Supporting Information

Schumann et al. 10.1073/pnas.1017288108

SI Text

Methods for Cohorts with Categorical Data on Alcohol Intake and Alcohol-Dependency Samples. Three cohorts [DeCODE, Prevention of Renal and Vascular End-Stage Disease (PREVEND), and regional study on population from Eastern Germany of Slovanic Origin (Sorbs)] had only categorical data on alcohol intake and were analyzed separately. Alcohol intake in each category was estimated using either midpoint (DeCODE) or data derived from the Northern Finland Birth Cohort 1966 study program (NFBC1966; PREVEND and Sorbs)—alcohol intake in NFBC1966 was categorized to mimic the distribution in each study, and then, mean value for each category in NFBC1966 was used. Two alcohol dependence samples were also examined: the European Ancestry subset of the Collaborative Study on the Genetics of Alcoholism (COGA; $n = 847$ alcohol-dependent cases and $n = 552$ unrelated controls) and a clinical case-control sample from DeCODE ($n = 3{,}324$ alcohol-dependent cases and $n = 34{,}947$ population controls).

ACKNOWLEDGMENTS. Financial support for the Atherosclerosis Risk in Young Adults (ARYA) cohort was received from The Netherlands Organisation for Health Research and Development (2100.0008, 2100.0042). Genotyping for this project was funded through an Incentive Grant from the Board of the UMC Utrecht.

The Collaborative Study on the Genetics of Alcoholism (COGA), Principal Investigators B. Porjesz, V. Hesselbrock, H. Edenberg, L. Bierut, includes ten different centers: University of Connecticut (V. Hesselbrock); Indiana University (H.J. Edenberg, J. Nurnberger Jr., T. Foroud); University of Iowa (S. Kuperman, J. Kramer); SUNY Downstate (B. Porjesz); Washington University in St. Louis (L. Bierut, A. Goate, J. Rice, K. Bucholz); University of California at San Diego (M. Schuckit); Howard University (R. Taylor); Rutgers University (J. Tischfield); Southwest Foundation (L. Almasy), and Virginia Commonwealth University (D. Dick). Q. Max Guo is the NIAAA Staff Collaborator. They continue to be inspired by their memories of Henri Begleiter and Theodore Reich, founding PI and Co-PI of COGA, and also owe a debt of gratitude to other past organizers of COGA, including Ting-Kai Li, P. Michael Conneally, and Raymond Crowe, for their critical contributions. This national collaborative study is supported by NIH Grant U10AA008401 from the National Institute on Alcohol Abuse and Alcoholism (NIAAA) and the National Institute on Drug Abuse (NIDA). Funding support for GWAS genotyping, which was performed at the Johns Hopkins University Center for Inherited Disease Research, was provided by the National Institute on Alcohol Abuse and Alcoholism, the NIH GEI (U01HG004438), and the NIH contract "High throughput genotyping for studying the genetic contributions to human disease" (HHSN268200782096C). The authors thank Kim Doheny and Elizabeth Pugh from CIDR and Justin Paschall from the NCBI dbGaP staff for valuable assistance with genotyping and quality control in developing the dataset available at dbGaP.

The COLAUS study was supported by research grants from GlaxoSmithKline, from the Swiss National Science Foundation (Grant number 33CSCO-122661) and from the Faculty of Biology and Medicine of Lausanne, Switzerland. The authors would like to thank Yolande Barreau, Mathieu Firmann, Vladimir Mayor, Anne-Lise Bastian, Binasa Ramic, Martine Moranville, Martine Baumer, Marcy Sagette, Jeanne Ecoffey and Sylvie Mermoud for data collection.

The DeCODE study was supported in part by a grant from the European Commission, the FP6 Integrated Project GENADDICT (LSHM-CT-2004-005166).

The DESIR study is supported by the French Region "*Nord Pas De Calais*" ("*Contrat de Projets État-Région*") INSERM contracts with CNAMTS, Lilly, Novartis Pharma and sanofi-aventis, by INSERM (Réseaux en Santé Publique, Interactions entre les déterminants de la santé), Cohortes Santé TGIR), the Association Diabète Risque Vasculaire, the Fédération Française de Cardiologie, La Fondation de France, ALFEDIAM, ONIVINS, Ardix Medical, Bayer Diagnostics, Becton Dickinson, Cardionics, Merck Santé, Novo Nordisk, Pierre Fabre, Roche and Topcon. This study was partly supported by the European Union (Integrated Project EuroDia LSHM-CT-2006-518153 in the Framework Programme 6 [FP6] of the European Community).

