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

Short title: GWAS meta-analysis of gallstone disease

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†,‡], Andrew D. Johnson[4,20†,‡]

***equal contribution first co-author**

†equal contribution senior co-author

‡to whom correspondence should be addressed

Institutions:
1) Program in Genetic Epidemiology and Statistical Genetics, Harvard School of Public Health, Boston, MA; 2) Division of Gastroenterology, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA; 3) Clinical and Translational Epidemiology Unit, Massachusetts General Hospital Boston, MA; 4) The National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, MA; 5) Medical Department 1, University Hospital Dresden, TU Dresden, Dresden Germany; 6) Department of Clinical Biochemistry, Rigshospitalet, Copenhagen, Denmark; 7) Department of Internal Medicine, Erasmus Medical Center, Rotterdam, the Netherlands; 8) Department of Epidemiology, Erasmus Medical Center, Rotterdam, the Netherlands; 9) Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, MN; 10) Department of Medicine, Vanderbilt University, Nashville, TN; 11) Department of Cardiology, The Heart Centre, Rigshospitalet, Copenhagen University Hospital, Denmark; 12) Joseph J. Zilber School of Public Health, University of Wisconsin, Milwaukee; 13) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA; 14) Department of Internal Medicine I, Ulm University Hospital, Ulm, Germany; 15) Division of Rheumatology, Allergy, and Immunology, Massachusetts General Hospital, Boston, MA; 16) Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 17) Renal Division, Department of Medicine, Brigham and Women's Hospital, Boston, MA ; 18) Department of Biomedical Informatics, Vanderbilt University, Nashville, TN; 19)

Department of Epidemiology, Harvard School of Public Health, Boston, MA ; 20) Population Sciences Branch, National Heart, Lung, and Blood Institute, National Institutes of Health, Framingham, MA; 21) Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany; 22) Department of Medical Oncology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA.; 23) Department of Nutrition, Harvard School of Public Health, Boston, MA; 24) Department of Gastroenterology and Hepatology, Erasmus MC, Rotterdam, the Netherlands; 25) Toronto Centre for Liver Disease, Toronto Western and General Hospital, University Health Network, Toronto, Canada; 26) Institute of Epidemiology, Christian Albrechts Universität Kiel, Niemannsweg 11, Kiel, Germany; 27) The Copenhagen General Population Study and 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, MA; 31) Division of Preventive Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 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, University Medicine Greifswald, Germany; 34) Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany; 35) German Center for Cardiovascular Research, Partner Site Greifswald; 36) German Center for Diabetes Research, Site Greifswald; 37) The National Heart, Lung, and Blood Institute, DNA Sequencing Core Laboratory, Bethesda, MD

To whom correspondence should be addressed:
Amit D. Joshi, MBBS, PhD, Clinical and Translational Epidemiology Unit, Division of Gastroenterology, Massachusetts General Hospital, 55 Fruit Street, Boston , Massachusetts 02114, USA. Tel: +1 617 724 7558; Email:ajoshi@hsph.harvard.edu
Charlotte Andersson, MD, PhD, The Framingham Heart Study, 73 Mt Wayte Avenue, Framingham, Massachusetts 01702, USA. E-mail: ca@heart.dk
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, USA. Tel:+1 617 724 0283; Fax: +1 617 726 3673; Email: achan@mgh.harvard.edu
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 Mt. Wayte Ave., Suite #2, Framingham, MA, 01702, USA. Tel: +1 508 663 4082; Fax: +1 508 626 1262;
Email:johnsonad2@nhlbi.nih.gov

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), GCTA (genome-wide complex trait analysis), GWAS (genome-wide association studies), HPFS (Health Professionals Follow-up Study), 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).

Table of Contents

Supplementary Methods

Study sample, phenotypes, genotyping and imputation

Women's Genome Health Study (WGHS)

The Women's Genome Health Study (WGHS) 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 (WHS) who provided

a blood sample at baseline and consent for blood-based analyses. The WHS was a 2x2 trial beginning in 1992-1994 of vitamin E and low dose aspirin in prevention of cancer and cardiovascular disease with about 10 years of follow-up. Since the end of the trial, follow-up has continued in observational mode. Additional information related to health and lifestyle were collected by questionnaire throughout the WHS trial and continuing observational follow-up.

Genotyping in the WGHS 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 minor allele frequency (MAF) to saturate candidate genes for cardiovascular disease as well as to increase coverage of SNPs with known or suspected biological function, e.g. disease association, non-synonymous changes, substitutions at splice sites, etc. For quality control, all samples were required to have successful genotyping using the BeadStudio v. 3.3 software (Illumina, San Diego, CA) 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 using1443 ancestry informative markers in PLINK v. 1.06. In the final dataset of these individuals, a total of 339596 SNPs were retained with MAF >1%, successful genotyping in 90% of the subjects, and deviations from Hardy-Weinberg equilibrium not exceeding $P=10^{-6}$ 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 LD 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 U.S. 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 have consistently 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 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] Gallstones cases and non-cases for whom genotyping data was available from twelve studies for different primary traits within these Harvard cohorts were included in analysis for the present study. The primary traits were – 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 three broad platform categories – the earlier generation of Illumina arrays (HumanHap), the Illumina OmniExpress array and Affymetrix 6.0 array were grouped into three non-overlapping datasets –

HumanHap comprising six GWAS datasets, OmniExpress comprising four GWAS datasets and Affymetrix 6.0 comprising two GWAS datasets. Imputation was done separately for the three datasets using 1000 Genomes Project ALL Phase I Integrated Release Version 3 Haplotypes excluding monomorphic and singleton sites as reference panel. We obtained dataset specific effect size estimates for the risk of gallstone disease by logistic regression analysis assuming log-additive genetic effects, adjusting for age, cohort (includes gender), primary trait, and top for eigenvectors. We further adjusted for BMI in the sensitivity analysis. All analyses in were performed using ProbABEL[13].

Framingham Heart Study (FHS)

The Framingham Heart Study is a prospective community-based observational study aiming 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 into 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 exam 12 (1971-1974, mean age 64 years), 13, 17, and 18 (1983-85, mean age 74) for the original cohort, and for exam 6 (1995-98, mean age 59, and 7 (1998-2001, mean age 62) for the offspring cohort. Cases were defined as cases from the day where they first replied ‘yes’ to any of the questions and controls were defined as controls after the last exam in 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 (SHARe) project. FHS

used MACH 1.0 to impute ~2.54 million SNPs based on the HapMap CEU phased haplotypes (build 22). SNPs used in the imputation process for FHS met the following criteria: MAF ≥ 1%, HWE P > 1.0 X 10-6, SNP call rate >97.0%, MISHAP test P > 1.0 X 10-9, Mendelian errors <100.

Rotterdam Study

The Rotterdam Study is a prospective cohort study in a suburb (Ommoord) in Rotterdam, the Netherlands [16]. Between 1990-1991, all inhabitants aged 55 years and older were invited to participate. In total, 7,983 inhabitants agreed to participate (response rate 78%). At baseline, participants were enquired 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 5,974 Caucasian participants were successfully genotyped (Illumna 550K). Genotyped data was 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 (ARIC)

The ARIC study is a population-based prospective cohort study of cardiovascular disease. ARIC included 15,792 individuals aged 45-64 years at baseline (1987-89) from four US communities. Participants have been examined 5 times (1987-89, 1990-92, 1993-95, 1996-98, and 2011-13). For the present study we analyzed prevalent, self-reported cases at the study's baseline exam (1987-1989).

Information regarding prevalent gallbladder disease at baseline was ascertained retrospectively during the medical history phone interview (1994–96)[17] . During the interview, participants were asked two 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?". Those who responded "Yes" to the first item and whose response to the second item was an age younger than their age at the baseline exam 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 the second visit. A genome-wide scan was conducted with the Affymetrix Genome-Wide Human SNP Array 6.0 (Affymetrix, Santa Clara, CA, USA) in almost the whole ARIC cohort. QC at SNP level included exclusion of SNPs for not passing laboratory QC, no chromosome location, monomorphic, call rate <95%, and autosomal SNPs with HWE-p < 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 imputation: MAF ≥ 1%, call rate ≥ 95%, and Hardy-Weinberg equilibrium (HWE)-p ≥ 10^{-5}. 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, gender, and field centers. We further adjusted for BMI in the sensitivity analysis. All analyses in ARIC were performed by ProbABEL[13] .

Vanderbilt University BioVU case-control study

Cases and controls were identified from the Vanderbilt University, BioVU, which holds data on DNA extracted from blood remaining from routine clinical testing at Vanderbilt University hospital.[19] BioVU is linked to the Vanderbilt

electronic health record, which included discharge diagnoses from all hospitalizations registered on the international classification of diseases (ICD), 9th version.[20] For the present study, we identified cases as having ≥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. Controls comprised an age and sex-matched sample free from any prior gallstone diagnosis (ICD-9 574.X) or related procedures. All cases and controls were manually reviewed; a positive predictive value >95% was identified for both cases and controls. Relevant ethical committees approved the study.

SHIP and SHIP-TREND cohorts

SHIP and SHIP-TREND are two 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 undertaken between 1997 and 2001. Another sample of 4420 adults aged 20–79 years was subsequently included in the SHIP-TREND cohort (first examination in 2008-12). 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=67) and a full physical examination (exclusions due to missing ultrasound data or cholecystectomy scar, SHIP n=104, SHIP-TREND n=101). Prior to study participation, all individuals gave 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 gave written informed consent prior to the study, and the study protocol has been 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 has 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 <0.5% or INFO <0.1 were removed

Kiel case-control replication study

German cases were recruited through clinical centers at Kiel University and had all undergone cholecystectomy for cholecystolithiasis. German controls were all 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 two prospective studies of the Danish general population, the CGPS and CCHS, were combined, yielding a total of 60,988 participants, including 3,599 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 CGPS[26, 27] is a prospective study of the Danish general population initiated in 2003 with ongoing enrollment. Individuals were selected based on the National Danish Civil Registration System to reflect the adult Danish population aged 20-100 years. Data were 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 CCHS [26, 27] is a prospective study of the Danish general population initiated in 1976-78 with follow-up examinations in 1981-83, 1991-94, and 2001-03. Participants were recruited and examined exactly as in the CGPS. Blood samples for DNA extraction were drawn at the 1991-94 and 2001-2003 examinations. We included 8,272 consecutive participants in the present analysis.

