. Complex nature of SNP genotype effects on gene expression in primary human leucocytes. BMC Med Genomics 2009;2:1.
35. Emilsson V, Thorleifsson G, Zhang B, et al. Genetics of gene expression and its effect on disease. Nature 2008; 452:423–428.
36. Fehrmann RS, Jansen RC, Veldink JH, et al. Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. PLoS Genet 2011;7:e1002197.
37. **Mehta D, Heim K**, Herder C, et al. Impact of common regulatory single-nucleotide variants on gene expression profiles in whole blood. Eur J Human Genet 2013; 21:48–54.
38. Zhernakova DV, de Klerk E, Westra HJ, et al. DeepSAGE reveals genetic variants associated with alternative polyadenylation and expression of coding and non-coding transcripts. PLoS Genet 2013;9:e1003594.
39. Sasayama D, Hori H, Nakamura S, et al. Identification of single nucleotide polymorphisms regulating peripheral blood mRNA expression with genome-wide significance: an eQTL study in the Japanese population. PLoS One 2013;8:e54967.
40. Landmark-Hoyvik H, Dumeaux V, Nebdal D, et al. Genome-wide association study in breast cancer survivors reveals SNPs associated with gene expression of genes belonging to MHC class I and II. Genomics 2013; 102:278–287.
41. **Westra HJ, Peters MJ, Esko T**, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet 2013;45:1238–1243.
42. **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.
43. The Genotype-Tissue Expression (GTEx) project. Nat Genet 2013;45:580–585.
44. Battle A, Mostafavi S, Zhu X, et al. Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. Genome Res 2014; 24:14–24.
45. Benton MC, Lea RA, Macartney-Coxson D, et al. Mapping eQTLs in the Norfolk Island genetic isolate identifies candidate genes for CVD risk traits. Am J Hum Genet 2013;93:1087–1099.
46. Narahara M, Higasa K, Nakamura S, et al. Large-scale East-Asian eQTL mapping reveals novel candidate genes for LD mapping and the genomic landscape of transcriptional effects of sequence variants. PLoS One 2014;9:e100924.
47. **Quinlan J, Idaghdour Y**, Goulet JP, et al. Genomic architecture of sickle cell disease in West African children. Front Genet 2014;5:26.
48. **Wright FA, Sullivan PF**, Brooks AI, et al. Heritability and genomics of gene expression in peripheral blood. Nat Genet 2014;46:430–437.
49. **Schramm K, Marzi C, Schurmann C**, et al. Mapping the genetic architecture of gene regulation in whole blood. PLoS One 2014;9:e93844.

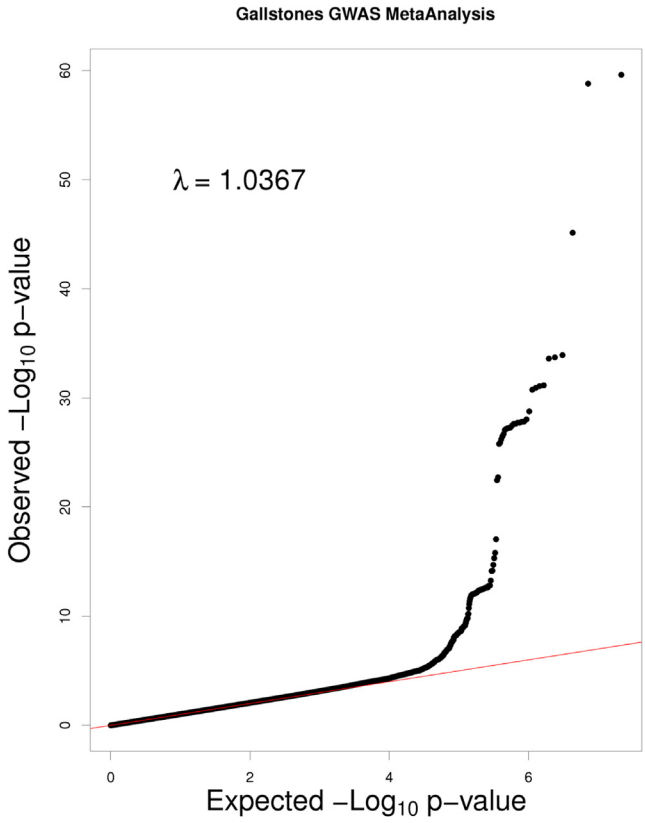
50. **Dixon AL, Liang L, Moffatt MF**, et al. A genome-wide association study of global gene expression. Nat Genet 2007;39:1202–1207.
51. Liang L, Morar N, Dixon AL, et al. A cross-platform analysis of 14,177 expression quantitative trait loci derived from lymphoblastoid cell lines. Genome Res 2013;23:716–726.
52. Stranger BE, Nica AC, Forrest MS, et al. Population genomics of human gene expression. Nat Genet 2007; 39:1217–1224.
53. Kwan T, Benovoy D, Dias C, et al. Genome-wide analysis of transcript isoform variation in humans. Nat Genet 2008;40:225–231.
54. **Dimas AS, Deutsch S, Stranger BE**, et al. Common regulatory variation impacts gene expression in a cell type-dependent manner. Science 2009;325:1246–1250.
55. Cusanovich DA, Billstrand C, Zhou X, et al. The combination of a genome-wide association study of lymphocyte count and analysis of gene expression data reveals novel asthma candidate genes. Hum Mol Genet 2012; 21:2111–2123.
56. **Grundberg E**, Small KS, Hedman AK, et al. Mapping cis- and trans-regulatory effects across multiple tissues in twins. Nat Genet 2012;44:1084–1089.
57. Gutierrez-Arcelus M, Lappalainen T, Montgomery SB, et al. Passive and active DNA methylation and the interplay with genetic variation in gene regulation. eLife 2013;2:e00523.
58. **Mangravite LM, Engelhardt BE**, Medina MW, et al. A statin-dependent QTL for GATM expression is associated with statin-induced myopathy. Nature 2013; 502:377–380.
59. Bryois J, Buil A, Evans DM, et al. Cis and trans effects of human genomic variants on gene expression. PLoS Genet 2014;10:e1004461.
60. Fairfax BP, Makino S, Radhakrishnan J, et al. Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. Nat Genet 2012;44:502–510.
61. Murphy A, Chu JH, Xu M, et al. Mapping of numerous disease-associated expression polymorphisms in primary peripheral blood CD4+ lymphocytes. Hum Mol Genet 2010;19:4745–4757.
62. Heinzen EL, Ge D, Cronin KD, et al. Tissue-specific genetic control of splicing: implications for the study of complex traits. PLoS Biol 2008;6:e1.
63. Zeller T, Wild P, Szymczak S, et al. Genetics and beyond–the transcriptome of human monocytes and disease susceptibility. PLoS One 2010;5:e10693.
64. Fairfax BP, Humburg P, Makino S, et al. Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. Science 2014; 343:1246949.
65. **Barreiro LB, Tailleux L**, Pai AA, et al. Deciphering the genetic architecture of variation in the immune response to Mycobacterium tuberculosis infection. Proc Natl Acad Sci U S A 2012;109:1204–1209.
66. **Lee MN, Ye C**, Villani AC, et al. Common genetic variants modulate pathogen-sensing responses in human dendritic cells. Science 2014;343:1246980.
67. Huang RS, Gamazon ER, Ziliak D, et al. Population differences in microRNA expression and biological implications. RNA Biol 2011;8:692–701.
68. **Degner JF, Pai AA, Pique-Regi R**, et al. DNase I sensitivity QTLs are a major determinant of human expression variation. Nature 2012;482:390–394.
69. Greenawalt DM, Dobrin R, Chudin E, et al. A survey of the genetics of stomach, liver, and adipose gene expression from a morbidly obese cohort. Genome Res 2011;21:1008–1016.
70. Kompass KS, Witte JS. Co-regulatory expression quantitative trait loci mapping: method and application to endometrial cancer. BMC Med Genomics 2011;4:6.
71. Li Q, Seo JH, Stranger B, et al. Integrative eQTL-based analyses reveal the biology of breast cancer risk loci. Cell 2013;152:633–641.
72. **Schadt EE, Molony C, Chudin E**, et al. Mapping the genetic architecture of gene expression in human liver. PLoS Biol 2008;6:e107.
73. **Innocenti F, Cooper GM**, Stanaway IB, et al. Identification, replication, and functional fine-mapping of expression quantitative trait loci in primary human liver tissue. PLoS Genet 2011;7:e1002078.
74. Schroder A, Klein K, Winter S, et al. Genomics of ADME gene expression: mapping expression quantitative trait loci relevant for absorption, distribution, metabolism and excretion of drugs in human liver. Pharmacogenomics J 2013;13:12–20.
75. Wang X, Tang H, Teng M, et al. Mapping of hepatic expression quantitative trait loci (eQTLs) in a Han Chinese population. J Med Genet 2014;51:319–326.
76. **Grundberg E, Kwan T**, Ge B, et al. Population genomics in a disease targeted primary cell model. Genome Res 2009;19:1942–1952.
77. Kabakchiev B, Silverberg MS. Expression quantitative trait loci analysis identifies associations between genotype and gene expression in human intestine. Gastroenterology 2013;144:1488–1496; 1496 e1-3.
78. Ongen H, Andersen CL, Bramsen JB, et al. Putative cis-regulatory drivers in colorectal cancer. Nature 2014; 512:87–90.
79. Keildson S, Fadista J, Ladenvall C, et al. Expression of phosphofructokinase in skeletal muscle is influenced by genetic variation and associated with insulin sensitivity. Diabetes 2014;63:1154–1165.
80. Quigley DA, Fiorito E, Nord S, et al. The 5p12 breast cancer susceptibility locus affects MRPS30 expression in estrogen-receptor positive tumors. Mol Oncol 2014; 8:273–284.
81. **Hao K, Bosse Y, Nickle DC**, et al. Lung eQTLs to help reveal the molecular underpinnings of asthma. PLoS Genet 2012;8:e1003029.
82. Gao C, Tignor NL, Salit J, et al. HEFT: eQTL analysis of many thousands of expressed genes while simultaneously controlling for hidden factors. Bioinformatics 2014;30:369–376.
83. **Ding J, Gudjonsson JE**, Liang L, et al. Gene expression in skin and lymphoblastoid cells: refined statistical method reveals extensive overlap in cis-eQTL signals. Am J Human Genet 2010;87:779–789.