The DESIR Study Group: B. Balkau, INSERM U780, Villejuif; P. Ducimetière, INSERM U780, Villejuif; E. Eschwège, INSERM U780, Villejuif; F. Alhenc-Gelas, INSERM U367, Paris; Y. Gallois, CHU Angers; A. Girault, CHU Angers; M Marre, Bichat Hospital, Paris; F. Fumeron, Bichat Hospital, Paris; F Bonnet, CHU Rennes; P. Froguel, CNRS UMR8090 Lille; J. Cogneau, Institute de Recherche Médecine Générale; C. Born, Institute inter-Regional pour la Santé La Riche; E. Caces, Institute inter-Regional pour la Santé La Riche; M. Cailleau, Institute inter-Regional pour la Santé La Riche; J. G. Moreau, Institute inter-Regional pour la Santé La Riche; O. Lantieri, Institute inter-Regional pour la Santé La Riche; F. Rakotozafy, Institute inter-Regional pour la Santé La Riche; J. Tichet, Institute inter-Regional pour la Santé La Riche; S. Vol, Institute inter-Regional pour la Santé, La Riche.

In Estonian Genome Project of University of Tartu (EGPUT), A.M. and T.E. received support from FP7 grants (201413 ENGAGE, 212111 BBMRI, ECOGENE and OPENGENE. AM and TE also received targeted financing from Estonian Government SF0180142s08 and by EU via the European Regional Development Fund, in the frame of Centre of Excellence in Genomics. The genotyping of the Estonian Genome Project samples were performed in Estonian Biocentre Genotyping Core Facility, A.M. and T.E. thank Mari Nelis and Viljo Soo for their contributions.

The EPIC Norfolk Study was funded by programme grants from the Medical Research Council UK and Cancer Research UK; and by additional support from the European Union; Stroke Association; British Heart Foundation; Department of Health; Food Standards Agency; and the Wellcome Trust.

The EPIC-Turin study was made possible by a grant from the Environmental Cancer Risk Nutrition and Individual Susceptibility (P.V., G.M.) project, a network of excellence operating within the European Union sixth Framework Program, Priority 5: `Food Quality and Safety' (Contract No.513943). A grant contribution was also given by the Compagnia di San Paolo (Turin, Italy; P.V.) and by Regione Piemonte (G.M.).

The Erasmus Rucphen Family (ERF) study was supported in part by the EU funded Adams program (242257) and by Center for Medical Systems Biology.

The Rotterdam Study (ERGO) is supported by the Erasmus Medical Center and Erasmus University Rotterdam, The Netherlands Organization for Scientific Research (NWO), The Netherlands Organization for Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), The Netherlands Genomics Initiative, the Ministry of Education, Culture and Science, the Ministry of Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. The contribution of inhabitants, general practitioners and pharmacists of the Ommoord district, the contributions of previous principal investigators, including Frank van den Ouweland (Endocrine diseases), Diederick Grobbee (Cardiovascular diseases), Paulus de Jong (Ophthalmic diseases) and Huibert Pols (Endocrine diseases) to the Rotterdam Study are gratefully acknowledged. The contribution of Guy Brusselle to the study of respiratory diseases is also gratefully acknowledged.

The Fenland Study was funded by the Wellcome Trust and the Medical Research Council. We are grateful to all the volunteers for their time and help, and to the General Practitioners and practice staff for help with recruitment. We thank the Fenland Study Co-ordination team, the Field Epidemiology team and the Fenland Study Investigators.

The Finnish Twin Cohort (FTC) study received financial support from the Academy of Finland Center of Excellence in Complex Disease Genetics, ENGAGE project and grant agreement HEALTH-F4-2007-201413 and the GenomeEUtwin project, which was supported by the European Commission under the program 'Quality of Life and Management of the Living Resources' of 5th Framework Program (no. QLG2-CT-2002-01254). The DNA extractions, sample quality controls, biobank up-keep and aliquotting was performed in the National Public Health Institute, Helsinki, Finland.

The Kooperative Gesundheitsforschung in der Region Augsburg (KORA) research platform was initiated and financed by the Helmholtz Center Munich, German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research and by the State of Bavaria. The work of KORA is supported by the German Federal Ministry of Education and Research (BMBF) in the context of the German National Genome Research Network (NGFN-2 and NGFN-plus). Our research was supported within the Munich Center of Health Sciences (MC Health) as part of LMUinnovativ.