In both studies, diagnoses of symptomatic gallstone disease (ICD8: 574-575; ICD10: K80-K81) were collected from the National Danish Patient Registry and the National Danish Causes of Death Registry from January 1, 1977 to May 10th, 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 U.S.-wide 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 AAs and 5,469 self-identified HAs. Details of the study design and cohort characteristics have been previously described [Hays J., Hunt J.R., Hubbell F.A., Anderson G.L., Limacher M., Allen C., Rossouw J.E. The Women's Health Initiative recruitment methods and results. Ann. Epidemiol. 2003;13(9, Suppl):S18–S77.] 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 8,515 self-identified AA and 3,642 self-identified HA participants from WHI, who had consented to genetic research, were selected for WHI SHARe (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 differ from self-reported ethnicities and one individual from each close relative pair were excluded. In total, 11,740 individuals passed all genotype and sample QC criteria (8,153 AA, 3,587 HA). Details of the QC procedures have been described in previous WHI-SHARe studies [Reiner A.P., Beleza S., Franceschini N., Auer P.L., Robinson J.G., Kooperberg C., Peters U., Tang H. 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. Carty C.L., Johnson N.A., Hutter C.M., Reiner A.P., Peters U., Tang H., Kooperberg C. 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.] The sample analyzed in the current study included African American and Hispanic American WHI women for whom both DNA samples were successfully genotyped, 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 $P<5x10^{-06}$ were queried for overlap with ENCODE regulatory features using HaploReg v3 available at http://www.broadinstitute.org/mammals/haploreg/haploreg_v3.php.[28] Blood cell related eQTL studies included fresh lymphocytes,[29] fresh leukocytes,[30] leukocyte samples in individuals with Celiac disease,[31] whole blood samples,[32-46] lymphoblastoid cell lines (LCL) derived from asthmatic children,[47, 48] HapMap LCL from 3 populations,[49] a separate study on HapMap CEU LCL,[50] additional LCL population samples,[51-56] CD19+ B cells,[57] primary PHA-stimulated T cells,[51, 54] CD4+ T cells,[58] peripheral blood monocytes,[57, 59, 60] and CD14+ monocytes before and after stimulation with LPS or interferon-gamma,[61] CD11+ dendritic cells before and after Mycobacterium tuberculosis infection,[62] and a separate study of dendritic cells before or after stimulation with LPS, influenza or interferon-beta.[63] Micro-RNA QTLs,[64] and DNase-I QTLs were also queried for LCL.[65]

Non-blood cell tissue eQTLs searched included omental and subcutaneous adipose,[32, 40, 53, 66] stomach,[66] endometrial carcinomas,[67] ER+ and ER- breast cancer tumor cells,[68] liver,[66, 69-72] osteoblasts,[73] intestine,[74] and normal and cancerous colon,[75] skeletal muscle,[76] breast tissue (normal and cancer),[77, 78] lung,[40, 78, 79] skin,[40, 53, 80] primary fibroblasts,[51, 54, 81] sputum,[82] pancreatic islet cells,[83] and heart tissue from left ventricles and left and right atria.[40, 84, 85] Micro-RNA QTLs were also queried for

gluteal and abdominal adipose,[86] and liver.[87] Further mRNA and micro-RNA QTLs were queried from ER+ invasive breast cancer samples, colon-, kidney renal clear-, lung- and prostate-adenocarcinoma samples.[88]

Brain eQTL studies included brain cortex,[59, 89, 90] cerebellar cortex,[91] cerebellum,[90, 92-95] frontal cortex,[91, 92, 94] gliomas,[96] hippocampus,[91, 94] inferior olivary nucleus (from medulla),[91] intralobular white matter,[91] occiptal cortex,[91] parietal lobe,[93] pons,[92] pre-frontal cortex,[94, 95, 97, 98] putamen (at the level of anterior commussure),[91] substantia nigra,[91] temporal cortex,[90-92, 94] thalamus,[94] and visual cortex.[95]

Additional eQTL data was integrated from online sources including ScanDB, the Broad Institute GTex browser, and the Pritchard Lab (eqtl.uchicago.edu). Cerebellum, parietal lobe and liver eQTL data was downloaded from ScanDB and cis-eQTLs were limited to those with $P<1.0x10^{-6}$ and trans-eQTLs with $P<5.0x10^{-8}$. The top 1000 eQTL results were downloaded from the GTex Browser at the Broad Institute for 9 tissues on 11/26/2013: thyroid, leg skin (sun exposed), tibial nerve, tibial artery, skeletal muscle, lung, heart (left ventricle), whole blood, and subcutaneous adipose.[40] All GTex results had associations with $p<8.4 x10^{-7}$.

Genetic risk score and discriminative ability

In the Kiel dataset, the weighted GRS ranged from -2.57 to + 4.27, with a median of -0.047. After adjusting for age, gender and BMI, an increase in 1 standard deviation of weighted GRS was associated with an increased risk of gallstone disease with an OR = 1.50, 95% CI = 1.39, 1.61. The addition of weighted GRS to a risk prediction model with age, gender and BMI, showed modest improvements in the Nagelkerke's R2 from 0.323 to 0.351 and the area under curve (AUC) for the receiver operating characteristic (ROC) plot from 0.783 to 0.798. (**Supplementary Fig. 4**).

These improvements in risk prediction measures were similar among males and females in the Kiel cohort.

In the NHS/HPFS replication dataset, the weighted GRS ranged from -2.71 to + 4.69, with a median of -0.195. The relative risk associated with a standard deviation increase in genetic risk score was 1.33 (1.23, 1.43), after adjusting for age, gender and BMI at blood draw. The improvement in Nagelkerke's R^2 was from 0.085 to 0.103, and improvement in AUC of the ROC plot from 0.663 to 0.679. The addition of a GRS yieled a greater improvement in risk prediction the NHS (women) compared to the HPFS (men) (**Supplementary Fig. 4**).

Acknowledgements

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

This research was funded by NIH grants – ACG Junior Faculty Development Award NIH K23DK103119 (MG), NHGRI U01HG004728 (LRP), NEI R01EY015473(LRP); K24DK098311 (ATC); German Ministry of Education and Research (BMBF) and the Deutsche Forschungsgemeinschaft (DFG) 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 no. 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 (GANI_MED)' 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 no. 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 (nr. 175.010.2005.011, 911-03-012). This study is funded by the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands

Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) project nr. 050-060-810. The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), 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 carried out 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 Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research.

The dataset used from the Vanderbilt University Medical Center's BioVU was supported by institutional funding and by the Vanderbilt CTSA grant UL1 TR000445 from NCATS/NIH, with additional support from grants U01 HG004603 and RC2 GM092618.

The Nurses' Health Study (NHS) is supported by the National Cancer Institute (NCI, P01CA087969, UM1 CA186107, R01CA137178), the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK, R01DK058845), and the National Heart, Lung, and Blood Institute (NHLBI, R01HL35464), the National Eye Institute (NEI, R01EY015473), with additional support for the collection and management of

genetic data. The NHS Breast Cancer GWAS (dbGaP:phs000147.v1.p1) was performed as part of the Cancer Genetic Markers of Susceptibility initiative of the NCI (R01CA40356, U01CA98233). The NHS type 2 diabetes (T2D, dbGaP:phs000091.v2.p1) and primary open-angle glaucoma (POAG, dbGaP:phs000308.v1.p1) GWAS were funded as part of the Gene Environment-Association Studies (GENEVA) project under the NIH Genes, Environment, and Health Initiative (T2D: U01HG004399, GA: 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 NIDDK (5P01DK070756). The NHS colon cancer GWAS (dbGAP: in progress) was funded as part of the Colorectal Cancer GWAS Consortium funded by the NCI (U01 CA137088).

The Health Professionals Follow-up Study (HPFS) is supported by the National Cancer Institute (P01CA055075, UM1 CA167552, R01CA137178), the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK, R01DK058845), and the National Heart, Lung, and Blood Institute (NHLBI, R01HL35464), with additional support for the collection and management of genetic data. The HPFS type 2 diabetes (T2D, dbGaP:phs000091.v2.p1) and primary open-angle glaucoma (POAG, dbGaP:phs000308.v1.p1) GWAS were funded as part of the Gene Environment-Association Studies (GENEVA) project under the NIH Genes, Environment, and Health Initiative (T2D: U01HG004399, POAG: U01HG004728). Genotyping for the HPFS coronary heart disease GWAS was supported by Merck/Rosetta Research Laboratories, North Wales, PA. The HPFS kidney stone GWAS (dbGaP:phs000460.v1.p1) was supported by NIDDK (5P01DK070756). The HPFS colon cancer GWAS (dbGAP: in progress) was funded as part of the Colorectal

Cancer GWAS Consortium supported by the NCI (U01 CA137088, R01CA059045). A Harvard Medical School Distinguished Ophthalmology Scholar Award supports Dr. Louis Pasquale. Dr. Pasquale is also supported by the Harvard Glaucoma Center of Excellence.

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 No. N01-HC-25195) and its contract with Affymetrix, Inc. for genotyping services (Contract No. N02-HL-6-4278). A portion of this research utilized the Linux Cluster for Genetic Analysis (LinGA-II) funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center.

The dataset from Vanderbilt University Medical Center used for the analyses described were obtained from Vanderbilt's BioVU, which is supported by institutional funding and by the Vanderbilt CTSA grant ULTR000445 from NCATS/NIH. Genotyping of the BioVU samples was performed by VANTAGE, which is supported by the Vanderbilt Ingram Cancer Center (P30 CA68485), the Vanderbilt Vision Center (P30 EY08126), and NIH/NCRR (G20 RR030956). In addition, funding was also received through the NHGRI grant: U01HG006378.

The WGHS is supported by HL043851 and HL080467 from the National Heart, Lung, and Blood Institute and 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.

The CGPS and CCHS were supported by the Danish Medical Research Council (grant no. 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 Women's Health Initiative (WHI) program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, and the United States Department of Health and Human Services through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C. The authors thank the WHI investigators and staff for their dedication and the studyparticipants for making the program possible.

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;54:249-55.
2. Tsai CJ, Leitzmann MF, Willett WC, et al. Statin use and the risk of cholecystectomy in women. Gastroenterology 2009;136:1593-600.
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;39:870-4.
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;41:986-90.
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;20:4707-13.
6. De Vivo I, Prescott J, Setiawan VW, et al. Genome-wide association study of endometrial cancer in E2C2. Hum Genet 2014;133:211-24.
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;131:1877-88.

9. Schumacher FR, Berndt SI, Siddiq A, et al. Genome-wide association study identifies new prostate cancer susceptibility loci. Hum Mol Genet 2011;20:3867-75.
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;19:2706-15.
11. Jensen MK, Pers TH, Dworzynski P, et al. Protein interaction-based genome-wide analysis of incident coronary heart disease. Circ Cardiovasc Genet 2011;4:549-56.
12. Stevens KN, Lindstrom S, Scott CG, et al. Identification of a novel percent mammographic density locus at 12q24. Hum Mol Genet 2012;21:3299-305.
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;41:279-81.
15. Kannel WB, Feinleib M, McNamara PM, et al. An investigation of coronary heart disease in families. The Framingham offspring study. American journal of epidemiology 1979;110:281-90.
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;7:403-22.
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;12:131-40.
18. Li Y, Willer CJ, Ding J, et al. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet Epidemiol 2010;34:816-34.
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;86:560-72.
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 2011:265-75.
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;55:831-5.
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;39:995-9.
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;11:6800-6.

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;63:2121-8.
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. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic acids research 2012;40:D930-4.
29. Goring HH, Curran JE, Johnson MP, et al. Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. Nature genetics 2007;39:1208-16.
30. Idaghdour Y, Czika W, Shianna KV, et al. Geographical genomics of human leukocyte gene expression variation in southern Morocco. Nature genetics 2010;42:62-7.
31. Heap GA, Trynka G, Jansen RC, et al. Complex nature of SNP genotype effects on gene expression in primary human leucocytes. BMC medical genomics 2009;2:1.
32. Emilsson V, Thorleifsson G, Zhang B, et al. Genetics of gene expression and its effect on disease. Nature 2008;452:423-8.
33. 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 genetics 2011;7:e1002197.
34. **Mehta D, Heim K**, Herder C, et al. Impact of common regulatory single-nucleotide variants on gene expression profiles in whole blood. European journal of human genetics : EJHG 2013;21:48-54.
35. 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 genetics 2013;9:e1003594.
36. 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.
37. 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;102:278-87.
38. **Westra HJ, Peters MJ, Esko T,** et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nature genetics 2013;45:1238-43.
39. **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.
40. The Genotype-Tissue Expression (GTEx) project. Nature genetics 2013;45:580-5.
41. Battle A, Mostafavi S, Zhu X, et al. Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. Genome research 2014;24:14-24.