84. Wagner JR, Busche S, Ge B, et al. The relationship between DNA methylation, genetic and expression inter-individual variation in untransformed human fibroblasts. Genome Biol 2014;15:R37.
85. Qiu W, Cho MH, Riley JH, et al. Genetics of sputum gene expression in chronic obstructive pulmonary disease. PLoS One 2011;6:e24395.
86. Fadista J, Vikman P, Laakso EO, et al. Global genomic and transcriptomic analysis of human pancreatic islets reveals novel genes influencing glucose metabolism. Proc Natl Acad Sci U S A 2014;111:13924–13929.
87. **Koopmann TT, Adriaens ME**, Moerland PD, et al. Genome-wide identification of expression quantitative trait loci (eQTLs) in human heart. PLoS One 2014;9:e97380.
88. Lin H, Dolmatova EV, Morley MP, et al. Gene expression and genetic variation in human atria. Heart Rhythm 2014;11:266–271.
89. Rantalainen M, Herrera BM, Nicholson G, et al. MicroRNA expression in abdominal and gluteal adipose tissue is associated with mRNA expression levels and partly genetically driven. PLoS One 2011;6:e27338.
90. Gamazon ER, Innocenti F, Wei R, et al. A genome-wide integrative study of microRNAs in human liver. BMC Genomics 2013;14:395.
91. Li Q, Stram A, Chen C, et al. Expression QTL-based analyses reveal candidate causal genes and loci across five tumor types. Hum Mol Genet 2014;23:5294–5302.
92. **Webster JA, Gibbs JR**, Clarke J, et al. Genetic control of human brain transcript expression in Alzheimer disease. Am J Hum Genet 2009;84:445–458.
93. Zou F, Chai HS, Younkin CS, et al. Brain expression genome-wide association study (eGWAS) identifies human disease-associated variants. PLoS Genet 2012;8:e1002707.
94. **Ramasamy A, Trabzuni D, Guelfi S**, et al. Genetic variability in the regulation of gene expression in ten regions of the human brain. Nat Neurosci 2014;17:1418–1428.
95. **Gibbs JR, van der Brug MP, Hernandez DG**, et al. Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. PLoS Genet 2010;6:e1000952.
96. Gamazon ER, Badner JA, Cheng L, et al. Enrichment of cis-regulatory gene expression SNPs and methylation quantitative trait loci among bipolar disorder susceptibility variants. Mol Psychiatry 2013;18:340–346.
97. Kim S, Cho H, Lee D, et al. Association between SNPs and gene expression in multiple regions of the human brain. Transl Psychiatry 2012;2:e113.
98. **Zhang B, Gaiteri C, Bodea LG**, et al. Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. Cell 2013;153:707–720.
99. Shpak M, Hall AW, Goldberg MM, et al. An eQTL analysis of the human glioblastoma multiforme genome. Genomics 2014;103:252–263.
100. **Colantuoni C, Lipska BK**, Ye T, et al. Temporal dynamics and genetic control of transcription in the human prefrontal cortex. Nature 2011;478:519–523.
101. Liu C, Cheng L, Badner JA, et al. Whole-genome association mapping of gene expression in the human prefrontal cortex. Mol Psychiatry 2010;15:779–784.

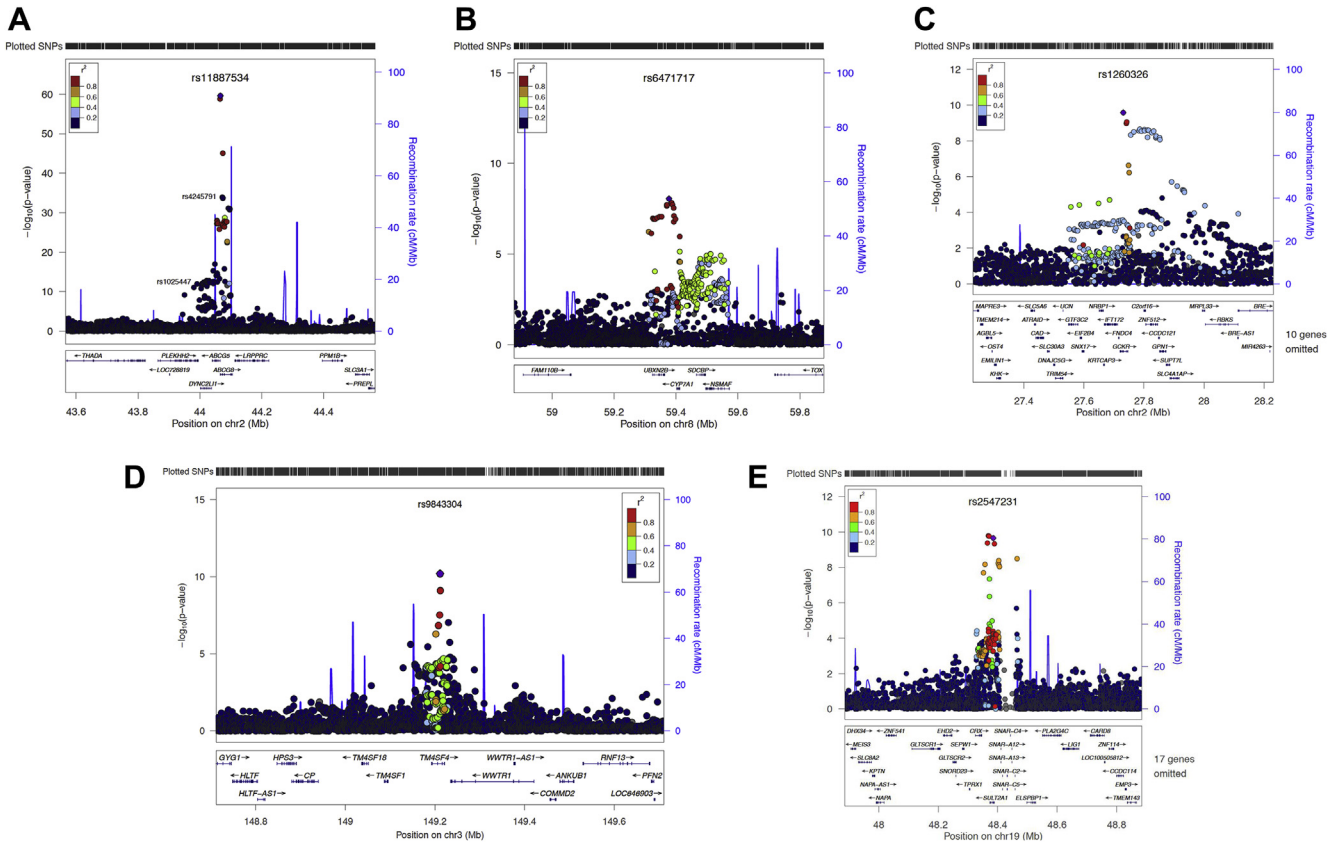
Author names in bold designate shared co-first authors.



