Supplementary Materials of Heritability of liver enzyme levels estimated from genome-wide SNP data

Materials and Methods

Participants

For 6,945 subjects who participated in the NTR biobank study (performed in twin-families) [1], valid genotype data and data on one or more liver enzyme concentrations were available. Data were excluded for nine individuals with known liver disease (ICD-10 codes K70-K77) and 239 individuals from non-European descent. Data were selected from unrelated individuals, which resulted in a sample of 3,309 NTR participants (60.6% females; year of birth 1914-1987). For 2,533 individuals who participated in the NESDA biobank study [2], liver enzyme concentrations and genotype data were available. Data were excluded for 155 individuals from non-European descent and for 266 persons who were (closely or distantly) related to other NESDA or NTR participants, resulting in 2,112 unrelated individuals whose data were included in the analyses (66.6% females; year of birth 1939-1988). The total NTR/NESDA sample thus consisted of 5,421 unrelated individuals (3,309 + 2,112). Permission for the biobank studies was obtained from the Central Ethics Committee on Research Involving Human Subjects of the VU University Medical Center Amsterdam, and informed consent was obtained from all participants [1;2].

Meta-analysis summary statistics (*z*-scores and *p*-values) for GGT and ALT levels originated from a large meta-analysis on data from 52,350 individuals with Caucasian ethnicity (including 1,721 NTR and 1,724 NESDA participants; <5% of the total sample size), as well as from 8,739 participants with an Indian-Asian background [3].

To compare the performance of the DE method, SNP-based heritability estimates were also estimated for BMI (which served as a bench mark trait since its additive genetic variance

explained by SNPs has been studied before [4]). For nearly all NTR/NESDA participants, data on BMI (*N*=5,406) were assessed at the same time as their liver enzyme data. Meta-analysis summary statistics (*p*-values) from large GWA studies on BMI (*N*=249,796, including 3,516 participants from NTR/NESDA) [5] were downloaded from http://www.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files. Table S1 gives a summary of all data that were available for each phenotype, with the respective numbers of participants.

Genotyping and quality control

Genotyping in the combined NTR/NESDA sample was performed on five platforms: Affymetrix 6.0 (*N*=3,180), Affymetrix 5.0-Perlegen (*N*=2,845) (Affymetrix, Santa Clara California), Illumina 660 (*N*=926), Illumina Omni Express 1M (*N*=225), and Illumina 370 (*N*=139) (Illumina, San Diego, California). Note that the number of individuals add up to >5,421 since 1,817 individuals were typed on two or more platforms. Overlap in SNPs between different arrays was between 110K and 500K; ~80K SNPs were present on all arrays.

SNP data quality control (QC) was performed in several steps. First, for all platforms, SNPs were aligned to the GIANT 1000 Genomes build 37 (HG19) ALL reference set (Phase I Integrated Release Version 3 Haplotypes) [6] using the liftOver [7] tool. SNPs were removed if they matched to none or multiple positions, or did not have matching or had opposite strand alleles, if allele frequencies differed >.20 compared to the reference set, and/or if alleles were C/G or A/T, and MAF >.35. A next round of QC was platform-specific. SNPs were removed if SNP missingness was >5%, allele frequencies deviated from Hardy-Weinberg equilibrium (HWE; p<.00001), Mendelian error rate was >1%, double samples typed error rate was >1%, or MAF was <.01. Samples were excluded if missing rate was >10%, the Plink *F* inbreeding coefficient <-.10 or >.10, or sex status did not match genotypic sex. Data of individual

platforms were then merged and identity by descent (IBD) and identity by state (IBS) relations were calculated for each pair of individuals. Family structure was updated if unexpected IBD/IBS relations were consistent with biographic information. Otherwise samples with unexpected IBD/IBS relations were removed. For individuals or MZ twin pairs who were genotyped on multiple platforms, samples were removed if DNA differed >1% over platform. In addition, samples were removed if the Mendelian error rate was >2%.

Imputation was performed in two stages. Phasing and imputation of genotype platform specific SNPs was performed with Mach [8]. Imputation was performed with Minimac (http://genome.sph.umich.edu/wiki/minimac; an extention of MaCH software which takes haplotypes as input) with the GIANT 1000 Phase I Integrated release version 3 ALL panel as reference set (GIANT.phase1_release_v3.20101123.snps_indels_svs.genotypes.refpanel.ALL.vcf.gz.tgz) [6]. In a stringent final round of quality control, SNPs were removed if imputation quality was low (r^2_hat <.8), SNPs showed platform effects based on testing individuals who were genotyped on one platform ('cases') against individuals who were genotyped on all other platforms ('controls') (p<. 0001), or MAF <.001, resulting in a final dataset of 5,994,956 autosomal SNPs.