42. Benton MC, Lea RA, Macartney-Coxson D, et al. Mapping eQTLs in the Norfolk Island genetic isolate identifies candidate genes for CVD risk traits. American journal of human genetics 2013;93:1087-99.
43. 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.
44. **Quinlan J, Idaghdour Y**, Goulet JP, et al. Genomic architecture of sickle cell disease in West African children. Frontiers in genetics 2014;5:26.
45. **Wright FA, Sullivan PF,** Brooks AI, et al. Heritability and genomics of gene expression in peripheral blood. Nature genetics 2014;46:430-7.
46. **Schramm K, Marzi C, Schurmann C,** et al. Mapping the genetic architecture of gene regulation in whole blood. PloS one 2014;9:e93844.
47. **Dixon AL, Liang L, Moffatt MF**, et al. A genome-wide association study of global gene expression. Nature genetics 2007;39:1202-7.
48. 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 research 2013;23:716-26.
49. Stranger BE, Nica AC, Forrest MS, et al. Population genomics of human gene expression. Nature genetics 2007;39:1217-24.
50. Kwan T, Benovoy D, Dias C, et al. Genome-wide analysis of transcript isoform variation in humans. Nature genetics 2008;40:225-31.
51. **Dimas AS, Deutsch S**, **Stranger BE,** et al. Common regulatory variation impacts gene expression in a cell type-dependent manner. Science 2009;325:1246-50.
52. 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. Human molecular genetics 2012;21:2111-23.
53. **Grundberg E**, Small KS, Hedman AK, et al. Mapping cis- and trans-regulatory effects across multiple tissues in twins. Nature genetics 2012;44:1084-9.
54. 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.
55. **Mangravite LM, Engelhardt BE**, Medina MW, et al. A statin-dependent QTL for GATM expression is associated with statin-induced myopathy. Nature 2013;502:377-80.
56. Bryois J, Buil A, Evans DM, et al. Cis and trans effects of human genomic variants on gene expression. PLoS genetics 2014;10:e1004461.
57. 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. Nature genetics 2012;44:502-10.
58. Murphy A, Chu JH, Xu M, et al. Mapping of numerous disease-associated expression polymorphisms in primary peripheral blood CD4+ lymphocytes. Human molecular genetics 2010;19:4745-57.
59. **Heinzen EL, Ge D, Cronin KD**, et al. Tissue-specific genetic control of splicing: implications for the study of complex traits. PLoS biology 2008;6:e1.

60. Zeller T, Wild P, Szymczak S, et al. Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. PloS one 2010;5:e10693.
61. 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.
62. **Barreiro LB, Tailleux L**, Pai AA, et al. Deciphering the genetic architecture of variation in the immune response to Mycobacterium tuberculosis infection. Proceedings of the National Academy of Sciences of the United States of America 2012;109:1204-9.
63. **Lee MN, Ye C**, Villani AC, et al. Common genetic variants modulate pathogen-sensing responses in human dendritic cells. Science 2014;343:1246980.
64. Huang RS, Gamazon ER, Ziliak D, et al. Population differences in microRNA expression and biological implications. RNA biology 2011;8:692-701.
65. **Degner JF, Pai AA, Pique-Regi R**, et al. DNase I sensitivity QTLs are a major determinant of human expression variation. Nature 2012;482:390-4.
66. 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 research 2011;21:1008-16.
67. Kompass KS, Witte JS. Co-regulatory expression quantitative trait loci mapping: method and application to endometrial cancer. BMC medical genomics 2011;4:6.
68. Li Q, Seo JH, Stranger B, et al. Integrative eQTL-based analyses reveal the biology of breast cancer risk loci. Cell 2013;152:633-41.
69. **Schadt EE, Molony C, Chudin E**, et al. Mapping the genetic architecture of gene expression in human liver. PLoS biology 2008;6:e107.
70. **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 genetics 2011;7:e1002078.
71. 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. The pharmacogenomics journal 2013;13:12-20.
72. Wang X, Tang H, Teng M, et al. Mapping of hepatic expression quantitative trait loci (eQTLs) in a Han Chinese population. Journal of medical genetics 2014;51:319-26.
73. **Grundberg E, Kwan T**, Ge B, et al. Population genomics in a disease targeted primary cell model. Genome research 2009;19:1942-52.
74. Kabakchiev B, Silverberg MS. Expression quantitative trait loci analysis identifies associations between genotype and gene expression in human intestine. Gastroenterology 2013;144:1488-96, 1496 e1-3.
75. Ongen H, Andersen CL, Bramsen JB, et al. Putative cis-regulatory drivers in colorectal cancer. Nature 2014;512:87-90.
76. 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;63:1154-65.

77. Quigley DA, Fiorito E, Nord S, et al. The 5p12 breast cancer susceptibility locus affects MRPS30 expression in estrogen-receptor positive tumors. Molecular oncology 2014;8:273-84.
78. **Hao K, Bosse Y, Nickle DC**, et al. Lung eQTLs to help reveal the molecular underpinnings of asthma. PLoS genetics 2012;8:e1003029.
79. 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;30:369-76.
80. **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. American journal of human genetics 2010;87:779-89.
81. Wagner JR, Busche S, Ge B, et al. The relationship between DNA methylation, genetic and expression inter-individual variation in untransformed human fibroblasts. Genome biology 2014;15:R37.
82. Qiu W, Cho MH, Riley JH, et al. Genetics of sputum gene expression in chronic obstructive pulmonary disease. PloS one 2011;6:e24395.
83. Fadista J, Vikman P, Laakso EO, et al. Global genomic and transcriptomic analysis of human pancreatic islets reveals novel genes influencing glucose metabolism. Proceedings of the National Academy of Sciences of the United States of America 2014;111:13924-9.
84. **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.
85. Lin H, Dolmatova EV, Morley MP, et al. Gene expression and genetic variation in human atria. Heart rhythm : the official journal of the Heart Rhythm Society 2014;11:266-71.
86. 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.
87. Gamazon ER, Innocenti F, Wei R, et al. A genome-wide integrative study of microRNAs in human liver. BMC genomics 2013;14:395.
88. Li Q, Stram A, Chen C, et al. Expression QTL-based analyses reveal candidate causal genes and loci across five tumor types. Human molecular genetics 2014;23:5294-302.
89. **Webster JA, Gibbs JR**, Clarke J, et al. Genetic control of human brain transcript expression in Alzheimer disease. American journal of human genetics 2009;84:445-58.
90. Zou F, Chai HS, Younkin CS, et al. Brain expression genome-wide association study (eGWAS) identifies human disease-associated variants. PLoS genetics 2012;8:e1002707.
91. **Ramasamy A, Trabzuni D, Guelfi S**, et al. Genetic variability in the regulation of gene expression in ten regions of the human brain. Nature neuroscience 2014;17:1418-28.
92. **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 genetics 2010;6:e1000952.
93. 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. Molecular psychiatry 2013;18:340-6.

94. Kim S, Cho H, Lee D, et al. Association between SNPs and gene expression in multiple regions of the human brain. Translational psychiatry 2012;2:e113.
95. **Zhang B, Gaiteri C, Bodea LG**, et al. Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. Cell 2013;153:707-20.
96. Shpak M, Hall AW, Goldberg MM, et al. An eQTL analysis of the human glioblastoma multiforme genome. Genomics 2014;103:252-63.
97. **Colantuoni C, Lipska BK**, Ye T, et al. Temporal dynamics and genetic control of transcription in the human prefrontal cortex. Nature 2011;478:519-23.
98. Liu C, Cheng L, Badner JA, et al. Whole-genome association mapping of gene expression in the human prefrontal cortex. Molecular psychiatry 2010;15:779-84.

Author names in bold designate shared co-first authors.

Supplementary Tables

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 due 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 ≥2 ICD-9 codes 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.
SPC (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 due to missing ultrasound data or cholecystectomy scar).
All discovery samples	
Replication studies	
CCHS and CGPS	Diagnoses of symptomatic gallstone disease (ICD8: 574-575; ICD10: 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 had all undergone cholecystectomy for cholecystolithiasis. German controls were all 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.

Supplementary Table 2. Annotation of nominally significant ($P < 5x10^{-6}$) GWAS SNPs after conditional analysis using GCTA.

Chr	SNP	HG38 location	Ref Allele	Freq Ref	Beta	*P*-meta-analysis	*P*-conditional	Gene	Annotation
2	rs1260326	2:27508073	T	0.412	-0.113	$2.55 X10^{-10}$	$2.65 X10^{-10}$	*GCKR*	Missense variant; splice region variant
2	rs1025447	2:43795831	T	0.831	0.165	$4.21 X10^{-12}$	$6.14 X10^{-10}$	*DYNC2LI1*	intron variant
2	rs11887534	2:43839108	C	0.066	0.527	$2.44 X10^{-60}$	$2.01 X10^{-47}$	*ABCG8*	missense variant
2	rs4299376	2:43845437	T	0.685	0.237	$1.18 X10^{-34}$	$3.39 X10^{-21}$	*ABCG8*	intron variant
3	rs4234161	3:72266437	C	0.762	0.099	$4.44 X10^{-06}$	$4.66 X10^{-06}$	-	intergenic variant
3	rs9843304	3:149493600	T	0.547	-0.113	$6.09 X10^{-11}$	$5.54 X10^{-11}$	*TM4SF4*	intron variant
6	rs6927914	6:60664964	T	0.223	-0.041	0.05689	$3.91 X10^{-06}$	-	intergenic variant
6	rs6904350	6:60992067	C	0.392	0.084	$2.34 X10^{-06}$	$1.46 X10^{-13}$	-	intergenic variant
6	rs1577631	6:61237979	A	0.611	0.043	0.01584	$3.03 X10^{-26}$	-	intergenic variant
6	rs1855933	6:61571349	A	0.610	0.040	0.02242	$1.03 X10^{-47}$	*MTRNR2L9*	upstream gene variant
8	rs6471717	8:58464798	A	0.655	-0.108	$8.84 X10^{-09}$	$9.30 X10^{-09}$	-	intergenic variant
11	rs1462565	11:23502288	A	0.015	0.327	$4.10 X10^{-06}$	$4.06 X10^{-06}$	-	intergenic variant
12	rs11061712	12:1367741	A	0.422	-0.082	$2.74 X10^{-06}$	$2.80 X10^{-06}$	*ERC1*	intron variant
12	rs2277368	12:53714444	C	0.313	0.096	$6.56 X10^{-07}$	$6.60 X10^{-07}$	*CALCOCO1*	intron variant
16	rs11644920	16:11551157	A	0.687	-0.097	$1.80 X10^{-07}$	$1.90 X10^{-07}$	*LITAF*	intron variant
16	rs2216730	16:78799555	T	0.840	-0.118	$1.56 X10^{-06}$	$1.53 X10^{-06}$	*WWOX*	intron variant
18	rs12605943	18:48137422	A	0.366	-0.101	$4.89 X10^{-06}$	$5.12 X10^{-06}$	*ZBTB7C*	upstream gene variant
19	rs296391	19:47865277	T	0.844	0.168	$1.59 X10^{-10}$	$1.54 X10^{-10}$	-	intergenic variant
21	rs9979307	21:35635881	A	0.872	-0.134	$6.26 X10^{-07}$	$6.11 X10^{-07}$	-	intergenic variant

Annotations were obtained from UCSC variant annotation integrator genome.ucsc.edu
Nominally significant = 10Mb windows around SNPs with $P<5x10^{-6}$

Supplementary Table 3a. Imputation quality scores[a] 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 Affy
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

[a]Imputation quality scores were obtained using MaCH software in ARIC, Rotterdam, Framingham, WGHS, SHIP, SHIP-TREND and NHS/HPFS studies. Imputation quality scores in SPC2 and BioVU were obtained using IMPUTEv2.

Supplementary Table 3b. 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]OR = odds ratios. Odds ratio were adjusted for age, gender and BMI in each discovery study and for study specific additional covariates.

[b]Proxy SNP for rs4299376 (R^2 = 0.995, D' = 0.999 among 1,753 NHS participants)

[c]Proxy SNP for rs296391 (R^2 = 0.904, D' = 0.969 among 1,753 NHS participants)

Supplementary Table 4. Discriminative accuracy of genetic risk score in the replication datasets.

	Odds ratio, per 1 standard deviation 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

Footnote: Odds ratio estimates were adjusted for age, sex, and BMI at blood draw. GRS, genetic risk score; AUC, area under curve.