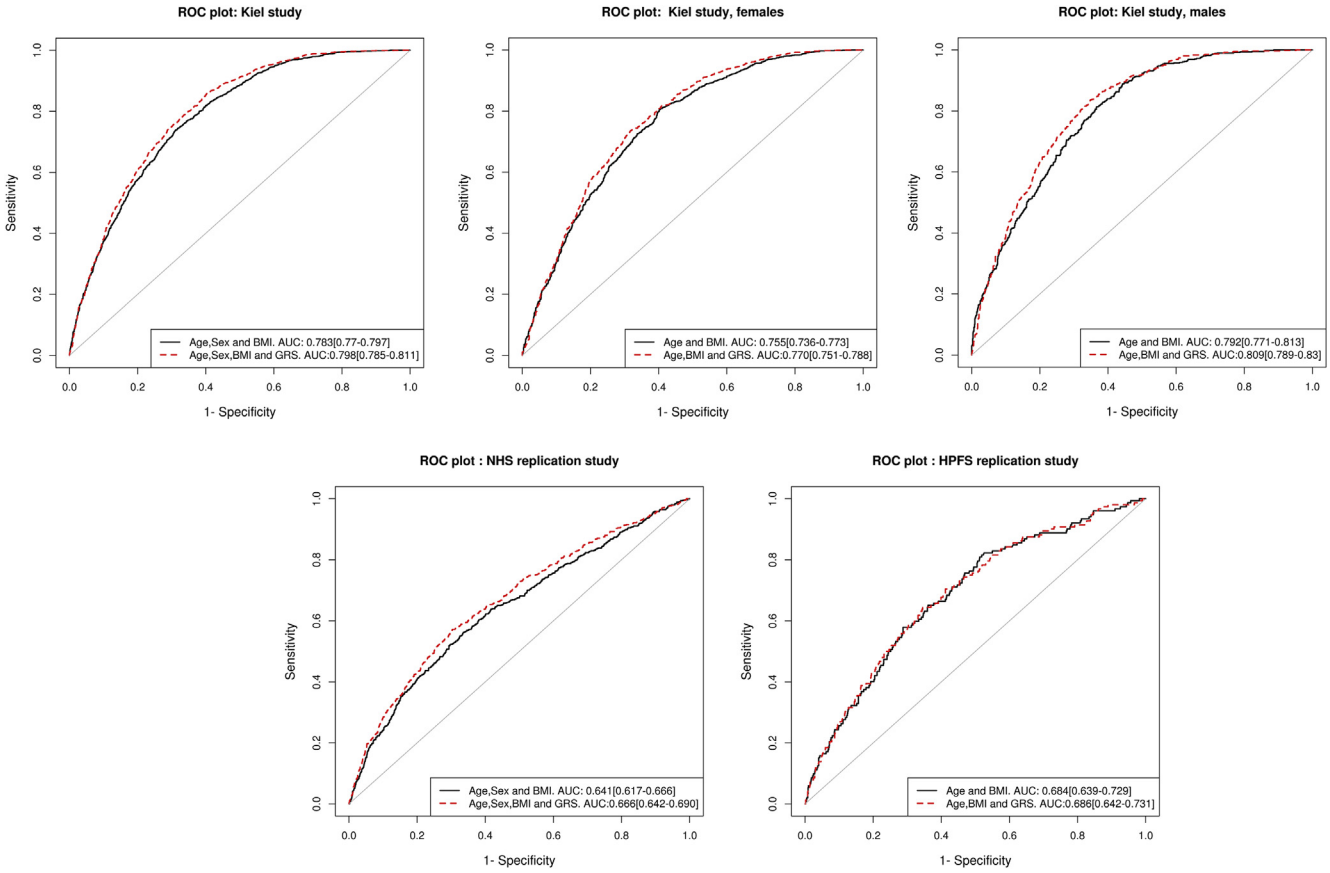
Supplementary Figure 1. Flow chart of study cohorts and methods in the discovery and replication stages. GCTA, Genome-wide complex trait analysis; LD, linkage disequilibrium.



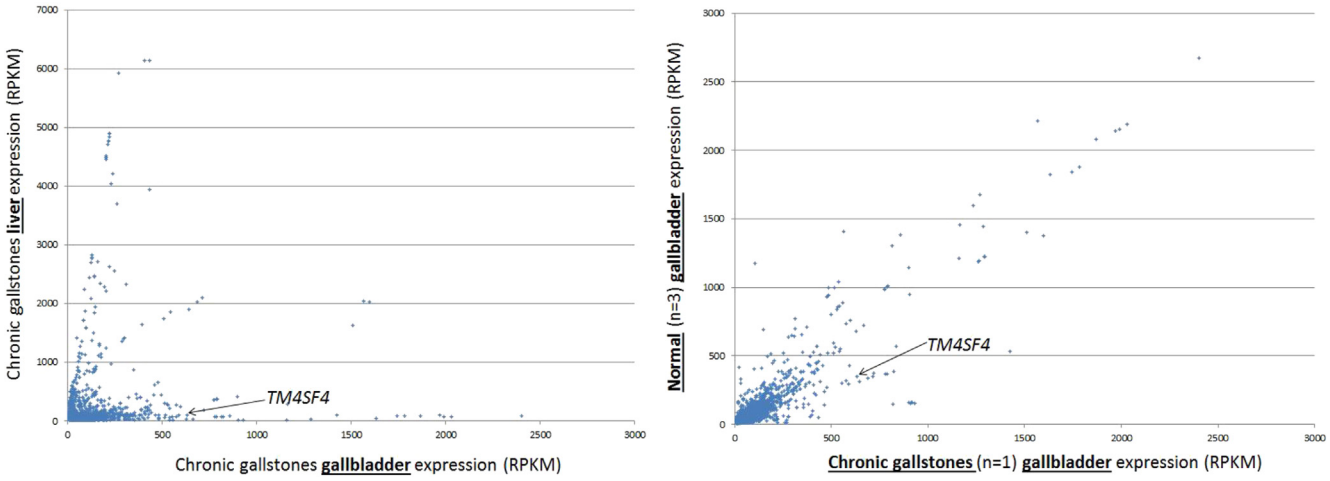
Supplementary Figure 2. Quantile-Quantile plot of gallstone disease GWAS meta-analysis.



Supplementary Figure 3. Regional association plots for discovered loci in GWAS meta-analysis. (*A*) ABCG8 locus, (*B*) CYP7A1 locus, (*C*) GCKR locus, (*D*) TM4SF4 locus, and (*E*) SULT2A1 locus.



Supplementary Figure 4. Receiver operator characteristic plots in replication studies. (*A*) Kiel study (total), (*B*) women in the Kiel study, (*C*) men in the Kiel study, (*D*) NHS replication study, and (*E*) HPFS replication study.



Supplementary Figure 5. RNA sequencing results from gallbladder and liver from chronic gallstones case and normal gallbladders. Comparison of reads per kilobase of transcript per million (RPKM) values for expressed genes in chronic gallstone gallbladder vs chronic gallstone liver (*left panel*) and chronic gallstone gallbladder vs normal (*nongallstones*) gallbladder. The point corresponding to *TM4SF4* expression is indicated. The following genes were excluded from the plots because of their high RPKM values: *MTRNR2L8*, *ALB*, and *APOA2*.