Phenotypes

Liver enzymes were determined in heparin plasma tubes that were collected after overnight fasting. Before the start of the blood sample collection, the NTR and NESDA biobank protocols for processing and storage of blood samples were harmonized [9]. For the NTR biobank project, blood samples were collected in the morning at the participant's home or, if preferred, at his/her work and processing took place at one central laboratory within six hours (average transport time 196 minutes). Blood was collected in two heparin plasma tubes, which

were stored in melting ice during transportation. When the samples arrived at the laboratory, plasma was collected and six samples of 500 µL were snap-frozen and stored at -30 °C. Liver enzyme levels were determined with Vitros assays (Vitros 250, Ortho-Clinical Diagnostics; Johnson & Johnson, Rochester, USA) in units per liter (U/L) [1]. The average time between blood sample collection and liver enzyme determination was 37.0 months (*SD*=12.0). For NESDA, blood sampling took place during the NESDA baseline assessment at seven field sites that were within walking distance of a laboratory. At the laboratory, processing of blood samples and determination of liver enzyme levels was performed within the hour [2;9]. Liver enzyme levels were determined with Roche Diagnostic assays [3].

Statistical analyses

Preparatory analyses

Creating sample of unrelated individuals To create a sample of independent individuals, for all NTR and NESDA individuals with valid genotype and liver enzyme level data a genetic relatedness matrix (GRM) was estimated (option --make-grm) using the free software package GCTA (v1.24.2) [10]. This GRM was then pruned for relatedness at a level of .025 (option --grm-cutoff 0.025), resulting in a set of 5,421 individuals with estimated relatedness <.025 (for each pair of individuals with an estimated relatedness >.025, one individual was removed).

Fixed effects of source and sex Liver enzyme values were *log*-transformed to approximate normality. Differences in liver enzyme levels were examined with respect to source (NTR, NESDA) and sex (male, female) by independent-samples *t*-tests. Based on these analyses, regression analyses were carried out with source (NTR/NESDA), age, sex, time in months between blood draw and liver enzyme assessment, platform effects, three PCs representing population stratification, and lab effects as covariates. Lab effects (among

NESDA participants) were (significant and) included as covariate for AST levels only, not for GGT and ALT levels. All regression analyses were performed in SPSS 19.0 [11]. The residuals from these regression analyses were used in all subsequent analyses.

GWA In the NTR/NESDA dataset, SNP associations were tested in a linear model assuming additive SNP effects using Plink (v1.07) [12]. GWA results are the input for the methods to estimate heritability, and were therefore inspected by quantile-quantile (QQ) and Manhattan plots with the R-package qqman [13]. QQ plots depict the observed distribution of *p*-values against the expected *p*-values. Manhattan plots show the *p*-values of the individual association tests for each SNP. Thresholds for suggestive and genome-wide significance were $1x10^{-5}$ and $5x10^{-8}$ respectively.

Main analyses

SNP heritability based on the NTR/NESDA sample

GRM method A linear mixed model with the genetic relationships as a random effect was used to estimate the phenotypic variance that is due to the genetic relatedness captured by the GRM, using the software package GCTA (v1.24.2) [10]. Estimation was performed using restricted maximum likelihood (REML) (option --reml). In additional analyses, the variance that can be explained by SNPs on each individual chromosome was estimated by genetic relatedness matrices that were estimated for each chromosome separately.

The GRM method provides an unbiased estimate of the variance that can be explained by the joint additive effects of all SNPs if the causal SNPs are perfectly correlated with the measured/imputed SNPs (i.e., if linkage disequilibrium, LD, is 1). In the case of imperfect LD, for instance when the distribution of minor allele frequencies (MAF) differs for the causal and measured/imputed SNPs, estimates are biased downwards. To avoid confounding of additive SNP effects with environment effects shared by family members and possible

causal variants that are not tagged by SNPs, the genetic relatedness was calculated for unrelated individuals only [14].