Supplementary Table 5. *Post hoc* analysis in 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)	0.0552	7.7	683/2461	1.28(0.94,1.73)	0.119	7.28	217/1560	1.35(0.77,2.35)	0.29	8.76
	AA	2506/10030	1.46(1.12,1.89)	0.00451	24.2	1913/6063	1.45(1.08,1.96)	0.0144	23.7	593/3967	1.47(0.85,2.52)	0.165	24.7
	CC(ref) vs CA/AA		1.41(1.09,1.83)	0.00905			1.4(1.04,1.89)	0.0263			1.43(0.84,2.46)	0.19	
	CC/CA(ref) vs AA		1.15(1.06,1.25)	0.00121			1.16(1.05,1.28)	0.00234			1.12(0.94,1.32)	0.212	
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)	0.71	0.966	1256/4283	1.09(0.96,1.23)	0.197	4.21	374/2774	0.81(0.66,1)	0.0454	NA
	CC	1260/4813	1.17(1.05,1.3)	0.00599	5.35	980/2951	1.23(1.08,1.4)	0.00196	7.18	280/1862	0.95(0.76,1.17)	0.616	NA
	TT(ref) vs TC/CC		1.08(0.98,1.19)	0.132			1.14(1.02,1.29)	0.0257			0.86(0.72,1.05)	0.134	
	TT/TC(ref) vs CC		1.15(1.06,1.24)	0.000454			1.16(1.06,1.27)	0.00175			1.1(0.93,1.29)	0.273	
rs11887534	GG	2827/12824	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)	0.00000254	7.2
	CC	33/63	2.45(1.6,3.76)	0.0000385	0.628	23/39	2.12(1.26,3.56)	0.00442	0.495	10/12	3.43(1.59,7.4)	0.0017	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)	0.000000214	
	GG/GC(ref) vs CC		2.26(1.48,3.46)	0.00018			1.94(1.16,3.26)	0.012			3.19(1.48,6.87)	0.00311	
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)	0.0794	5.49	1086/3956	1.15(0.98,1.35)	0.0819	6.33	359/2511	1.18(0.88,1.57)	0.267	7.36
	TT	1728/6481	1.35(1.18,1.55)	0.0000144	13.6	1331/3847	1.45(1.24,1.69)	0.00000326	16.5	397/2634	1.24(0.93,1.65)	0.146	10
	CC(ref) vs CT/TT		1.24(1.09,1.41)	0.00122			1.3(1.12,1.51)	0.000674			1.21(0.92,1.59)	0.18	
	CC/CT(ref) vs TT		1.22(1.13,1.32)	0.000000171			1.29(1.18,1.41)	$7.88x10^{-9}$			1.08(0.93,1.26)	0.327	
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)	0.00603	5.97	1305/4325	1.11(1,1.22)	0.0565	5.14	412/2736	1.24(1.03,1.49)	0.0265	10.3
	CC	827/3022	1.28(1.16,1.43)	0.00000349	5.53	614/1765	1.27(1.13,1.44)	0.00011	5.15	213/1257	1.42(1.15,1.76)	0.00138	8.49
	TT(ref) vs TC/CC		1.18(1.08,1.28)	0.000127			1.15(1.05,1.27)	0.00361			1.3(1.09,1.55)	0.00406	

	TT/TC(ref) vs CC		1.19(1.09,1.3)	0.00016			1.2(1.08,1.33)	0.000723			1.24(1.04,1.48)	0.0168	
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)	0.0984	3.01	1178/3864	1.06(0.97,1.16)	0.217	2.57	377/2553	1.14(0.96,1.34)	0.129	5.91
	GG	474/1715	1.22(1.08,1.37)	0.00118	2.54	363/1063	1.18(1.03,1.35)	0.018	2.13	111/652	1.32(1.03,1.68)	0.0261	3.54
	AA(ref) vs AG/GG		1.1(1.02,1.19)	0.013			1.09(0.99,1.19)	0.0649			1.17(1,1.37)	0.0455	
	AA/AG(ref) vs GG		1.17(1.05,1.31)	0.00436			1.14(1.01,1.3)	0.0387			1.23(0.98,1.55)	0.0699	

Supplementary Table 6. *Post hoc* analysis in NHS and HPFS cohorts: Haplotype Analyses at the ABCG5/8 locus in relation to gallstone disease risk.

rs11887534(G/C*) -rs4245791(C/T*)	All samples (Males and females)				Females only				Males only			
Haplotype combinations	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)	0.369	2.79	941/3626	1.08(0.92,1.27)	0.327	3.2	314/2306	1.11(0.83,1.49)	0.466	4.27
G-T/ G-T	1270/5382	1.19(1.03,1.37)	0.0148	6.61	986/3254	1.27(1.08,1.48)	0.00394	9.1	284/2128	1.08(0.81,1.45)	0.597	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)	0.000000867	3.06	45/205	1.83(1.19,2.82)	0.00611	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)	0.00114	6.2
C-T/ C-T	32/63	2.64(1.69,4.13)	0.000021	0.71	22/39	2.33(1.36,4.01)	0.00221	0.588	10/24	3.73(1.66,8.38)	0.00142	1.14

Supplementary Table 7. *Post hoc* analysis in NHS and HPFS cohorts: Gene-gene interactions (GxG) between GWAS significant SNPs.

GxG *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	-	0.041	0.899	0.406	0.103
	rs4245791	0.625	-	0.808	0.781	0.848
	rs9843304	0.448	0.305	-	0.323	0.927
	rs6471717	0.727	0.883	0.737	-	0.560
	rs2547231	0.414	0.526	0.831	0.058	-

Supplementary Table 8. *Post hoc* analysis in 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 post-menopausal 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 9. *Post hoc* analysis in 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)	0.577	0.873	1157/ 3887	1.01 (0.92,1.1)	0.902	0.441	364/2485	1.08 (0.91,1.27)	0.368	3.38
	TT	417/ 1478	1.22 (1.08,1.38)	0.00145	2.2	307/ 899	1.16 (1,1.34)	0.0498	1.61	110/579	1.45 (1.14,1.85)	0.00284	4.38
	GG(ref) vs GT/TT		1.06 (0.98,1.14)	0.128			1.03 (0.95,1.13)	0.453			1.15 (0.98,1.34)	0.0813	
	GG/GT(ref) vs TT		1.21 (1.08,1.36)	0.0015			1.15 (1.00,1.32)	0.0428			1.39 (1.11,1.75)	0.00433	

Supplementary Table 10. Concordant *cis*-eQTLs at gallstone GWAS susceptibility loci.

All eQTL results (P<1.0E-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 where all three pairwise relationship with r^2>0.8 in HapMap CEU populations as defined by querying SNAP (http://www.broadinstitute.org/mpg/snap/).

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
			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

Supplementary Table 11. Regulatory annotations for gallstone SNPs with eQTL associations.

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).

rsID	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 HepG2,MAFK,Stanford	GATA_known10;GATA_known9;HDAC2_disc1;HDAC2_disc6;Smad3_2
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,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

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

Supplementary Table 12. Results of querying gallstone SNPs and proxies (r^2>0.8) in the GRASP GWAS database v. 2.0.

GallstoneProxy	GallstoneIndex	Distance	r^2	D'	Chr	Gallstone candidate gene	Pubmed ID	Results location	GWAS p-value	Trait
rs9921290	rs1260326	0	1	1	2	GCKR	20686565	Table S19	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	Table S19	2.30E-49	LDL cholesterol
rs9476368	rs4299376	0	1	1	2	ABCG8	20686565	Table S19	3.20E-47	Total cholesterol
rs9448882	rs1260326	0	1	1	2	GCKR	23263486	Table6Supp	1.25E-44	Serum urate
rs9446581	rs1260326	0	1	1	2	GCKR	23263486	Table15Supp	3.80E-43	C-reactive protein (CRP)
rs9446578	rs1260326	0	1	1	2	GCKR	22885924	Table S3a	2.17E-41	Fasting blood glucose
rs9382866	rs1260326	0	1	1	2	GCKR	19936222	Table S1	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	Table S3	2.79E-28	VLDL cholesterol large lipoprotein fraction concentration
rs9352216	rs1260326	0	1	1	2	GCKR	20686565	Table S19	4.40E-28	Total cholesterol
rs9350568	rs1260326	10297	0.933	1	2	GCKR	20081858	Table 1	3.00E-24	HOMA-IR
rs9343302	rs1260326	0	1	1	2	GCKR	22885924	Table S2e	2.74E-22	Fasting insulin
rs9341417	rs1260326	0	1	1	2	GCKR	20081857	Table S2	2.26E-21	2 hour glucose
rs9294905	rs1260326	0	1	1	2	GCKR	23022100	Table S4	4.10E-19	Serum albumin
rs9294231	rs296381	17900	0.941	1	19	SULT2A1	21533175	Table S1	1.96E-18	Serum dehydroepiandrosterone sulphate (DHEAS)
rs9294080	rs1260326	10297	0.933	1	2	GCKR	19936222	Table S3	1.07E-17	APOB assay lipoprotein fraction concentration
rs8192870	rs1260326	11663	0.932	1	2	GCKR	22829776	Table 1	2.20E-16	Sex hormone-binding globulin (SHBG) concentrations
rs780094	rs1260326	0	1	1	2	GCKR	20383146	Table 2	3.00E-14	Serum creatinine estimated glomerular filtration rate (eGFR)
rs780094	rs6471717	65660	0.922	1	8	Intergenic, close to CYP7A1	20686565	Table S19	2.50E-13	Total cholesterol

rs780094	rs1260326	0	1	1	2	GCKR	22001757	Table 1	3.90E-13	Gamma-glutamyl transferase (GGT)
rs780094	rs1260326	10297	0.933	1	2	GCKR	19936222	Table S1	9.80E-13	LDL cholesterol mean size lipoprotein fraction concentration in fasting sample
rs780094	rs1260326	11663	0.932	1	2	GCKR	21386085	Table S6	1.90E-12	Waist circumference and Triglycerides
rs780094	rs1260326	0	1	1	2	GCKR	19060906	Table S7	8.70E-12	APOC3 (apolipoprotein C-III)
rs780094	rs1260326	10297	0.933	1	2	GCKR	21194676	Table 1	2.20E-11	Height (adults)
rs780094	rs1260326	0	1	1	2	GCKR	19936222	Table S3	2.87E-11	APOA1 assay lipoprotein fraction concentration
rs780094	rs1260326	11663	0.932	1	2	GCKR	21102463	Table 2	4.70E-11	Crohn's disease
rs780094	rs4299376	1305	1	1	2	ABCG8	23202125	Table S9	2.76E-10	Coronary artery disease (CAD)
rs780094	rs1260326	11663	0.932	1	2	GCKR	21386085	Table 2	3.00E-10	Triglycerides and Blood pressure
rs780094	rs11887534	0	1	1	2	ABCG8	19936222	Table S3	3.48E-10	APOB assay lipoprotein fraction concentration
rs780094	rs1260326	10297	0.933	1	2	GCKR	19936222	Table S3	6.98E-10	IDL total lipoprotein fraction concentration
rs780094	rs1260326	0	1	1	2	GCKR	22139419	Table 1	9.12E-10	Platelet count (PLT)
rs780094	rs1260326	11663	0.932	1	2	GCKR	23362303	Table S3	9.80E-10	Plasma palmitoleic acid
rs780094	rs1260326	10297	0.933	1	2	GCKR	20081858	Table 2	1.30E-09	Type 2 diabetes
rs780094	rs6471717	65660	0.922	1	8	Intergenic, close to CYP7A1	20686565	Table S19	1.90E-09	LDL cholesterol
rs780094	rs1260326	10297	0.933	1	2	GCKR	21829377	Text	2.52E-09	Plasma docosapentaenoic acid levels
rs780094	rs1025447	0	1	1	2	DYNC2LI1	20686565	FullScan	2.76E-09	LDL cholesterol
rs780093	rs1260326	0	1	1	2	GCKR	23118302	TableS2	9.40E-09	Lipoprotein-associated phospholipase A2 mass (Lp-PLA2)
rs780093	rs1025447	0	1	1	2	DYNC2LI1	20686565	FullScan	1.13E-08	Total cholesterol
rs780093	rs1260326	10297	0.933	1	2	GCKR	21423719	Table S5	2.59E-08	Nonalcoholic fatty liver disease

GRASP database : (http://apps.nhlbi.nih.gov/grasp/).