The London Life Sciences Population (LOLIPOP) study is supported by the National Institute for Health Research Comprehensive Biomedical Research Centre at Imperial College Healthcare NHS Trust, the British Heart Foundation (SP/04/002), the Medical Research Council (G0700931), the Wellcome Trust (084723/Z/08/Z), and the National Institute for Health Research (RP-PG-0407-10371). We thank the participants and research staff who made the study possible.

The Northern Finland Birth Cohort (NFBC) 1966 study program received financial support from the Academy of Finland (project grants 104781, 120315, 110143, 132071, Sigrid Juselius Foundation, NARSAD, Stanley Medical Research Institute and Center of Excellence in Complex Disease Genetics), University Hospital Oulu, Biocenter, University of Oulu, Finland, NHLBI NIH grant 1-R01HL087679-01,through the STAMPEED program, NIMH NIH grant (1RL1MH083268-01), ENGAGE project and grant agreement HEALTH-F4-2007-201413, the Medical Research Council (studentship grant G0500539, centre grant G0600705), the Wellcome Trust (project grant GR069224), UK, and the Research Council UK fellowship. The DNA extractions, sample

quality controls, biobank up-keep and aliquotting were performed in the National Public Health Institute, Biomedicum Helsinki, Finland and supported financially by the Academy of Finland and Biocentrum Helsinki. Genotyping was supported by grant 5RL1MH083268 from the National Institute of Mental Health.

The Netherlands Twin Registry (NTR) and the Netherlands Study of Depression and Anxiety (NESDA) acknowledge the support received from NWO/ZonMW: Genetic basis of anxiety and depression (904-61-090); Resolving cause and effect in the association between exercise and well-being (904-61-193); Twin-family database for behaviour genomics studies (480-04-004); Twin research focusing on behaviour (400-05-717), Center for Medical Systems Biology (NWO Genomics); Spinozapremie (SPI 56-464-14192). J.M.V. is financially supported by NWO (VENI 451-06-004). ABS was funded by the Center for Medical Systems Biology (CMSB). NTRNESDA also acknowledges support from: Geestkracht program (10-000-1002); matching funds from universities and mental health care institutes involved in NESDA (GGZ Buitenamstel-Geestgronden, Rivierduinen, University Medical Center Groningen, GGZ Lentis, GGZ Friesland, GGZ Drenthe); Centre for Neurogenomics and Cognitive Research (CNCR-VU). Genotyping was funded by the Genetic Association Information Network (GAIN) of the Foundation for the US National Institutes of Health, and analysis was supported by grants from GAIN and the NIMH (MH081802). Genetics of Mental Illness: A lifespan approach to the genetics of childhood and adult neuropsychiatric disorders and comorbid conditions (ERC), Genetic determinants of risk behavior in relation to alcohol use and alcohol use disorder: a developmental perspective (ZonMW (Addiction) 31160008).

The Prevention of Renal and Vascular End-stage disease (PREVEND) study is supported by the Dutch Kidney Foundation (Grant E033), EU project grant GENECURE (FP-6 LSHM CT 2006 037697), and NWO VENI (grant number 916.76.170).

Prospect-The European Prospective Investigation Into Cancer and Nutrition (EPIC; The Netherlands) was funded by the European Commission - Europe Against Cancer: WHO AEP/90/05; The Dutch Ministry of Health; The Dutch Prevention Funds; the LK Research Funds; and the WCRF funds (WCRF 98A04 and WCRF 2000/30). Genotyping for this project was funded through an Incentive Grant from the Board of the UMC Utrecht.

SORBS (regional study on population from Eastern Germany of Slovanic Origin) received financial support from the German Research Council (KFO-152), IZKF (B27) and the German Diabetes Association. The authors would like to thank Peter Kovacs and Knut Krohn (Microarray Core Facility of the Interdisciplinary Centre for Clinical Research, University of Leipzig) for the genotyping/analytical support and Joachim Thiery (Institute of Laboratory Medicine, Clinical Chemistry and Molecular Diagnostics, University of Leipzig) for clinical chemistry services.

Recruitment and data collection for the Australian Semi-Structured Assessment for Genetics of Alcoholism (SSAGA) Study was supported by grants from NIH (AA07535 to A.C.H) and the Australian National Health and Medical Research Council (to N.G.M); genotyping was funded by the European Community's Seventh Framework Programme (FP7/2007-2013), ENGAGE project, grant agreement HEALTH-F4-2007-201413.