Supplementary Table 1. Outcome Assessment in Discovery and Replication Studies

Discovery studies	Ascertainment of gallstone disease in discovery/replication studies
WGHS	Cases were identified based on questionnaires asking direct questions about gallbladder surgery
NHS1/2/HPFS (Affymetrix) NHS1/2/HPFS (Illumina)	Cases were identified based on self-report in questionnaires that asked about having physician-diagnosed gallstone disease or having undergone cholecystectomy in each follow-up cycle
SHIP	Participants underwent an abdominal ultrasound to identify gallstones and a full physical examination (and participants were excluded owing to missing ultrasound data or cholecystectomy scar)
ARIC	"Have you ever been diagnosed by a doctor as having gallstones or a gallbladder attack?"
Rotterdam	Participants were linked to a hospital admission registry in the region for cases of cholelithiasis, gallbladder disease, cholecystitis, cholecystectomy, or biliary obstruction (ICD codes, 574–576).
FHS	Cases were identified as having a history of gallstones based on questionnaires asking direct questions about prior gallstones, gallbladder disease, or gallbladder surgery
BioVU	For the present study, cases were identified as having $\geq$2 ICD-9 codes (574.X, calculus of gallbladder with acute cholecystitis) or a history of cholecystectomy (ICD-9 codes 51.22 [open cholecystectomy], 51.23 [laparoscopic cholecystectomy], or 51.24 [laparoscopic partial cholecystectomy]) that were not performed in conjunction with other intra-abdominal surgeries
PopGen	All cases with gallstone disease had undergone cholecystectomy or were diagnosed with cholecystolithiasis using B-mode ultrasonography; the gallstone-free controls were confirmed to be gallstone-free by ultrasonography
SHIP-TREND	Participants underwent an abdominal ultrasound to identify gallstones and a full physical examination (and participants were excluded because of missing ultrasound data or cholecystectomy scar)
All discovery samples	
Replication studies	
CCHS and CGPS	Diagnoses of symptomatic gallstone disease (ICD-8, 574–575; ICD-10, K80–K81) were collected from the National Danish Patient Registry and the National Danish Causes of Death Registry
Kiel University	Hospital-based, case-control study in which German cases were recruited through clinical centers at Kiel University and all had undergone cholecystectomy for cholecystolithiasis German controls all were confirmed to be gallstone-free by ultrasonography and were drawn from a randomly selected urban population sample
WHI	Cases were identified based on questionnaires asking about prior gallstones or gallbladder disease

BioVU, Vanderbilt DNA Biobank; CCHS, Copenhagen City Heart Study; CGPS, Copenhagen General Study Population; WGHS, Women's Genome Health Study.

Supplementary Table 2. Annotation of Nominally Significant ($P < 5 \times 10^{-6}$) GWAS SNPs After Conditional Analysis Using Genome-Wide Complex Trait Analysis

Chromosome	SNP	HG38 location	Reference allele	Freq reference	β	*P* meta-analysis	*P* conditional	Gene	Annotation
2	rs1260326	2:27508073	T	0.412	-0.113	2.55×10^{-10}	2.65×10^{-10}	*GCKR*	Missense variant; splice region variant
2	rs1025447	2:43795831	T	0.831	0.165	4.21×10^{-12}	6.14×10^{-10}	*DYNC2LI1*	Intron variant
2	rs11887534	2:43839108	C	0.066	0.527	2.44×10^{-60}	2.01×10^{-47}	*ABCG8*	Missense variant
2	rs4299376	2:43845437	T	0.685	0.237	1.18×10^{-34}	3.39×10^{-21}	*ABCG8*	Intron variant
3	rs4234161	3:72266437	C	0.762	0.099	4.44×10^{-06}	4.66×10^{-06}	-	Intergenic variant
3	rs9843304	3:149493600	T	0.547	-0.113	6.09×10^{-11}	5.54×10^{-11}	*TM4SF4*	Intron variant
6	rs6927914	6:60664964	T	0.223	-0.041	.05689	3.91×10^{-06}	-	Intergenic variant
6	rs6904350	6:60992067	C	0.392	0.084	2.34×10^{-06}	1.46×10^{-13}	-	Intergenic variant
6	rs1577631	6:61237979	A	0.611	0.043	.01584	3.03×10^{-26}	-	Intergenic variant
6	rs1855933	6:61571349	A	0.610	0.040	0.02242	1.03×10^{-47}	*MTRNR2L9*	Upstream gene variant
8	rs6471717	8:58464798	A	0.655	-0.108	8.84×10^{-09}	9.30×10^{-09}	-	Intergenic variant
11	rs1462565	11:23502288	A	0.015	0.327	4.10×10^{-06}	4.06×10^{-06}	-	Intergenic variant
12	rs11061712	12:1367741	A	0.422	-0.082	2.74×10^{-06}	2.80×10^{-06}	*ERC1*	Intron variant
12	rs2277368	12:53714444	C	0.313	0.096	6.56×10^{-07}	6.60×10^{-07}	*CALCOCO1*	Intron variant
16	rs11644920	16:11551157	A	0.687	-0.097	1.80×10^{-07}	1.90×10^{-07}	*LITAF*	Intron variant
16	rs2216730	16:78799555	T	0.840	-0.118	1.56×10^{-06}	1.53×10^{-06}	*WWOX*	Intron variant
18	rs12605943	18:48137422	A	0.366	-0.101	4.89×10^{-06}	5.12×10^{-06}	*ZBTB7C*	Upstream gene variant
19	rs296391	19:47865277	T	0.844	0.168	1.59×10^{-10}	1.54×10^{-10}	-	Intergenic variant
21	rs9979307	21:35635881	A	0.872	-0.134	6.26×10^{-07}	6.11×10^{-07}	-	Intergenic variant

NOTE. Annotations were obtained from University of California, Santa Cruz (UCSC) variant annotation integrator genome.ucsc.edu. Nominally significant = 10 Mb windows around SNPs with $P < 5 \times 10^{-6}$.

Supplementary Table 3. Imputation Quality Scores in Each Study of SNPs Associated With Gallstone Disease in Discovery Sets

SNP	ARIC study	Rotterdam study	SPC2 study	Framingham study	WGHS study	SHIP study	SHIP-TREND study	BioVU study	NHS-HPFS (Illumina)	NHS-HPFS (Affymetrix)
rs4245791	0.98	1.00	0.83	0.89	1.00	0.98	1.00	1.00	0.99	1.00
rs1025447	0.99	1.00	1.00	1.00	1.00	0.99	0.99	1.00	1.00	0.99
rs9843304	0.96	1.00	0.88	1.00	0.99	0.97	0.99	1.00	1.00	0.98
rs1260326	0.98	0.96	0.91	0.99	0.98	0.98	0.98	1.00	1.00	0.97
rs2547231	0.87	1.00	0.80	1.01	1.00	0.81	1.00	1.00	1.00	0.69
rs6471717	1.00	Not imputed	1.00	0.98	0.98	0.99	0.99	0.99	0.99	0.99
rs11887534	0.92	1.00	0.72	0.40	1.00	0.95	0.98	0.97	0.96	0.86

NOTE. Imputation quality scores were obtained using MaCH software in the ARIC, Rotterdam, Framingham, Women's Genome Health Study, SHIP, SHIP-TREND, and NHS/HPFS studies. Imputation quality scores in SPC2 and the Vanderbilt DNA Biobank were obtained using IMPUTEv2.
BioVU, Vanderbilt DNA Biobank; WGHS, Women's Genome Health Study.

Supplementary Table 4. Results of SNPs Associated With Gallstone Disease in Discovery Sets After Adjusting for BMI

SNP	Hg38/dbSNP 142 location	Gene, variant	Risk allele	OR[a]	*P* value
rs11887534	chr2:43839108	*ABCG8*, D19H	C	1.72	7.74×10^{-62}
rs4245791[b]	chr2:43847292	*ABCG8*, intron	T	1.26	8.79×10^{-33}
rs1025447	chr2:43795831	*DYNC2LI1*, intron	T	1.18	7.32×10^{-12}
rs9843304	chr3:149493600	*TM4SF4*, intron	C	1.12	2.41×10^{-11}
rs1260326	chr2:27508073	*GCKR*, P446L	C	1.12	1.39×10^{-09}
rs2547231[c]	chr19:47881800	*SULT2A1*, intron	A	1.18	1.00×10^{-10}
rs6471717	chr8:58464798	*CYP7A1/UBXN2B*, intergenic	G	1.11	2.75×10^{-09}

[a]Odds ratio were adjusted for age, sex, and BMI in each discovery study and for study-specific additional covariates.
[b]Proxy SNP for rs4299376 ($R^2 = 0.995$ and D' = 0.999 among 1753 NHS participants).
[c]Proxy SNP for rs296391 ($R^2 = 0.904$ and D' = 0.969 among 1753 NHS participants).

Supplementary Table 5. Discriminative Accuracy of Genetic Risk Score in the Replication Data Sets

	Odds ratio, per 1-SD increase in GRS	AUC: age, sex, and BMI	AUC: age, sex, BMI, and GRS
NHS/HPFS	1.33 (1.23–1.43)	0.663	0.679
Copenhagen cohorts	1.35 (1.31–1.40)	0.671	0.691
Kiel case-control	1.50 (1.39–1.61)	0.783	0.798

NOTE. Odds ratio estimates were adjusted for age, sex, and BMI at blood draw.
AUC, area under the curve.