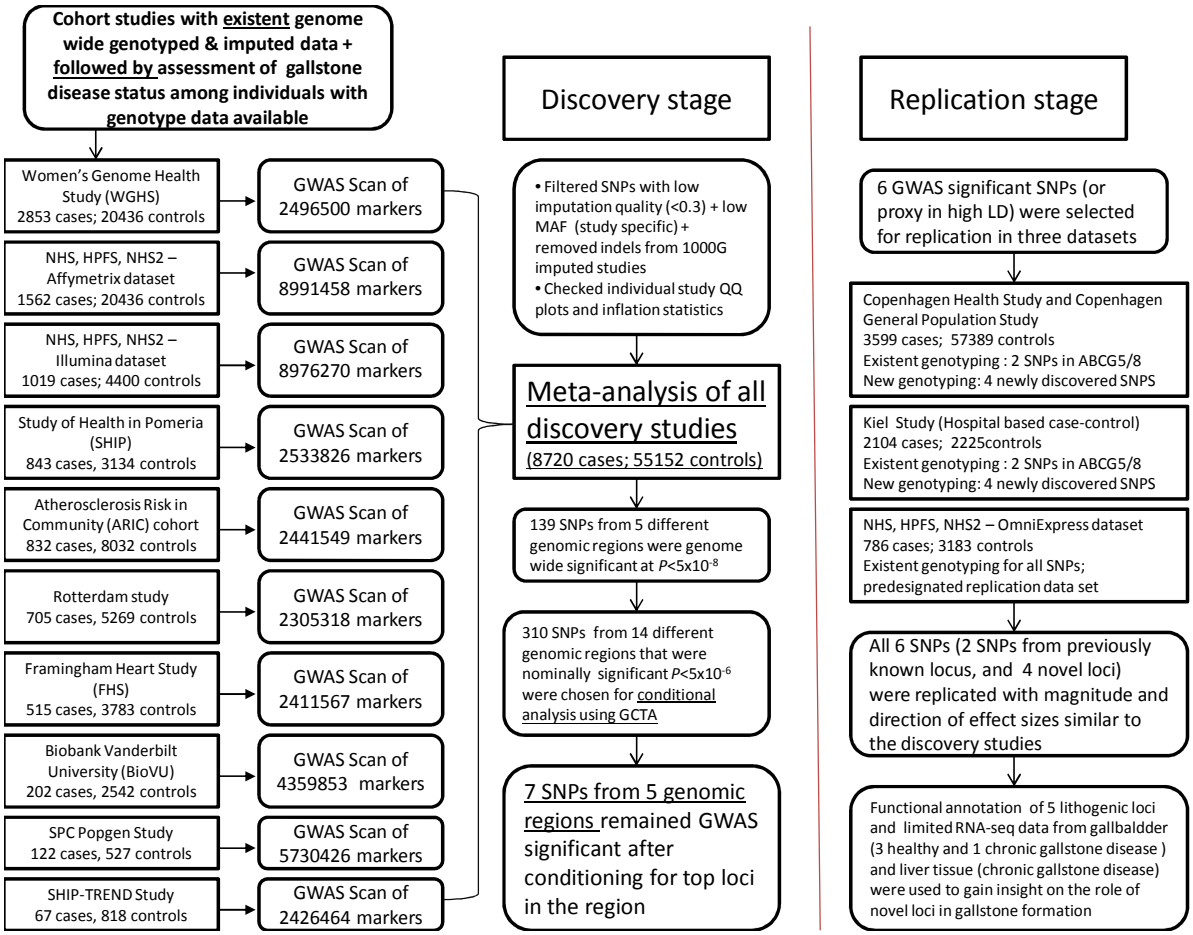
Supplementary Table 13. Results of querying gallstone SNPs and proxies (r^2>0.8) in the atlas of genetic influences on human blood metabolites.

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-3beta,17beta-diol disulfate 2	7,240	A/C	0.83	0.141 (0.005)	3.06E-191	Yes	23093944
15. GCKR (2p23.3)	rs1260326	glucose/ mannose	7,310	T/C	0.41	0.041 (0.002)	2.50E-148	Yes	23362303; 23362303; 21829377; 21886157; 22286219

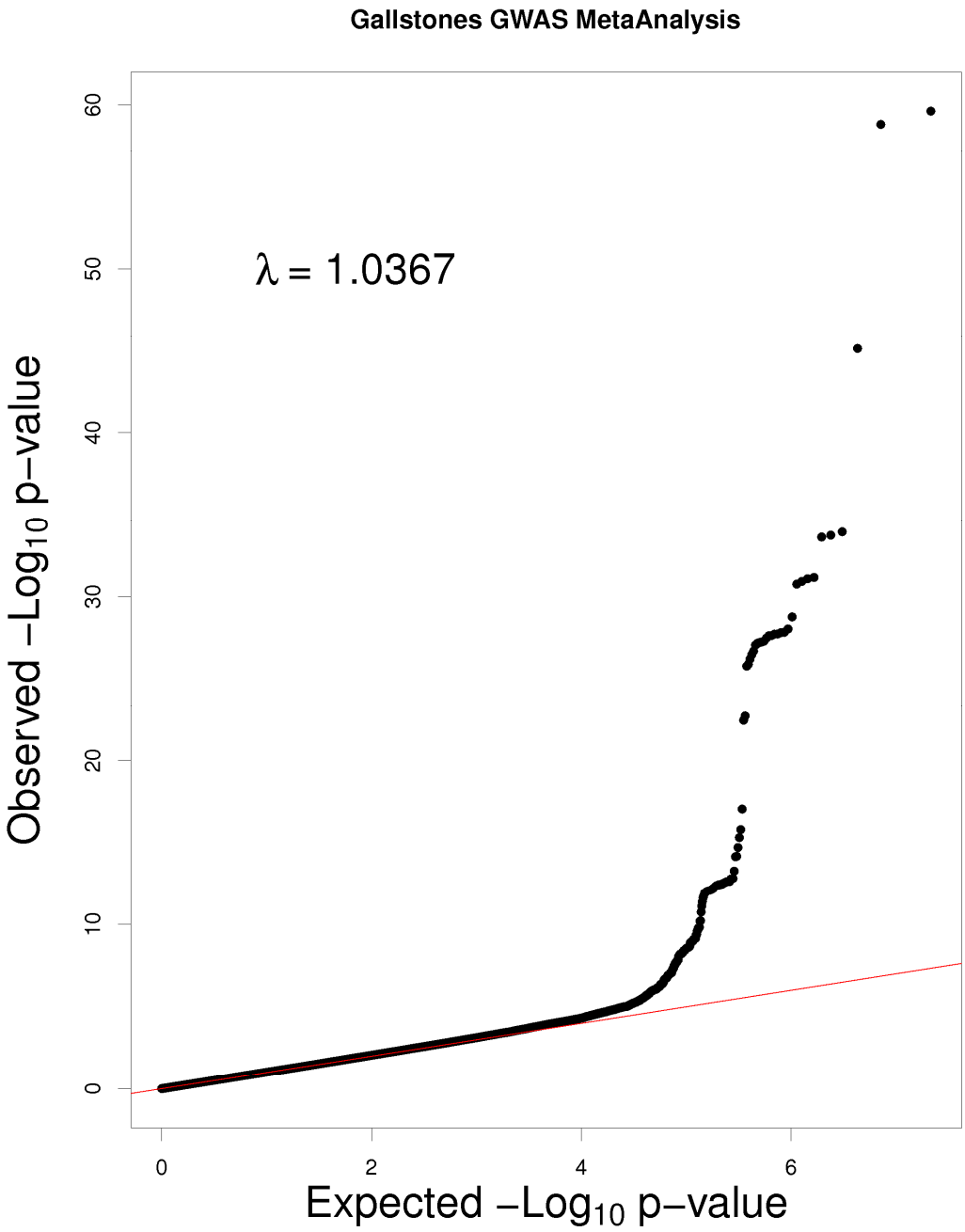
Data retrieved from: – PMID: 24816252 – An atlas of genetic influences on human blood metabolites Supplementary Table 4.

Supplementary Figures

Supplementary Figure 1 Flow chart of study cohorts and methods in the discovery and replication stages

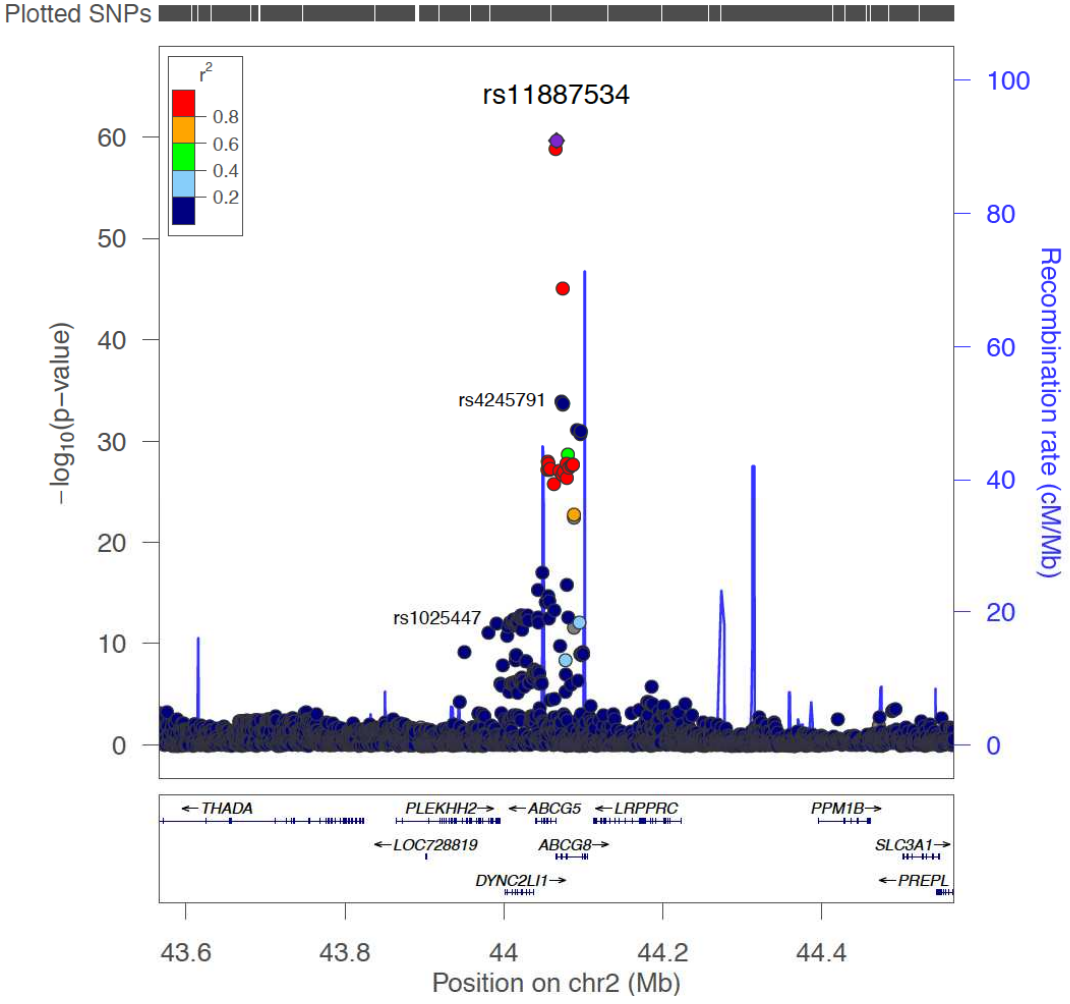


Supplementary Figure 2 Q-Q plot of gallstones disease GWAS meta-analysis.