The Twins UK study was funded by the Wellcome Trust; European Community's Seventh Framework Programme (FP7/2007-2013)/grant agreement HEALTH-F2-2008-201865-GEFOS and (FP7/2007-2013), ENGAGE project grant agreement HEALTH-F4-2007-201413 and the FP-5 GenomEUtwin Project (QLG2-CT-2002-01254). The study also receives support from the Dept of Health via the National Institute for Health Research (NIHR) comprehensive Biomedical Research Centre award to Guy's & St Thomas' NHS Foundation Trust in partnership with King's College London. T.D.S. is an NIHR senior Investigator. The project also received support from a *Biotechnology and Biological Sciences Research Council* (BBSRC) project grant (G20234).The authors acknowledge the funding and support of the National Eye Institute via an NIH/CIDR genotyping project (PI: Terri Young). The authors thank the staff from the Genotyping Facilities at the Wellcome Trust Sanger Institute for sample preparation, Quality Control and Genotyping led by Leena Peltonen and Panos Deloukas; Le Centre National de Génotypage, France, led by Mark Lathrop, for genotyping; Duke University, North Carolina, USA, led by David Goldstein, for genotyping; and the Finnish Institute of Molecular Medicine, Finnish Genome Center, University of Helsinki, led by Aarno Palotie. Genotyping was also performed by CIDR as part of an NEI/NIH project grant.

The Utrecht Health Project (UHP) was funded with grants the Ministry of Health, Welfare, and Sports (VWS), the University of Utrecht, the Province of Utrecht, the Dutch Organisation of Care Research (ZON), the University Medical Center of Utrecht (UMC Utrecht) and the Dutch College of Healthcare Insurance Companies (CVZ). Genotyping for this project was funded through an Incentive Grant from the Board of the UMC Utrecht.

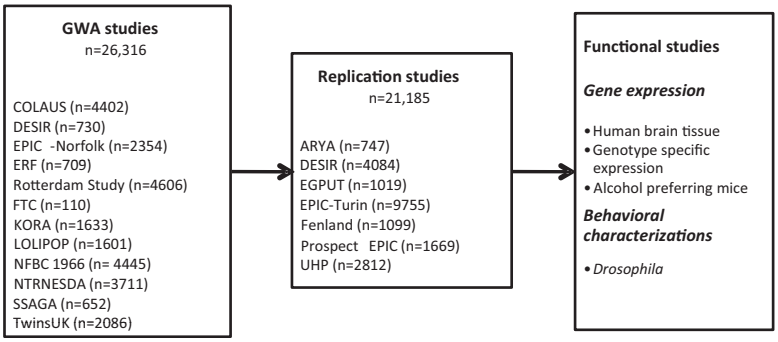

Fig. S1. Project outline and design for genome-wide association (GWAS), replication, and functional studies.

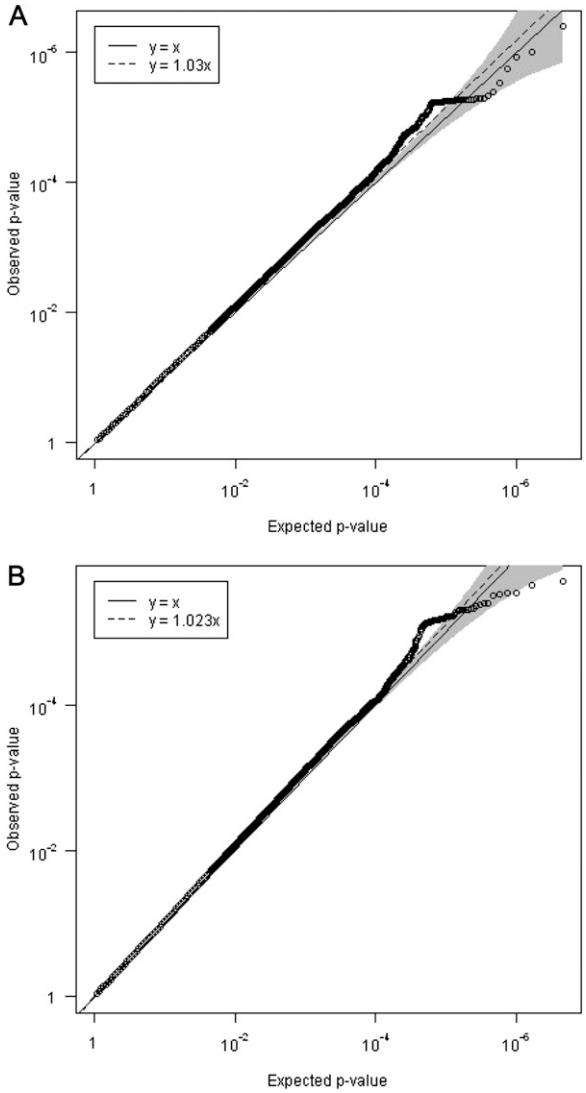

Fig. S2. (Continued)