Supplementary Table 6. Post Hoc Analysis in the NHS and HPFS Cohorts Assuming Dominant/Recessive Modes of Action for GWAS Significant SNPs and Genotype-Specific Population Attributable Risk

		All samples (males and females)				Females only				Males only			
SNP	Genotype	Ca/Co	OR (95% CI)	*P*	PAR%	Ca/Co	OR (95% CI)	*P*	PAR%	Ca/Co	OR (95% CI)	*P*	PAR%
rs2547231	CC	70/413	1.0 Ref			54/250	1.0 Ref			16/163	1.0 Ref		
	CA	900/4021	1.3 (0.99–1.69)	.0552	7.7	683/2461	1.28 (0.94–1.73)	.119	7.28	217/1560	1.35 (0.77–2.35)	.29	8.76
	AA	2506/10,030	1.46 (1.12–1.89)	.00451	24.2	1913/6063	1.45 (1.08–1.96)	.0144	23.7	593/3967	1.47 (0.85–2.52)	.165	24.7
	CC(ref) vs CA/AA		1.41 (1.09–1.83)	.00905			1.4 (1.04–1.89)	.0263			1.43 (0.84–2.46)	.19	
	CC/CA(ref) vs AA		1.15 (1.06–1.25)	.00121			1.16 (1.05–1.28)	.00234			1.12 (0.94–1.32)	.212	
rs1260326	TT	586/2594	1.0 Ref			414/1540	1.0 Ref			172/1054	1.0 Ref		
	TC	1630/7057	1.02 (0.92–1.13)	.71	0.966	1256/4283	1.09 (0.96–1.23)	.197	4.21	374/2774	0.81 (0.66–1)	.0454	NA
	CC	1260/4813	1.17 (1.05–1.3)	.00599	5.35	980/2951	1.23 (1.08–1.4)	.00196	7.18	280/1862	0.95 (0.76–1.17)	.616	NA
	TT(ref) vs TC/CC		1.08 (0.98–1.19)	.132			1.14 (1.02–1.29)	.0257			0.86 (0.72–1.05)	.134	
	TT/TC(ref) vs CC		1.15 (1.06–1.24)	.000454			1.16 (1.06–1.27)	.00175			1.1 (0.93–1.29)	.273	
rs11887534	GG	2827/12,824	1.0 Ref			2160/7848	1.0 Ref			667/4976	1.0 Ref		
	GC	616/1577	1.78 (1.61–1.98)	0	7.84	467/887	1.92 (1.7–2.17)	$<10^{-10}$	8.51	149/690	1.64 (1.34–2.02)	.00000254	7.2
	CC	33/63	2.45 (1.6–3.76)	.0000385	0.628	23/39	2.12 (1.26–3.56)	.00442	0.495	10/12	3.43 (1.59–7.4)	.0017	1.01
	GG(ref) vs GC/CC		1.81 (1.64–2)	0			1.93 (1.71–2.17)	$<10^{-10}$			1.71 (1.39–2.09)	.000000214	
	GG/GC(ref) vs CC		2.26 (1.48–3.46)	.00018			1.94 (1.16–3.26)	.012			3.19 (1.48–6.87)	.00311	
rs4245791	CC	303/1516	1.0 Ref			233/971	1.0 Ref			70/545	1.0 Ref		
	CT	1445/6467	1.13 (0.99–1.3)	.0794	5.49	1086/3956	1.15 (0.98–1.35)	.0819	6.33	359/2511	1.18 (0.88–1.57)	.267	7.36
	TT	1728/6481	1.35 (1.18–1.55)	.0000144	13.6	1331/3847	1.45 (1.24–1.69)	.00000326	16.5	397/2634	1.24 (0.93–1.65)	.146	10
	CC(ref) vs CT/TT		1.24 (1.09–1.41)	.00122			1.3 (1.12–1.51)	.000674			1.21 (0.92–1.59)	.18	
	CC/CT(ref) vs TT		1.22 (1.13–1.32)	.000000171			1.29 (1.18–1.41)	7.88×10^{-9}			1.08 (0.93–1.26)	.327	
rs9843304	TT	932/4381	1.0 Ref			731/2684	1.0 Ref			201/1697	1.0 Ref		
	TC	1717/7061	1.13 (1.04–1.24)	.00603	5.97	1305/4325	1.11 (1–1.22)	.0565	5.14	412/2736	1.24 (1.03–1.49)	.0265	10.3
	CC	827/3022	1.28 (1.16–1.43)	.00000349	5.53	614/1765	1.27 (1.13–1.44)	.00011	5.15	213/1257	1.42 (1.15–1.76)	.00138	8.49
	TT(ref) vs TC/CC		1.18 (1.08–1.28)	.000127			1.15 (1.05–1.27)	.00361			1.3 (1.09–1.55)	.00406	
	TT/TC(ref) vs CC		1.19 (1.09–1.3)	.00016			1.2 (1.08–1.33)	.000723			1.24 (1.04–1.48)	.0168	
rs6471717	AA	1447/6332	1.0 Ref			1109/3847	1.0 Ref			338/2485	1.0 Ref		
	AG	1555/6417	1.07 (0.99–1.16)	.0984	3.01	1178/3864	1.06 (0.97–1.16)	.217	2.57	377/2553	1.14 (0.96–1.34)	.129	5.91
	GG	474/1715	1.22 (1.08–1.37)	.00118	2.54	363/1063	1.18 (1.03–1.35)	.018	2.13	111/652	1.32 (1.03–1.68)	.0261	3.54
	AA(ref) vs AG/GG		1.1 (1.02–1.19)	.013			1.09 (0.99–1.19)	.0649			1.17 (1–1.37)	.0455	
	AA/AG(ref) vs GG		1.17 (1.05–1.31)	.00436			1.14 (1.01–1.3)	.0387			1.23 (0.98–1.55)	.0699	

Supplementary Table 7. Post Hoc Analysis in the NHS and HPFS Cohorts: Haplotype Analyses at the ABCG5/8 Locus in Relation to Gallstone Disease Risk

rs11887534(G/C*) -rs4245791(C/T*) Haplotype combinations	All samples (males and females)				Females only				Males only			
	Ca/Co	OR (95% CI)	*P*	PAR%	Ca/Co	OR (95% CI)	*P*	PAR%	Ca/Co	OR (95% CI)	*P*	PAR%
G-C/G-C	302/1510	1.0 Ref			233/968	1.0 Ref			69/542	1.0 Ref		
G-C/G-T	1255/5932	1.07 (0.93–1.23)	.369	2.79	941/3626	1.08 (0.92–1.27)	.327	3.2	314/2306	1.11 (0.83–1.49)	.466	4.27
G-T/G-T	1270/5382	1.19 (1.03–1.37)	.0148	6.61	986/3254	1.27 (1.08–1.48)	.00394	9.1	284/2128	1.08 (0.81–1.45)	.597	2.91
G-C/C-T	189/535	1.79 (1.45–2.21)	4.58E-08	2.84	144/330	1.84 (1.44–2.35)	.000000867	3.06	45/205	1.83 (1.19–2.82)	.00611	2.9
G-T/C-T	426/1036	2.09 (1.76–2.47)	$<10^{-10}$	7.24	323/554	2.43 (1.99–2.96)	$<10^{-10}$	8.28	103/482	1.78 (1.26–2.52)	.00114	6.2
C-T/C-T	32/63	2.64 (1.69–4.13)	.000021	0.71	22/39	2.33 (1.36–4.01)	.00221	0.588	10/24	3.73 (1.66–8.38)	.00142	1.14

Supplementary Table 8. Post Hoc Analysis in the NHS and HPFS Cohorts: Gene–Gene Interactions (G × G) Between GWAS Significant SNPs

G × G *P* values	SNP	Upper triangle: interaction *P* values in the HPFS study (males)				
		rs1260326	rs4245791	rs9843304	rs6471717	rs2547231
Lower triangle: interaction *P* values in the NHS study (females)	rs1260326	-	.041	.899	.406	.103
	rs4245791	0.625	-	.808	.781	.848
	rs9843304	0.448	.305	-	.323	.927
	rs6471717	0.727	.883	.737	-	.560
	rs2547231	0.414	.526	.831	.058	-

Supplementary Table 9. Post Hoc Analysis in the NHS and HPFS Cohorts of GWAS Significant SNPs After Adjusting for Potentially Confounding Medication Use