DE method Analyses with the DE method were performed in R3.0.2 [15] with the script for continuous traits obtained from the developer's website: https://sites.google.com/site/honcheongso/software/total-vg The method takes a vector with *z*-statistics as input that correspond to the effect sizes of SNPs obtained in a GWA study. Under a standard normal probability distribution, *z*-statistics correspond to specific *p*-values and vice versa (ignoring the sign of the *z*-statistics), with *z*-statistics defined as the standardized deviation of a value *x* to the mean $\bar{x}$, that is $z = \frac{x - \bar{x}}{sd(x)}$, with $\bar{x}$=0 and sd(*x*)=1. Using this relation, *z*-statistics were extracted from GWA *p*-values. The sign of the *z*-statistics was permuted 100 times, in order to avoid incorrect estimates in case all observed effects are in the same direction. So et al. [16] described two different estimates, based on conditional and unconditional kernel estimation respectively, but noted that unconditional Kernel estimation is more accurate and stable with pruned data than conditional Kernel estimation. This was in line with our preliminary results. Therefore reported results were based on unconditional Kernel density estimation. To avoid inflation of the variance explained by SNPs by redundant SNP effects, SNPs in high LD with other SNPs were removed by pruning. Note that pruning was independent of the trait associations. Based on samples containing ≤2.7M SNPs), So et al. [16] recommended LD-based SNP pruning at an r^2 level of .25 (--indep-pairwise 100 25 0.25). For the current study (with the NTR/NESDA set containing ~6M SNPs), this level of pruning was also applied, resulting in a set of 226,243 SNPs.

Since the DE method does not provide standard errors, we obtained an indication of the stability/variability of the heritability estimates across different sets of SNPs. To this end, the NTR/NESDA dataset was pruned 10 times (resulting in sets differing <50%). The analysis

was carried out on each pruned set, and results were then averaged. Note that the variability across 10 pruned sets should not be interpreted as a standard error.

SNP heritability in the single sample with meta-analysis results

DE method To compare SNP-based heritability for a single sample (NTR/NESDA) with that for the consortium GWA meta-analysis results, the same pruned set was used to calculate DE estimates for both datasets. This pruned set consisted of SNPs that were present in the GGT and ALT meta-analyses as well as in the NTR/NESDA dataset. Pruning was based on the LD pattern among SNPs in the NTR/NESDA dataset since the GWA meta-analysis results did not include raw SNP data, and was performed at a level of r^2 .25 as suggested by So et al. [16] (Plink options --indep-pairwise 100 25 0.25) resulting in a pruned set of 111,995 SNPs. Note that the size of this pruned set differed from that described above, since here a dataset of ~2.7M SNPs was pruned at r^2 .25; whereas for the comparison with GRM-based estimates described above, the entire dataset (containing ~6M SNPs) was pruned at r^2 .25.

SNP markers in the GWA meta-analyses were imputed against build 36 (HG18) of the Human reference genome [3;5], thus to match these results to the pruned set that was based on the LD pattern in the NTR/NESDA data, the genetic map of the SNP markers was lifted over to build 37 (HG19) that was used as a reference for the NTR/NESDA SNP dataset, using the liftOver [7] tool. SNPs were removed if they matched to none or multiple locations (for GGT, ALT, BMI respectively for 750, 740, 462 SNPs) or did not have matching alleles (for GGT, ALT, BMI respectively for 2,202, 2,201, 1,857 SNPs). For the GWA meta-analysis results on BMI which did not contain information on chromosome and base pair positions of SNPs, this information was obtained by merging the GWA meta-analysis results to the Hapmap 2 CEU reference set (release 22, hapmap-ceu-all.zip, downloaded from the Plink [12] website http://pngu.mgh.harvard.edu/~purcell/plink/res.shtml#hapmap) that had been used as a

reference set for imputation in the individual cohorts included in the meta-analysis. Note that when filtering the NTR/NESDA set on the ~2.7M SNPs that were included in the GGT and ALT meta-analyses, GWA meta-analysis results were left out for 30,065 (GGT), 30,057 (ALT), and 9,917 (BMI) SNPs, since these SNPs were not included in the NTR/NESDA dataset.

For the meta-analysis results, the DE method was applied to *z*-scores or *p*-values that were uncorrected for the overall genomic inflation factor correction (not for the study-specific genomic inflation factors). To this aim, for GGT and ALT [3], the uncorrected meta-analysis *z*-scores were analyzed. These *z*-scores were based on study-specific GWA analyses in which *log*-transformed liver enzyme levels were corrected for sex, age, PCs representing effects of population stratification (and in case-control studies, for case-control status). For BMI [5], meta-analysis *p*-values were analyzed that were 'uncorrected' for the overall genomic control inflation factor by transforming the *p*-values into χ^2 statistics, multiplying the χ^2 statistics by the genomic inflation factor (1.318) and then calculating new *p*-values from these new χ^2 statistics. These meta-analysis *p*-values were based on study-specific GWA analyses in which levels of BMI were corrected for age, age^2, and PCs representing effects of population stratification, stratified by sex and case-control status (in case-control studies).

As indicated above, to verify that DE estimates did not depend on a specific pruned set of SNPs, the NTR/NESDA SNP dataset was pruned 10 times (with <50% overlap in SNPs between sets) and DE estimates were averaged over these 10 pruned sets.