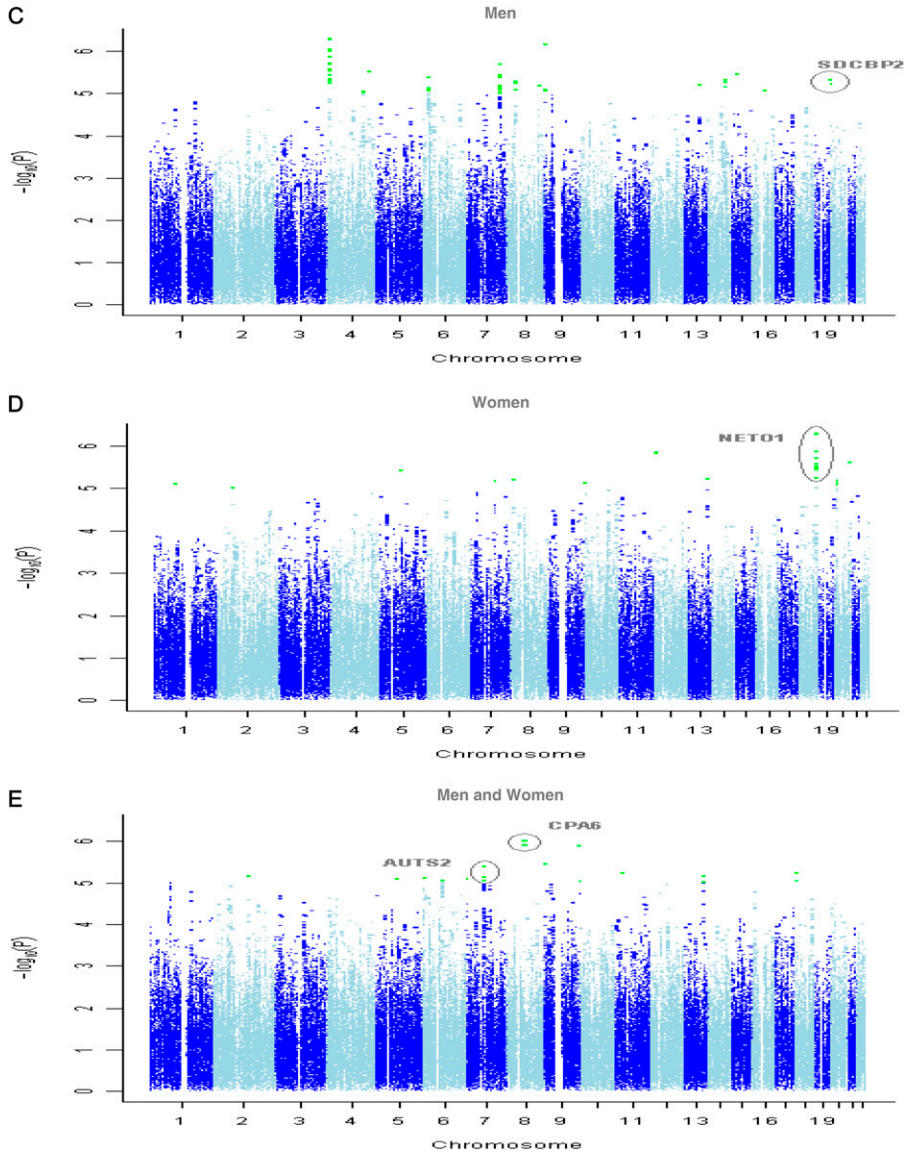

Fig. S2. (Continued)