SNP	Association in overall GWAS (age and sex adjusted)	Association in the NHS study (age and BMI adjusted)	Association in NHS after adjustment for age, BMI, and history of self-reported hypercholesterolemia	Association in NHS after adjustment for age, BMI, and cholesterol-lowering drug use	Association in NHS study after adjusting for age and postmenopausal hormone use	Association in the HPFS study (age and BMI adjusted)	Association in the HPFS study after adjusting for age, BMI, and history of self-reported hypercholesterolemia	Association in the HPFS study after adjusting for age, BMI, and cholesterol-lowering drug use
rs11887534	1.78 (1.70–1.86)	1.94 [1.73–2.19]	1.97 [1.75–2.22]	1.95 [1.73–2.19]	1.96 [1.74–2.22]	1.76 [1.44–2.14]	1.76 [1.45–2.15]	1.76 [1.45–2.14]
rs4245791	1.28 (1.25– 1.32)	1.23 [1.15–1.32]	1.24 [1.15–1.33]	1.23 [1.15–1.32]	1.23 [1.15–1.32]	1.08 [0.96–1.21]	1.08 [0.96–1.22]	1.08 [0.96–1.22]
rs9843304	1.11 (1.08–1.14)	1.14 [1.07–1.22]	1.14 [1.07–1.22]	1.14 [1.07–1.22]	1.14 [1.07–1.21]	1.19 [1.07–1.33]	1.19 [1.07–1.33]	1.19 [1.07–1.33]
rs1260326	1.12 (1.09–1.15)	1.11 [1.04–1.19]	1.12 [1.05–1.19]	1.11 [1.04–1.19]	1.11 [1.04–1.19]	0.99 [0.89–1.11]	0.99 [0.89–1.11]	0.99 [0.89–1.11]
rs2547231	1.17 (1.13–1.22)	1.18 [1.08–1.3]	1.19 [1.08–1.30]	1.18 [1.08–1.30]	1.19 [1.08–1.31]	1.11 [0.94–1.31]	1.11 [0.94–1.31]	1.11 [0.94–1.31]
rs6471717	1.11 (1.08–1.14)	1.08 [1.02–1.16]	1.08 [1.01–1.15]	1.08 [1.02–1.16]	1.09 [1.02–1.16]	1.15 [1.02–1.29]	1.15 [1.02–1.28]	1.15 [1.02–1.28]

Supplementary Table 10. Post Hoc Analysis in the NHS and HPFS Cohorts: Association of Previously Reported UGT1A1 SNP rs6742078

		All samples (males and females)				Females only				Males only			
SNP	Genotype	Ca/Co	OR (95% CI)	*P*	PAR%	Ca/Co	OR (95% CI)	*P*	PAR%	Ca/Co	OR (95% CI)	*P*	PAR%
rs6742078	GG	1538/6614	1.0 Ref			1186/3988	1.0 Ref			352/2626	1.0 Ref		
	GT	1521/6372	1.02 (0.94–1.11)	.577	0.873	1157/3887	1.01 (0.92–1.1)	.902	0.441	364/2485	1.08 (0.91–1.27)	.368	3.38
	TT	417/1478	1.22 (1.08–1.38)	.00145	2.2	307/899	1.16 (1–1.34)	.0498	1.61	110/579	1.45 (1.14–1.85)	.00284	4.38
	GG(ref) vs GT/TT		1.06 (0.98–1.14)	.128			1.03 (0.95–1.13)	.453			1.15 (0.98–1.34)	.0813	
	GG/GT(ref) vs TT		1.21 (1.08–1.36)	.0015			1.15 (1.00–1.32)	.0428			1.39 (1.11–1.75)	.00433	

Supplementary Table 11. Concordant *cis*-eQTLs at Gallstone GWAS Susceptibility Loci

Index SNP	Locus label	SNPlabel	Esnp	Esnp <> Index (r^2)	eSNP.p	Tissue	eQTL Transcript	Index <> best Esnp (r^2)	Esnp <> best Esnp (r^2)	Best Esnp	Best eQTL.p
rs1025447	ABCG5/8 + DYNC2LI1	Main	rs1025447	1	4.99E-10	Omental adipose	ABCG8	SameSNP	SameSNP	rs1025447	4.99E-10
					1.18E-06	Subcutaneous adipose	ABCG8	SameSNP	SameSNP	rs1025447	1.18E-06
rs9843304	TM4SF4	Main	rs12633863	1	2.83E-09	Liver (PMID: 21602305)	TM4SF4	SameSNP	0.967	rs12633863	2.83E-09
			rs6774253	0.966	2.98E-06	Liver (PMID: 21637794)	TM4SF4	SameSNP	0.9	rs6774253	2.98E-06
rs1260326	GCKR	Main	rs1260326	1	1.45E-09	Liver (PMID: 21602305)	C2orf16	SameSNP	SameSNP	rs1260326	1.45E-09
rs296381	SULT2A1	Main	rs2547231	0.945	1.14E-55	Cerebellum (all samples)	SULT2A1	SameSNP	0.866	rs2547231	1.14E-55
					2.07E-54	Liver (PMID: 21602305)	SULT2A1	SameSNP	0.866	rs2547231	2.07E-54
					3.56E-26	Cerebellum (Alzheimer's)	SULT2A1	SameSNP	0.866	rs2547231	3.56E-26
					7.07E-24	Visual cortex (all samples)	SULT2A1	SameSNP	0.866	rs2547231	7.07E-24
					6.25E-20	Prefrontal cortex (all samples)	SULT2A1	SameSNP	0.866	rs2547231	6.25E-20
					3.64E-16	Cerebellum (Huntington's)	SULT2A1	SameSNP	0.866	rs2547231	3.64E-16
					1.14E-14	Cerebellum (normal samples)	SULT2A1	SameSNP	0.866	rs2547231	1.14E-14
					2.16E-11	Liver (PMID: 18462017)	SULT2A1	SameSNP	0.866	rs2547231	2.16E-11
					1.50E-09	Visual cortex (Alzheimer's)	SULT2A1	SameSNP	0.866	rs2547231	1.50E-09
					1.29E-08	Prefrontal cortex (Alzheimer's)	SULT2A1	SameSNP	0.866	rs2547231	1.29E-08
					3.03E-08	Visual cortex (Huntington's)	SULT2A1	SameSNP	0.866	rs2547231	3.03E-08
					3.30E-06	Visual cortex (normal samples)	SULT2A1	SameSNP	0.866	rs2547231	3.30E-06
			rs296391	0.972	2.00E-16	Lung (PMID: 23209423)	SULT2A1	SameSNP	0.932	rs296391	2.00E-16
rs2547231	SULT2A1	Replication	rs2547231	1	1.14E-55	Cerebellum (all samples)	SULT2A1	SameSNP	SameSNP	rs2547231	1.14E-55
					2.07E-54	Liver (PMID: 21602305)	SULT2A1	SameSNP	SameSNP	rs2547231	2.07E-54
					3.56E-26	Cerebellum (Alzheimer's)	SULT2A1	SameSNP	SameSNP	rs2547231	3.56E-26
					7.07E-24	Visual cortex (all samples)	SULT2A1	SameSNP	SameSNP	rs2547231	7.07E-24
					6.25E-20	Prefrontal cortex (all samples)	SULT2A1	SameSNP	SameSNP	rs2547231	6.25E-20
					3.64E-16	Cerebellum (Huntington's)	SULT2A1	SameSNP	SameSNP	rs2547231	3.64E-16
					1.14E-14	Cerebellum (normal samples)	SULT2A1	SameSNP	SameSNP	rs2547231	1.14E-14
					2.16E-11	Liver (PMID: 18462017)	SULT2A1	SameSNP	SameSNP	rs2547231	2.16E-11
					1.50E-09	Visual cortex (Alzheimer's)	SULT2A1	SameSNP	SameSNP	rs2547231	1.50E-09
					1.29E-08	Prefrontal cortex (Alzheimer's)	SULT2A1	SameSNP	SameSNP	rs2547231	1.29E-08
					3.03E-08	Visual cortex (Huntington's)	SULT2A1	SameSNP	SameSNP	rs2547231	3.03E-08
					3.30E-06	Visual cortex (normal samples)	SULT2A1	SameSNP	SameSNP	rs2547231	3.30E-06
			rs296391	0.917	2.00E-16	Lung (PMID: 23209423)	SULT2A1	SameSNP	0.799	rs296391	2.00E-16
rs11644920	LITAF	Main	rs11074995	0.957	1.54E-38	Subcutaneous adipose	LITAF	0.955	1	rs3784924	2.37E-42
					1.90E-12	Liver (PMID: 18462017)	LITAF	SameSNP	0.955	rs11074995	1.90E-12
					7.83E-06	Subcutaneous adipose	SNN	0.955	SameSNP	rs11644920	2.55E-06
			rs11074996	0.957	1.48E-38	Subcutaneous adipose	LITAF	0.955	1	rs3784924	2.37E-42
					7.90E-06	Subcutaneous adipose	SNN	0.955	SameSNP	rs11644920	2.55E-06
			rs11644920	1	3.70E-42	Subcutaneous adipose	LITAF	1	1	rs3784924	2.37E-42
					2.55E-06	Subcutaneous adipose	SNN	SameSNP	SameSNP	rs11644920	2.55E-06
			rs12595973	0.957	6.61E-38	Subcutaneous adipose	LITAF	0.955	1	rs3784924	2.37E-42
					8.85E-22	Prefrontal cortex (all samples)	LITAF	SameSNP	0.955	rs12595973	8.85E-22
					4.17E-12	Prefrontal cortex (Alzheimer's)	LITAF	SameSNP	0.955	rs12595973	4.17E-12
			rs12596176	0.957	4.78E-39	Subcutaneous adipose	LITAF	0.955	1	rs3784924	2.37E-42