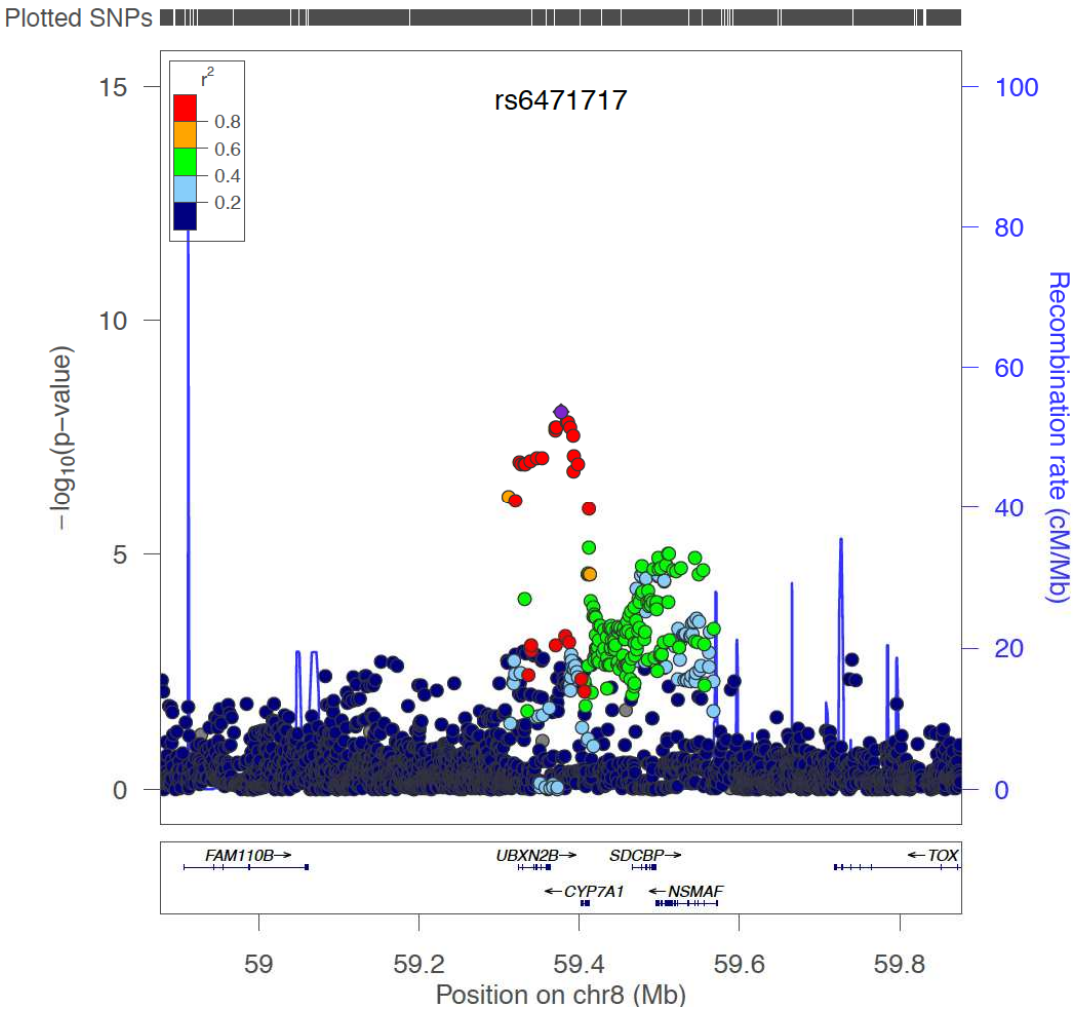


Supplementary Figure 3 Regional association plots for discovered loci in GWAS meta-analysis.