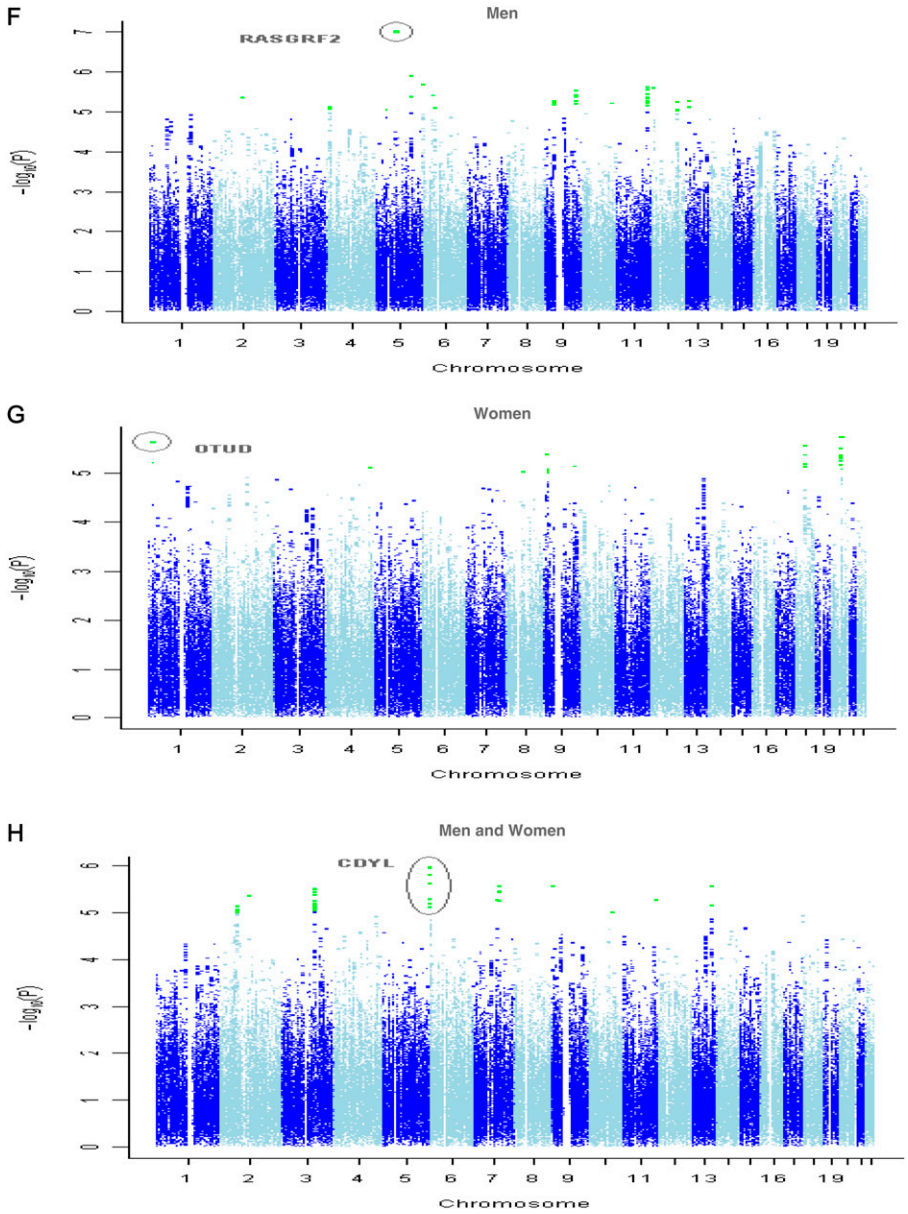

Fig. S2. Quantile–quantile plots and Manhattan plots (the association peaks that were taken through to the replication stage are indicated with circles on the Manhattan plots). (*A*) Quantile–quantile plot of test statistics from metaanalyses of quantile transformation (includes nondrinkers). The gray area represents the 95% confidence intervals around the null distribution of the test statistic. The dashed line shows the expected value after adjusting for genomic control. (*B*) Quantile–quantile plot of test statistics from metaanalyses of log transformation (drinkers only). The gray area represents the 95% confidence intervals around the null distribution of the test statistic. The dashed line shows the expected value after adjusting for genomic control. (*C*) Manhattan plots for quantile transformation (including nondrinkers) in men (the SNPs for replication were chosen before correction of a cohort-specific error that increased significance of a second association peak on chromosome 4). (*D*) Manhattan plots for quantile transformation (including nondrinkers) in women. (*E*) Manhattan plots for quantile transformation (including nondrinkers) in men and women combined. (*F*) Manhattan plots for log transformation (drinkers only) in men. (*G*) Manhattan plots for log transformation (drinkers only) in women (the SNPs for replication were chosen before correction of a cohort-specific error that increased significance of a second peak on chromosome 20). (*H*) Manhattan plots for log transformation (drinkers only) in men and women combined.