Supplementary Table 11. Continued

Index SNP	Locus label	SNPlabel	Esnp	Esnp <> Index (r^2)	eSNP.p	Tissue	eQTL Transcript	Index <> best Esnp (r^2)	Esnp <> best Esnp (r^2)	Best Esnp	Best eQTL.p
			rs3784924	1	4.53E-47	Subcutaneous adipose	LITAF	SameSNP	1	rs3784924	4.53E-47
					2.37E-42	Subcutaneous adipose	LITAF	SameSNP	1	rs3784924	2.37E-42
					9.34E-42	Omental adipose	LITAF	SameSNP	1	rs3784924	9.34E-42
					1.00E-16	Liver (PMID: 21637794)	LITAF	SameSNP	1	rs3784924	1.00E-16
					2.05E-13	Visual cortex (all samples)	LITAF	SameSNP	1	rs3784924	2.05E-13
					1.26E-10	Visual cortex (Alzheimer's)	LITAF	SameSNP	1	rs3784924	1.26E-10
					3.44E-08	Liver (PMID: 21637794)	LITAF	SameSNP	1	rs3784924	3.44E-08
					3.30E-06	Subcutaneous adipose	SNN	1	SameSNP	rs11644920	2.55E-06
			rs57792815	0.868	1.20E-09	Subcutaneous adipose	LITAF	SameSNP	0.868	rs57792815	1.20E-09

NOTE. All eQTL results ($P < 1.0 \times^{-05}$) for gallstone main index and replication SNPs are shown that display concordance between index, gallstone-selected eSNP, and best-known eSNP. Concordance was defined as either the same SNP or SNPs in which all 3 pairwise relationship (between eSNP, index SNP and best eSNP) with $r^2 > 0.8$ in HapMap CEU populations as defined by querying SNAP (http://www.broadinstitute.org/mpg/snap/).

Supplementary Table 12. Regulatory Annotations for Gallstone SNPs With eQTL Associations

SNP	Gallstone eQTL	eQTL tissues	Enhancer ENCODE	Enhancer roadmap	DNAse	Proteins	Motifs
rs1025447	ABCG8	Adipose					Foxj2_1;Irf_known10; Irf_known11; Irf_known5; Irf_known6; MIF-1;Nkx3_4
rs1260326	C2orf16	Liver		LIV.A,9_TxEnhG1 H1.BMP4DT,12_EnhWk2 GAS,9_TxEnhG1	AWG,HepG2		NRSF_known3
rs11074995	LITAF	Liver, adipose		BN.SN,14_Enh BN.ITL,14_Enh CCIP.LSMPTP,14_Enh PFM.1,11_EnhWk1 BN.CC,14_Enh PFM.2,11_EnhWk1	AWG,HMEC AWG,HSMM AWG,HSMMtube Duke,pHTE UW,HAEpiC UW,HAc UW,HCFaa UW,HEEpiC UW,HFF-Myc UW,HNPCEpiC UW,PrEC		
rs11074996	LITAF	Adipose		BN.SN,14_Enh BN.ITL,14_Enh CCIP.LSMPTP,14_Enh PFM.1,11_EnhWk1 BN.CC,14_Enh PFM.2,11_EnhWk1			HP1-site-factor;Hand1_1
rs11644920	LITAF	Adipose		CD15.P,9_TxEnhG1 ESO,14_Enh LIV.A,9_TxEnhG1 MSC.ADIPC,11_EnhWk1 R.MUC29,9_TxEnhG1 ADI.NUC,13_EnhA BN.AG,14_Enh DUO.MUC61,9_TxEnhG1 SK.MUS63,9_TxEnhG1 BN.CC,9_TxEnhG1 BN.AC,14_Enh PFM.2,9_TxEnhG1	AWG,A549 UW,AG09319 UW,HCF UW,HCM UW,HCPEpiC UW,HConF UW,HGF UW,HIPEpiC UW,HL-60 UW,HMVEC-LLy UW,HMVEC-dNeo UW,HVMF UW,Monocytes-CD14+_RO01746		ATF2;Nanog_disc1
rs12595973	LITAF	Adipose, brain		ADI.NUC,11_EnhWk1 PFF.2,12_EnhWk2 KID.FE,12_EnhWk2 BN.FE0,12_EnhWk2	AWG,LNCaP UW,HCPEpiC UW,HPdLF	HepG2,MAFF,Stanford HepG2,MAFK,Stanford	GATA_known10; GATA_known9; HDAC2_disc1; HDAC2_disc6; Smad3_2

Supplementary Table 12. Continued

SNP	Gallstone eQTL	eQTL tissues	Enhancer ENCODE	Enhancer roadmap	DNAse	Proteins	Motifs
rs12596176	LITAF	Adipose	H1,7_Weak_Enh	CCCRA.NP,11_EnhWk1 CC.TPC,12_EnhWk2 BR.H35,11_EnhWk1 CD4.NP,11_EnhWk1 IPS.18,11_EnhWk1 BR.MYO,11_EnhWk1 CD4.MP,12_EnhWk2 CD8.NP,11_EnhWk1	AWG,HMEC AWG,HeLa-S3 Duke,Fibrobl UW,HEEpiC UW,HMVEC-LLy UW,Monocytes-CD14+_RO01746 UW,PrEC UW,SAEC	MCF10A-Er-Src,STAT3,Harvard(Weissman), TAM_1uM_36hr	SREBP_known1
rs3784924	LITAF	Liver, adipose, brain		CD15.P,9_TxEnhG1 ESO,14_Enh LIV.A,9_TxEnhG1 MSC.ADIPC,11_EnhWk1 R.MUC29,9_TxEnhG1 ADI.NUC,13_EnhA BN.AG,14_Enh	Duke,Fibrobl	MCF10A-Er-Src,STAT3,Harvard(Weissman), EtOH_0.01pct_4hr	GR_known1
rs57792815	LITAF	Adipose		CD15.P,9_TxEnhG1 ESO,14_Enh LIV.A,9_TxEnhG1 MSC.ADIPC,11_EnhWk1 R.MUC29,9_TxEnhG1 ADI.NUC,13_EnhA BN.AG,14_Enh DUO.MUC61,9_TxEnhG1 SK.MUS63,9_TxEnhG1 BN.CC,9_TxEnhG1 BN.AC,14_Enh PFM.2,9_TxEnhG1			NRSF_disc4;Sp4
rs2547231	SULT2A1	Liver, brain		LIV.A,10_TxEnhG2			Hand1_1;Smad3_2
rs296391	SULT2A1	Lung		HUES48,14_Enh HUES6,14_Enh IPS.20,14_Enh IPS.DF19,14_Enh H1,14_Enh IPS.18,12_EnhWk2	Duke,Osteobl		
rs12633863	TM4SF4	Liver	HepG2,4_Strong_Enh	IPS.DF19,14_Enh IPS.DF6,11_EnhWk1 HUES48,12_EnhWk2 SK.MUS62,12_EnhWk2	AWG,HepG2	HepG2,ELF1,HudsonAlpha HepG2,FOXA1,HudsonAlpha HepG2,HDAC2,HudsonAlpha HepG2,HEY1,HudsonAlpha HepG2,HNF4A,HudsonAlpha HepG2,HNF4A,Stanford,forskolin HepG2,HNF4G,HudsonAlpha HepG2,P300,HudsonAlpha HepG2,SP1,HudsonAlpha HepG2,TAF1,HudsonAlpha HepG2,USF1,HudsonAlpha	Foxa_known2; HDAC2_disc6

Supplementary Table 12. Continued

SNP	Gallstone eQTL	eQTL tissues	Enhancer ENCODE	Enhancer roadmap	DNAse	Proteins	Motifs
rs6774253	TM4SF4	Liver	HepG2,4_Strong_Enh H1, 7_weak_Enh	IPS.DF19,14_Enh IPS.DF6,11_EnhWk1 HUES48,12_EnhWk2 SK.MUS62,12_EnhWk2 ES.I3,11_EnhWk1 NCC.GED2,12_EnhWk2 ADI.MSC,13_EnhA R.SMUS,12_EnhWk2 H1,11_EnhWk1 ES.WA7,12_EnhWk2 DUO.SMUS,12_EnhWk2 BM.MSC,11_EnhWk1 BN.SN,14_Enh PFF.1,11_EnhWk1	Duke,Urothelia,UT189		CIZ;Mef2_disc3

NOTE. All index and replication SNPs with concordant *cis*-eQTL associations (Supplementary Table 5) were queried against haploReg v.3.0 (http://www.broadinstitute.org/mammals/haploreg/haploreg_v3.php).