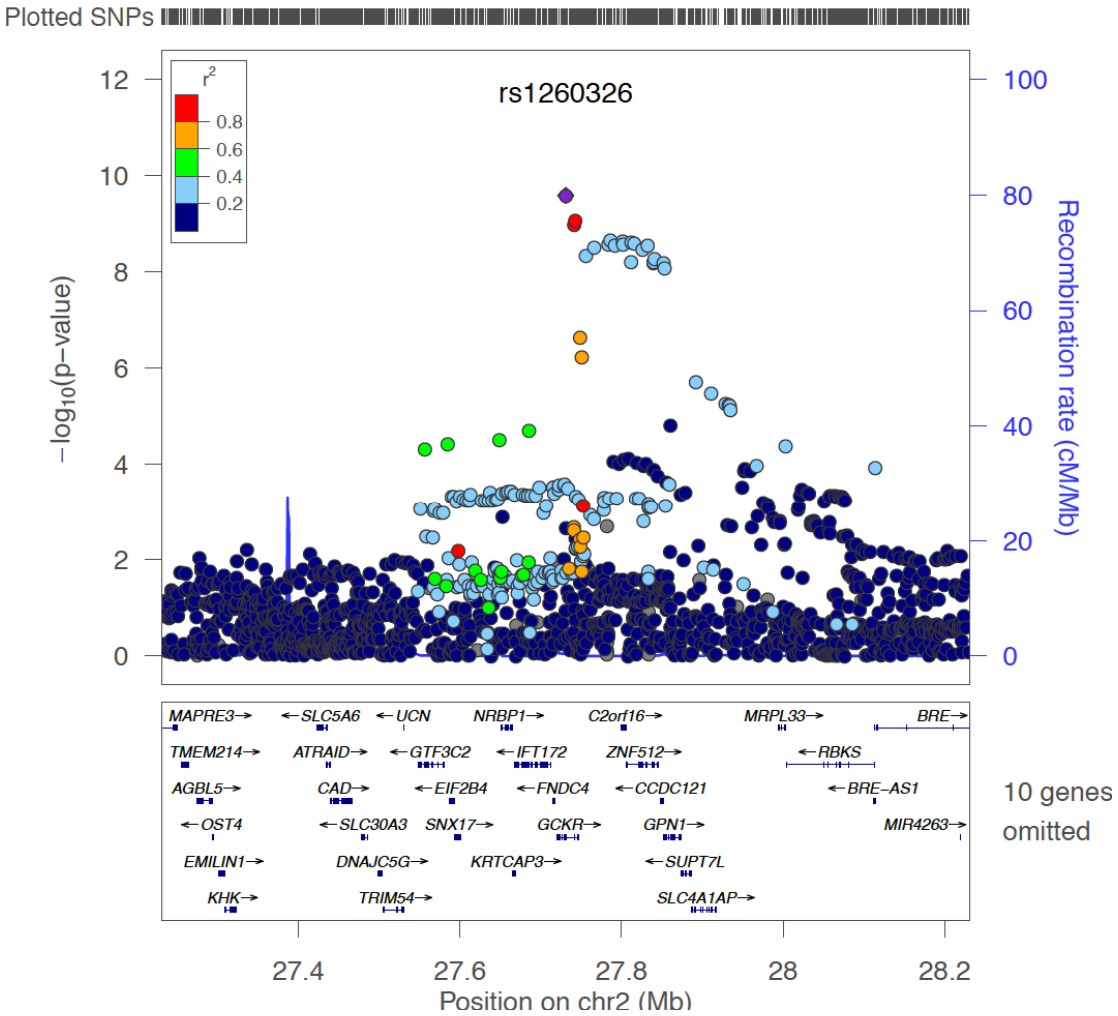
Supplementary Figure 3A – ABCG8 locus



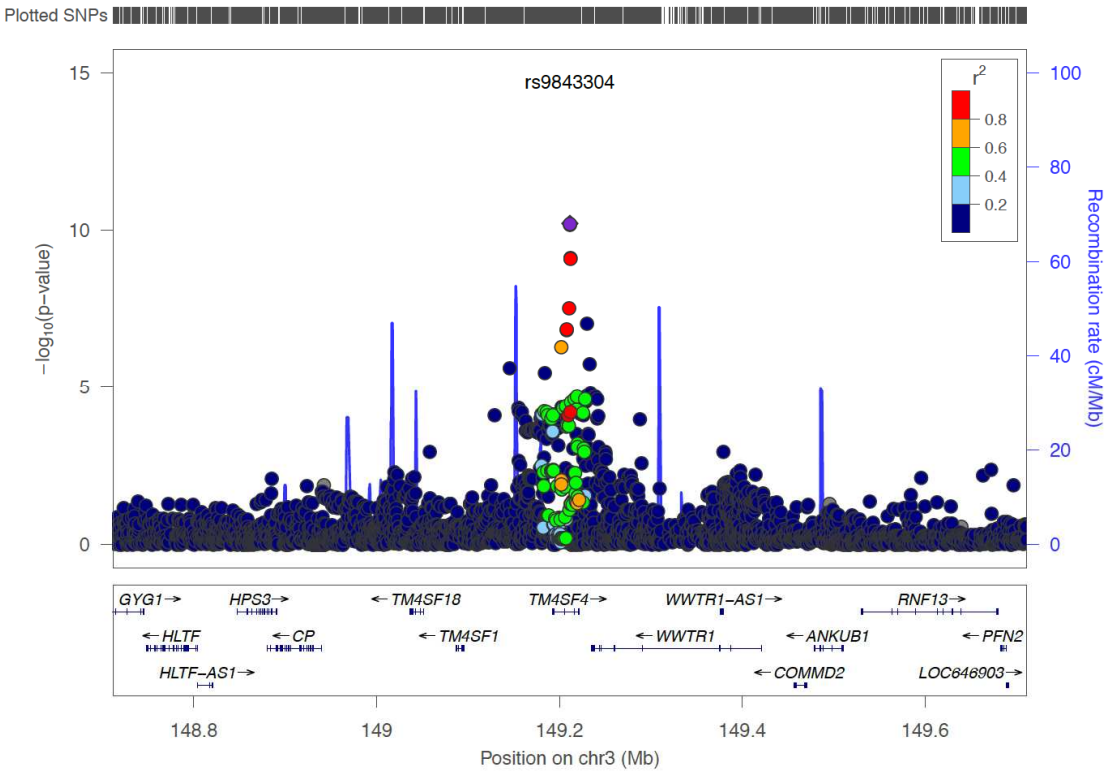
Supplementary Figure 3B – CYP7A1 locus



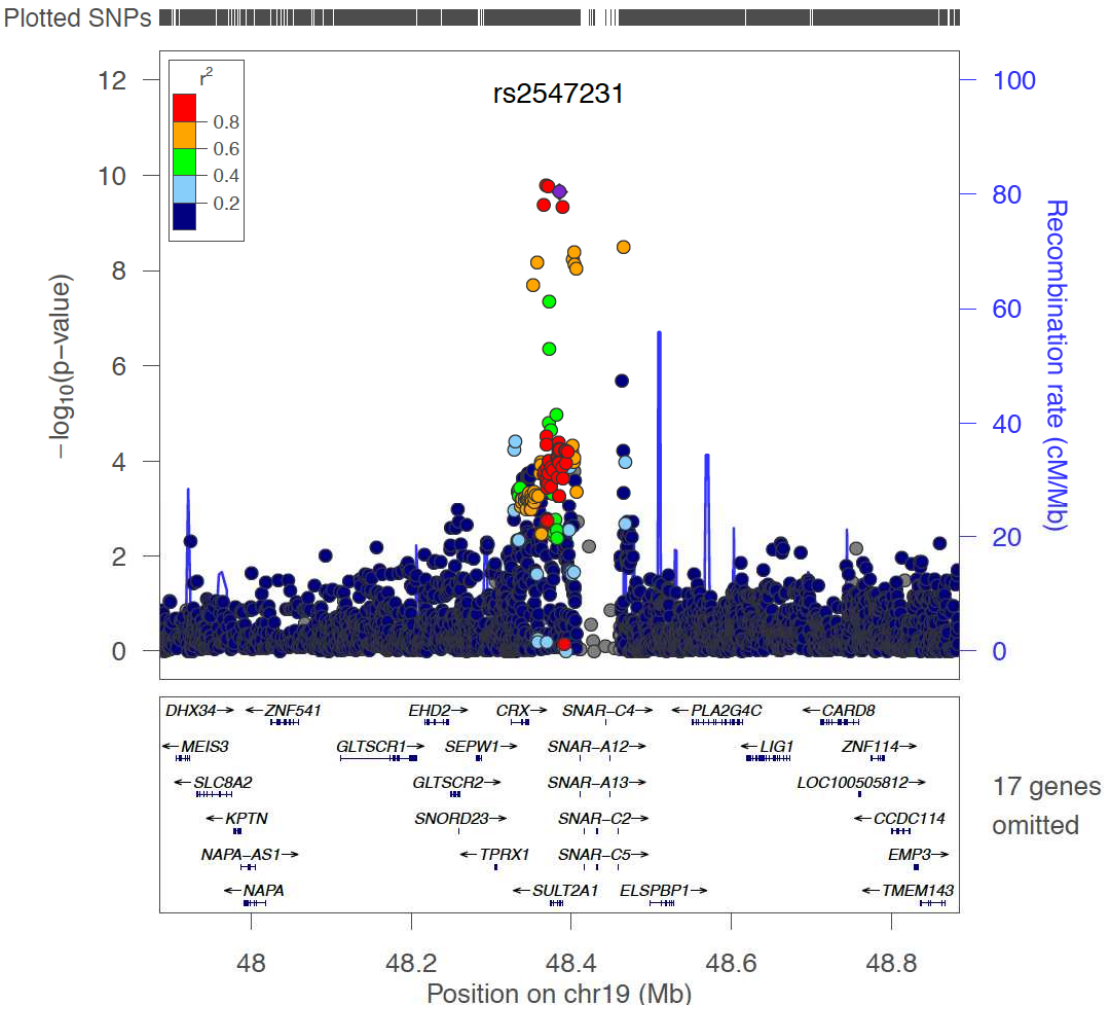
Supplementary Figure 3C – GCKR locus



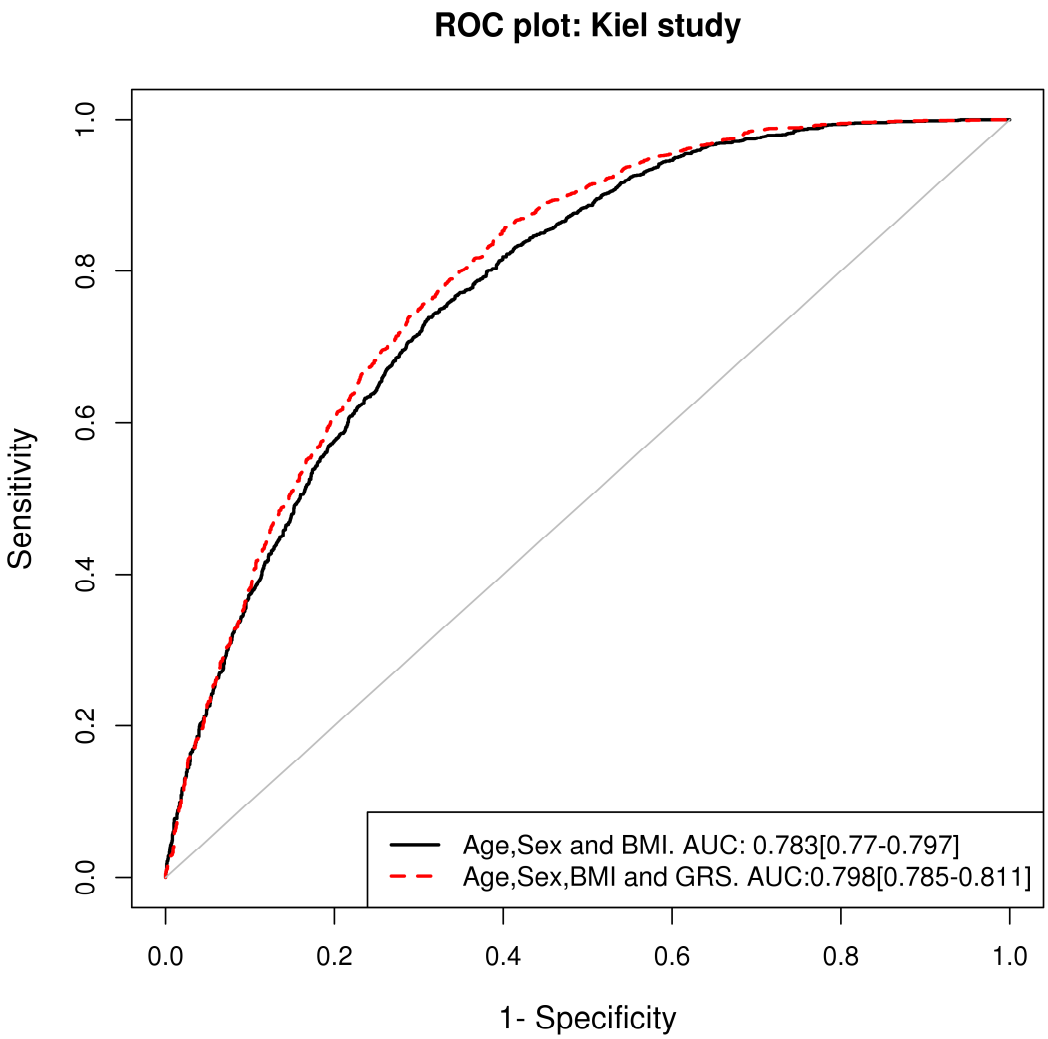
Supplementary Figure 3D – TM4SF4 locus

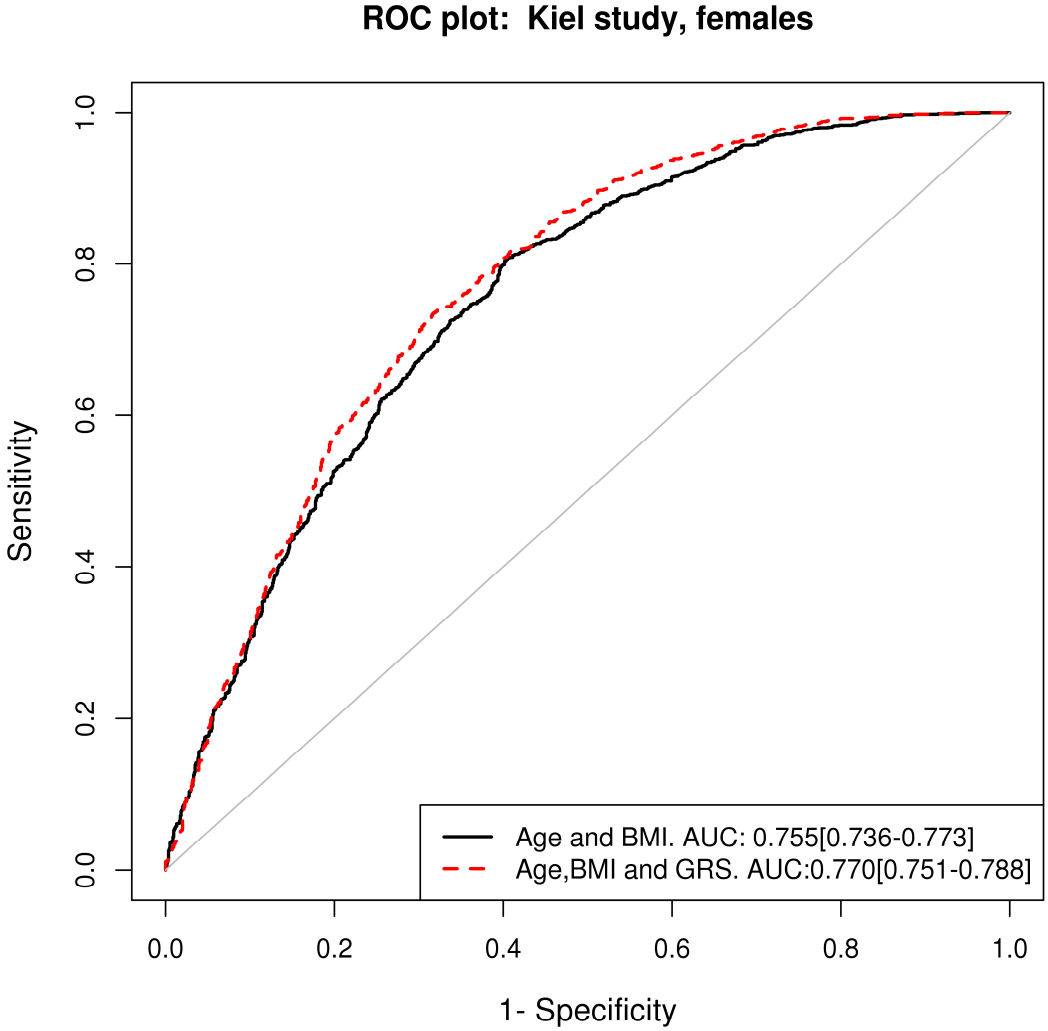


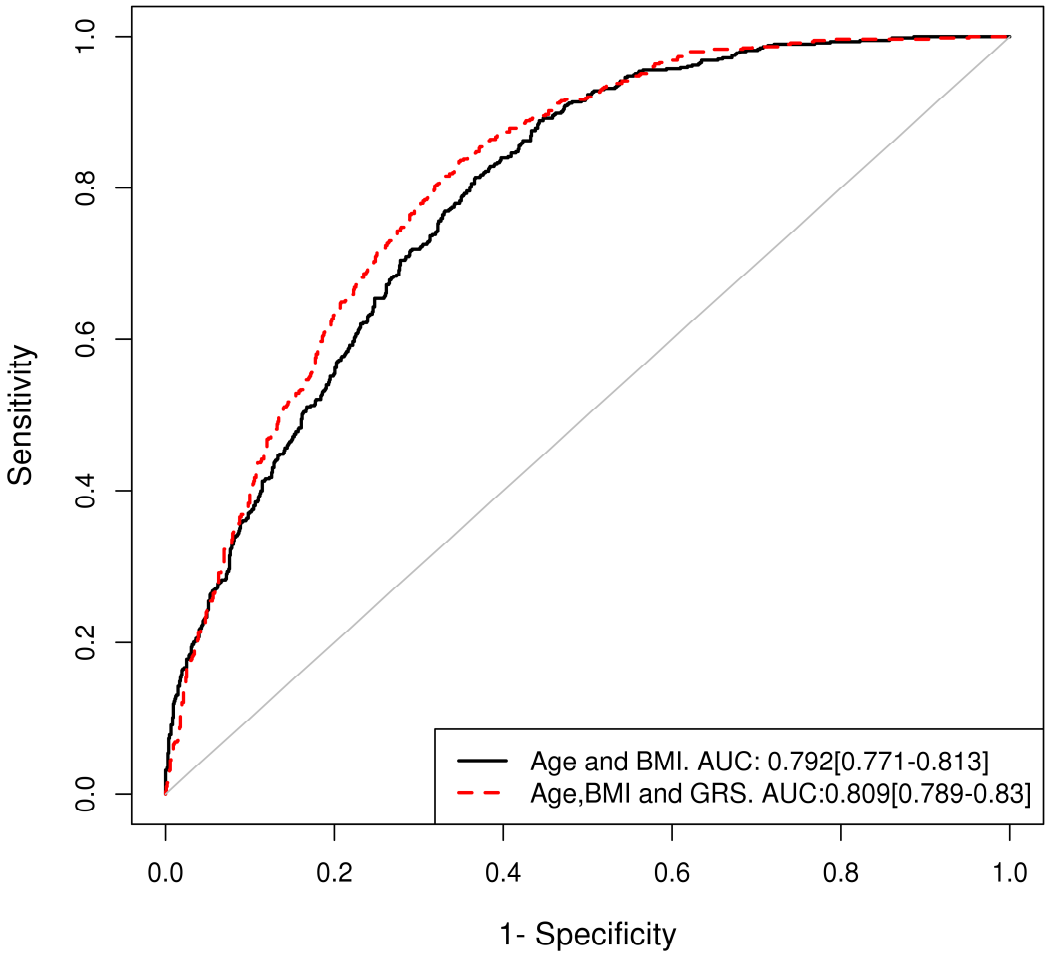
Supplementary Figure 3E – SULT2A1 locus

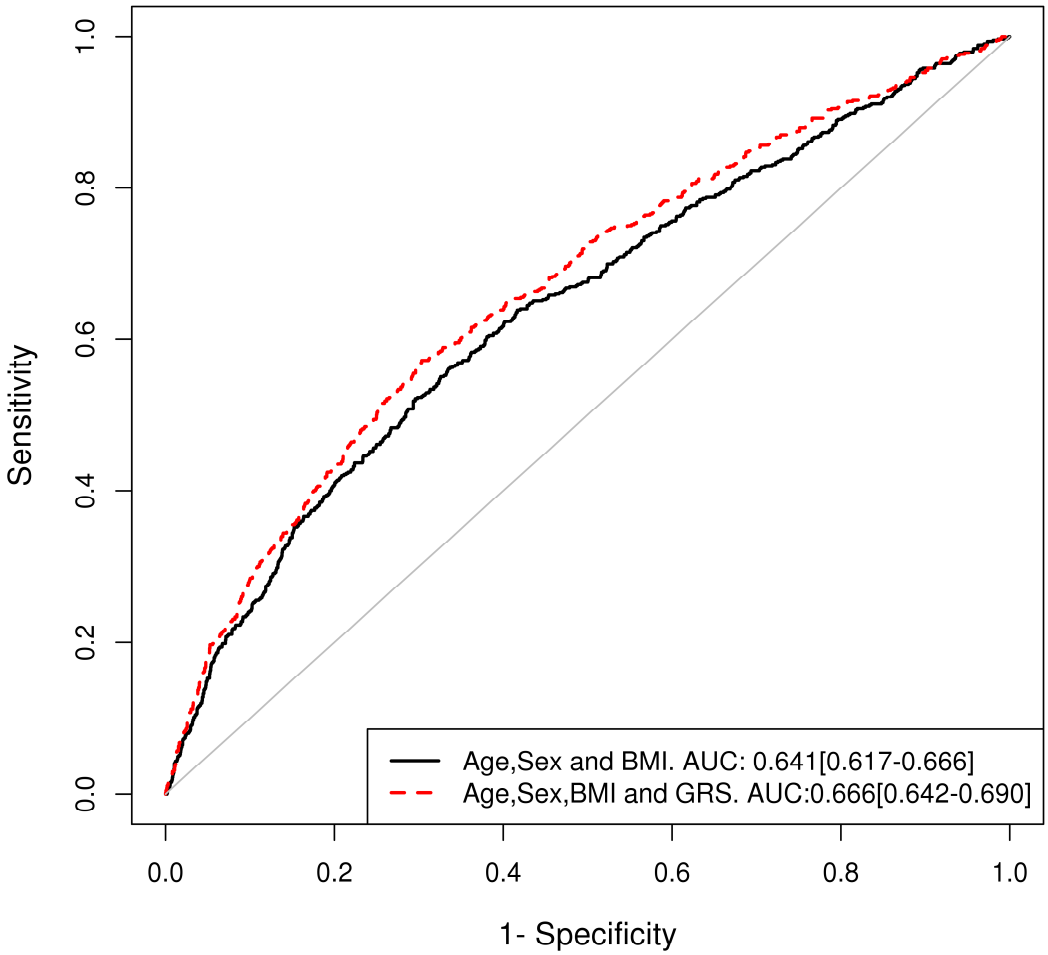


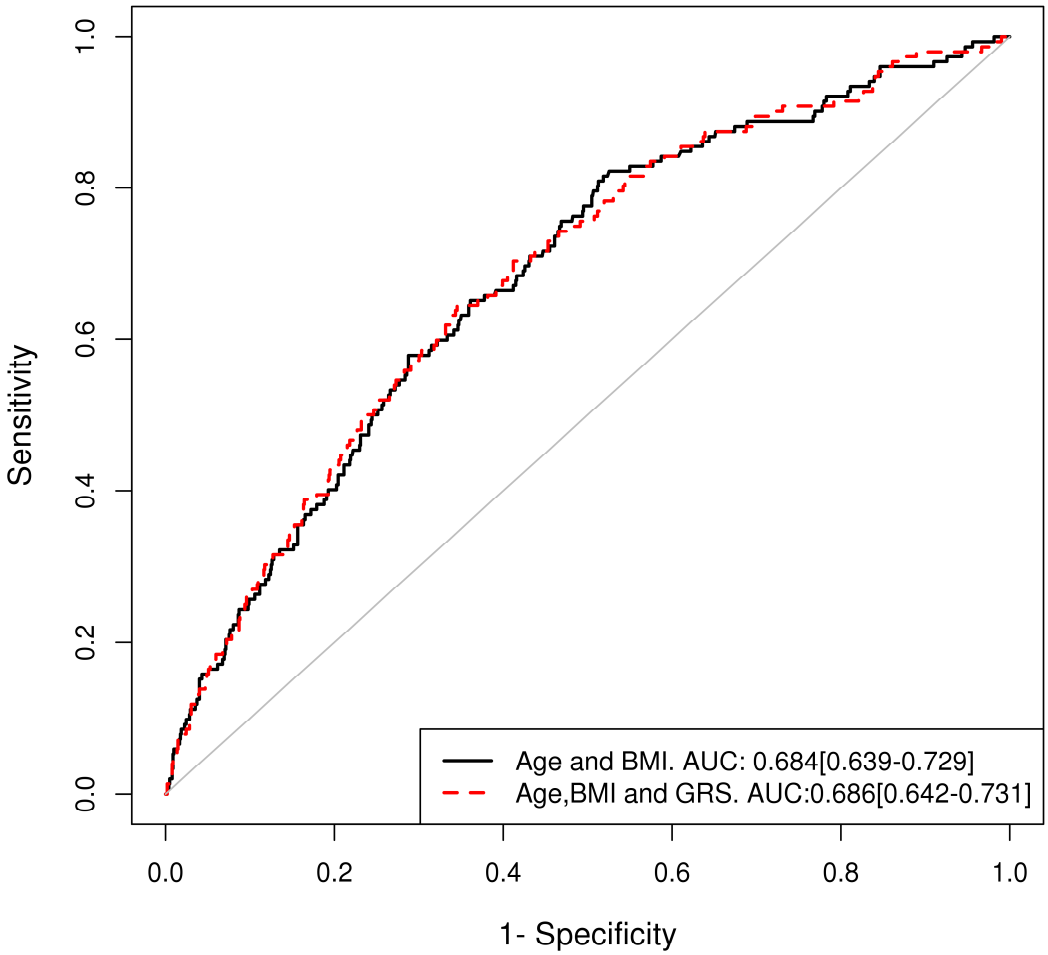
Supplementary Figure 4. Receiver operator characteristic plots in replication studies.



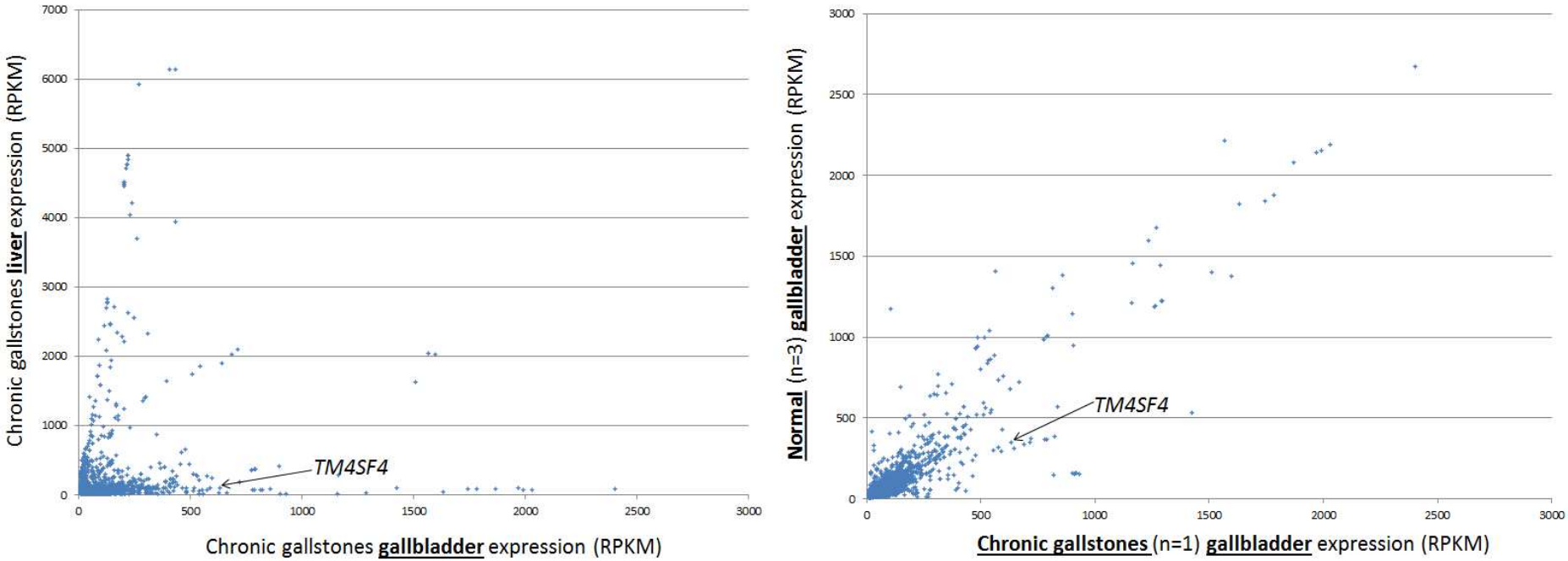

ROC plot: Kiel study, females
Sensitivity
1- Specificity
0.0
0.2
0.4
0.6
0.8
1.0
Age and BMI. AUC: 0.755[0.736-0.773]
Age,BMI and GRS. AUC:0.770[0.751-0.788]


ROC plot: Kiel study, males
Sensitivity
1- Specificity
0.0
0.2
0.4
0.6
0.8
1.0
Age and BMI. AUC: 0.792[0.771-0.813]
Age,BMI and GRS. AUC:0.809[0.789-0.83]


ROC plot : NHS replication study
Sensitivity
1- Specificity
0.0
0.2
0.4
0.6
0.8
1.0
Age,Sex and BMI. AUC: 0.641[0.617-0.666]
Age,Sex,BMI and GRS. AUC:0.666[0.642-0.690]


ROC plot : HPFS replication study
Sensitivity
1- Specificity
0.0
0.2
0.4
0.6
0.8
1.0
Age and BMI. AUC: 0.684[0.639-0.729]
Age,BMI and GRS. AUC:0.686[0.642-0.731]

Supplementary Figure 5. RNA sequencing results from gallbladder and liver from chronic gallstones case and normal gallbladders.



Comparison of RPKM values for expressed genes in chronic gallstone gallbladder versus chronic gallstone liver (left panel) and chronic gallstone gallbladder versus normal (non-gallstones) gallbladder. The point corresponding to *TM4SF4* expression is indicated. The following genes are excluded from the plots due to their high RPKM values: *MTRNR2L8, ALB,* and *APOA2*.