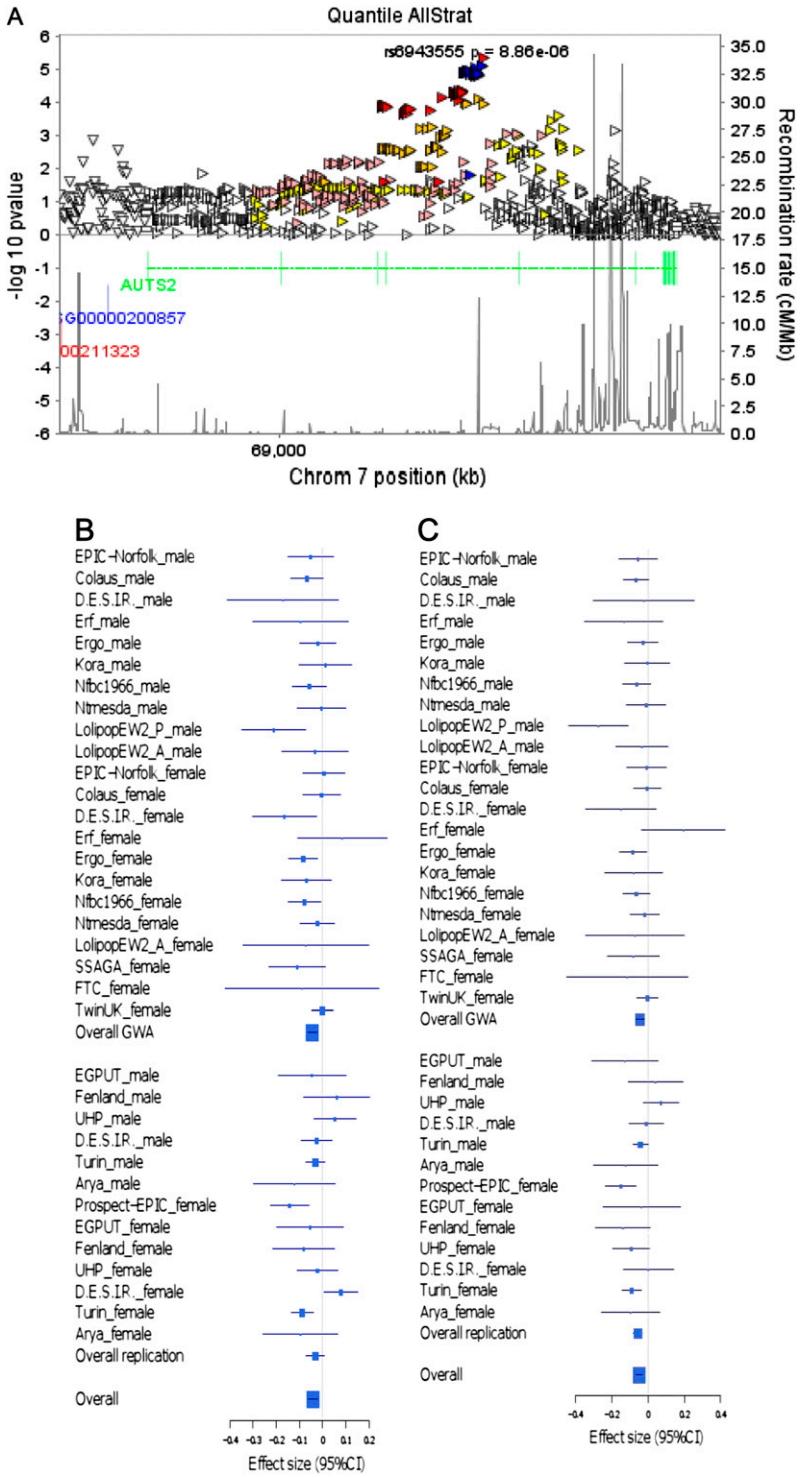

Fig. S3. (Continued)