Supplementary Table 13. Results of Querying Gallstone SNPs and Proxies ($r^2 > 0.8$) in the Genome-Wide Repository of Associations Between SNPs and Phenotypes GWAS Database v. 2.0

GallstoneProxy	GallstoneIndex	Distance (base-pairs)	r^2	D'	Chromosome	Gallstone candidate gene	Pubmed ID	Results location	GWAS *P* value	Trait
rs9921290	rs1260326	0	1	1	2	GCKR	20686565	Supplementary Table 19	1.30E-139	Triglycerides
rs9843304	rs4299376	1855	1	1	2	ABCG8	20529992	Table 1	1.40E-72	Serum phytosterol (sitosterol normalized to cholesterol)
rs9500809	rs4299376	0	1	1	2	ABCG8	20686565	Supplementary Table 19	2.30E-49	LDL cholesterol
rs9476368	rs4299376	0	1	1	2	ABCG8	20686565	Supplementary Table 19	3.20E-47	Total cholesterol
rs9448882	rs1260326	0	1	1	2	GCKR	23263486	Supplementary Table 6	1.25E-44	Serum urate
rs9446581	rs1260326	0	1	1	2	GCKR	23263486	Supplementary Table 15	3.80E-43	C-reactive protein
rs9446578	rs1260326	0	1	1	2	GCKR	22885924	Supplementary Table 3A	2.17E-41	Fasting blood glucose
rs9382866	rs1260326	0	1	1	2	GCKR	19936222	Supplementary Table 1	6.30E-36	HDL cholesterol total lipoprotein fraction concentration
rs9352458	rs1260326	0	1	1	2	GCKR	21676895	Table 2	1.70E-28	FVII
rs9352243	rs1260326	0	1	1	2	GCKR	19936222	Supplementary Table 3	2.79E-28	VLDL cholesterol large lipoprotein fraction concentration
rs9352216	rs1260326	0	1	1	2	GCKR	20686565	Supplementary Table 19	4.40E-28	Total cholesterol
rs9350568	rs1260326	10,297	0.933	1	2	GCKR	20081858	Table 1	3.00E-24	HOMA-IR
rs9343302	rs1260326	0	1	1	2	GCKR	22885924	Supplementary Table 2E	2.74E-22	Fasting insulin
rs9341417	rs1260326	0	1	1	2	GCKR	20081857	Supplementary Table 2	2.26E-21	2-hour glucose
rs9294905	rs1260326	0	1	1	2	GCKR	23022100	Supplementary Table 4	4.10E-19	Serum albumin
rs9294231	rs296381	17,900	0.941	1	19	SULT2A1	21533175	Supplementary Table 1	1.96E-18	Serum dehydroepiandrosterone sulfate
rs9294080	rs1260326	10,297	0.933	1	2	GCKR	19936222	Supplementary Table 3	1.07E-17	APOB assay lipoprotein fraction concentration
rs8192870	rs1260326	11,663	0.932	1	2	GCKR	22829776	Table 1	2.20E-16	Sex hormone-binding globulin concentrations
rs780094	rs1260326	0	1	1	2	GCKR	20383146	Table 2	3.00E-14	Serum creatinine estimated glomerular filtration rate
rs780094	rs6471717	65,660	0.922	1	8	Intergenic, close to CYP7A1	20686565	Supplementary Table 19	2.50E-13	Total cholesterol
rs780094	rs1260326	0	1	1	2	GCKR	22001757	Table 1	3.90E-13	γ-glutamyl transferase
rs780094	rs1260326	10,297	0.933	1	2	GCKR	19936222	Supplementary Table 1	9.80E-13	LDL cholesterol mean size lipoprotein fraction concentration in fasting sample
rs780094	rs1260326	11,663	0.932	1	2	GCKR	21386085	Supplementary Table 6	1.90E-12	Waist circumference and triglycerides
rs780094	rs1260326	0	1	1	2	GCKR	19060906	Supplementary Table 7	8.70E-12	APOC3 (apolipoprotein C-III)
rs780094	rs1260326	10,297	0.933	1	2	GCKR	21194676	Table 1	2.20E-11	Height (adults)
rs780094	rs1260326	0	1	1	2	GCKR	19936222	Supplementary Table 3	2.87E-11	APOA1 assay lipoprotein fraction concentration
rs780094	rs1260326	11,663	0.932	1	2	GCKR	21102463	Table 2	4.70E-11	Crohn's disease
rs780094	rs4299376	1305	1	1	2	ABCG8	23202125	Supplementary Table 9	2.76E-10	Coronary artery disease
rs780094	rs1260326	11,663	0.932	1	2	GCKR	21386085	Table 2	3.00E-10	Triglycerides and blood pressure

Supplementary Table 13. Continued

GallstoneProxy	GallstoneIndex	Distance (base-pairs)	r^2	D'	Chromosome	Gallstone candidate gene	Pubmed ID	Results location	GWAS *P* value	Trait
rs780094	rs11887534	0	1	1	2	ABCG8	19936222	Supplementary Table 3	3.48E-10	APOB assay lipoprotein fraction concentration
rs780094	rs1260326	10,297	0.933	1	2	GCKR	19936222	Supplementary Table 3	6.98E-10	IDL total lipoprotein fraction concentration
rs780094	rs1260326	0	1	1	2	GCKR	22139419	Table 1	9.12E-10	Platelet count
rs780094	rs1260326	11,663	0.932	1	2	GCKR	23362303	Supplementary Table 3	9.80E-10	Plasma palmitoleic acid
rs780094	rs1260326	10,297	0.933	1	2	GCKR	20081858	Table 2	1.30E-09	Type 2 diabetes
rs780094	rs6471717	65,660	0.922	1	8	Intergenic, close to CYP7A1	20686565	Supplementary Table 19	1.90E-09	LDL cholesterol
rs780094	rs1260326	10,297	0.933	1	2	GCKR	21829377	Text	2.52E-09	Plasma docosapentaenoic acid levels
rs780094	rs1025447	0	1	1	2	DYNC2LI1	20686565	Full GWAS scan	2.76E-09	LDL cholesterol
rs780093	rs1260326	0	1	1	2	GCKR	23118302	Supplementary Table 2	9.40E-09	Lipoprotein-associated phospholipase A2 mass
rs780093	rs1025447	0	1	1	2	DYNC2LI1	20686565	Full GWAS scan	1.13E-08	Total cholesterol
rs780093	rs1260326	10,297	0.933	1	2	GCKR	21423719	Supplementary Table 5	2.59E-08	Nonalcoholic fatty liver disease

Genome-wide Repository of Associations between SNPs and Phenotypes database: (http://apps.nhlbi.nih.gov/grasp/).
APOB, apolipoprotein B; FVII, Coagulation Factor VII; HDL, high-density lipoprotein; HOMA-IR, Homeostatic model assessment for insulin resistance; IDL, Intermediate-density lipoprotein; LDL, low-density lipoprotein; VLDL, very low density lipoprotein.

Supplementary Table 14. Results of Querying Gallstone SNPs and Proxies ($r^2 > 0.8$) in the Atlas of Genetic Influences on Human Blood Metabolites

Locus and gene ID (Cytoband)	SNP	Biochemical(s)	N	EA/OA[1]	EAF[2]	Effect (SE)	*P* value	eQTL	Reference (PMID)
136. SULT2A1 (*19q13.32*)	rs2547231	X-11440/4-androsten-3β, 17β-diol disulfate 2	7240	A/C	0.83	0.141 (0.005)	3.06E-191	Yes	23093944
15. GCKR (*2p23.3*)	rs1260326	Glucose/ mannose	7310	T/C	0.41	0.041 (0.002)	2.50E-148	Yes	23362303;23362303; 21829377;21886157;22286219

Data from PMID: 24816252; An atlas of genetic influences on human blood metabolites Supplementary Table 4.