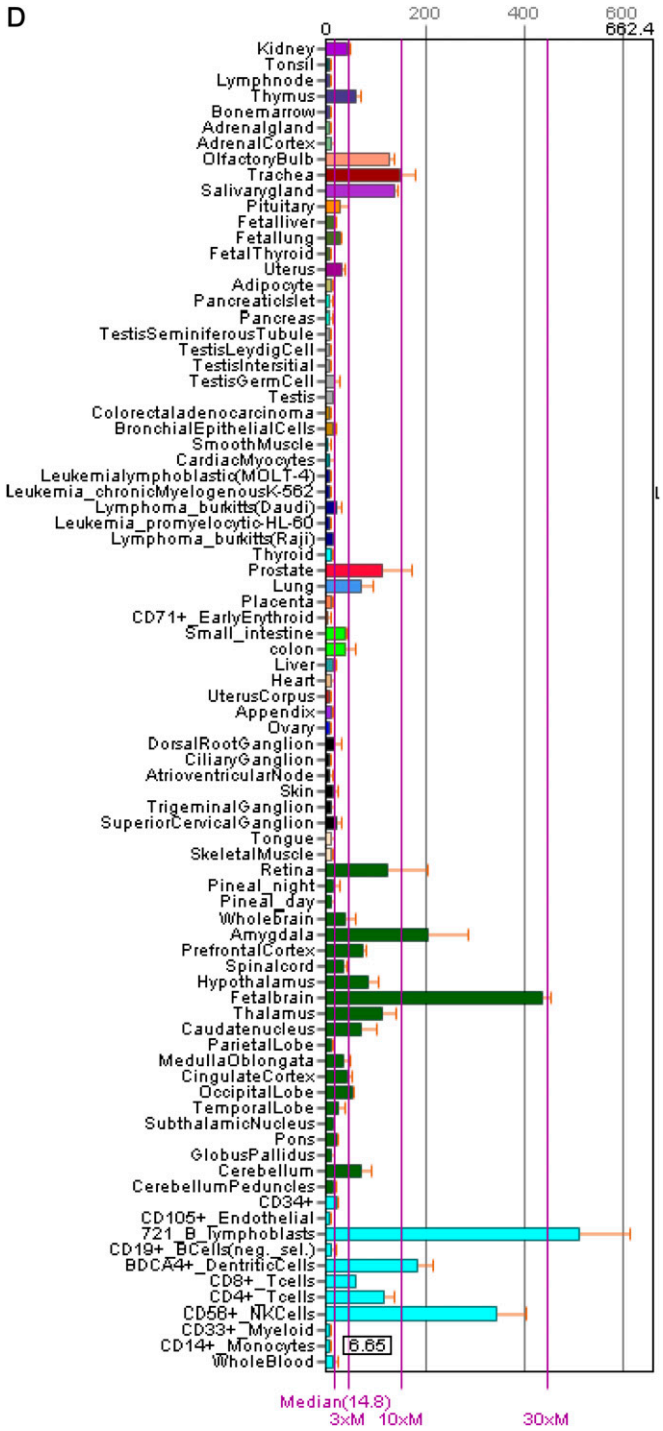

Fig. S3. Analyses of *autism susceptibility candidate 2* gene (*AUTS2*). (*A*) Regional association plot around rs6943555 in *AUTS2*. SNP is plotted by chromosomal position (National Center for Biotechnology Information build36) against association with alcohol intake in grams per day per kilogram ($-\log_{10} P$ value). SNP location in the gene is shown in symbols: right triangle, intronic; left triangle, stop gained or stop lost; down triangle, intergenic; horizontal rectangle, upstream or downstream; vertical rectangle, 3′-UTR or 5′-UTR. R^2 between the SNP and the top signal is color coded: blue, >0.99; red, 0.8–0.99; orange, 0.5–0.8; pink, 0.2–0.5; yellow, 0.1–0.2; white, 0–0.1. (*B*) Forest plot for rs6943555 showing effect size (95% confidence interval) from quantile transformation (includes nondrinkers). (*C*) Forest plot for rs6943555 showing effect size (95% confidence interval) from log transformation (drinkers only). (*D*) Bar chart obtained from BioGPS depicting relative *AUTS2* mRNA expression across a range of human tissues as measured by probe 212599_at on the Human Genome U133 Plus 2.0 array (Affymetrix). *AUTS2* expression values (upper axis) relate to fluorescence intensity after GC-RMA normalization; median *AUTS2* expression value (lower axis) across all tissues represents the standard.

Other Supporting Information Files

Table S1 (DOCX)
Table S2 (DOCX)
Table S3 (DOCX)
Table S4 (DOCX)
Table S5 (DOCX)
Table S6 (DOCX)
Table S7 (DOCX)