An overview of all analyses performed on the datasets is included in Table S1. SNP heritability estimates obtained with the GRM method were considered to be significant if *p*-values <.05. In the case that GRM- and DE-based estimates differed, we applied a conservative approach by focusing on the lower of the two estimates since Walters [17] has shown that DE-based heritability estimates could be overestimated when sample size is small.

The data used for this article will be made available on request to the NTR committee (ntr@psy.vu.nl). Genotype data used for this study are included in two datasets available by application to dbGaP (http://www.ncbi.nlm.nih.gov/gap/), accession numbers phs000020.v2.p1 and phs000486.v1.p1.

Supplementary Tables

Table S1. Number of participants with SNP data and information on liver enzyme levels and BMI, and methods applied to these phenotypes/ samples to estimate SNP-based heritability

	Phenotypes				Methods	
	GGT	ALT	AST	BMI	GRM-based	DE-based
NTR+NESDA	5,390	5,285	5,402	5,406	X	X
Meta-analysis summary statistics consortia[a]	57,690	55,486		~123,912		X

[a] For all markers, data was available for a subset of the participants in the study (see text). These numbers reflect the maximum number of observations that was available for one marker.

Table S2. Correlations between liver enzyme levels (*log*-transformed), BMI, and age, split over source (NTR, NESDA) and sex [a b]

	NTR								NESDA							
	Males				Females				Males				Females			
	ALT	AST	BMI	age	ALT	AST	BMI	age	ALT	AST	BMI	age	ALT	AST	BMI	age
GGT	**.26**	**.42**	**.27**	**.20**	**.30**	**.38**	**.25**	**.29**	**.53**	**.35**	**.36**	**.29**	**.53**	**.34**	**.27**	**.22**
ALT		**.56**	**.14**	**-.20**		**.52**	**.07**	**.06**		**.66**	**.35**	.08		**.59**	**.31**	**.23**
AST			**.18**	.00			.01	**.21**			**.16**	-.01			**.11**	**.15**
BMI				**.26**				**.28**				**.29**				**.25**

[a] Correlations significant at $p<.05$ shown in bold

[b] In the GWA meta-analysis by Chambers et al. [3], the correlation between *log*-transformed values of ALT and GGT was .64.

Table S3. GRM-based estimates (with standard error) on proportions of variance explained per chromosome for the NTR/NESDA sample [a] for liver enzyme levels and BMI

	GRM-based estimates per chromosome																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
GGT	.011 (.016)	**.027 (.017)**	**.025 (.015)**	.006 (.014)	.000 (.014)	.008 (.013)	.019 (.014)	.000 (.013)	.022 (.014)	**.029 (.014)**	.000 (.012)	.018 (.014)	.000 (.012)	.008 (.011)	.000 (.009)	.005 (.011)	.003 (.010)	.008 (.011)	.002 (.008)	**.023 (.011)**	.000 (.008)	**.013 (.008)**
ALT	.000 (.015)	.002 (.016)	.007 (.015)	.003 (.014)	.005 (.014)	.017 (.014)	.016 (.014)	.010 (.013)	.000 (.013)	.016 (.014)	.019 (.013)	.007 (.013)	.000 (.011)	.000 (.011)	.000 (.010)	.000 (.011)	.000 (.010)	.000 (.011)	.000 (.008)	.000 (.010)	.000 (.009)	.003 (.007)
AST	.012 (.016)	**.034 (.017)**	.000 (.014)	.000 (.014)	.019 (.015)	**.027 (.015)**	.008 (.014)	.013 (.014)	.000 (.013)	.003 (.012)	.003 (.009)	.004 (.013)	.000 (.011)	.007 (.010)	.000 (.010)	.000 (.011)	.000 (.010)	.008 (.011)	.007 (.008)	.000 (.009)	.005 (.008)	.004 (.007)
BMI	.000 (.016)	**.048* (.017)**	.000 (.015)	.008 (.014)	.000 (.014)	.017 (.015)	.022 (.014)	.022 (.014)	.012 (.013)	.000 (.013)	.000 (.011)	.015 (.013)	.000 (.011)	.007 (.010)	.000 (.010)	.010 (.012)	.004 (.010)	**.035* (.012)**	.001 (.008)	.000 (.010)	.006 (.008)	.004 (.008)

[a] Estimates significant at $p<.05$ shown in bold; at $p<.002$ indicated by *

Figure Captions and legends

Figure s1A,C-s4A,C: QQ plots with *p*-values resulting from GWAs based on NTR/NESDA data set and meta-analysis-based data from large consortia.

Figure s1B,D-s4B,D: Manhattan plots that show the (negative logarithm) of the *p*-value of association (on the *y* axis) for each SNP along the chromosome (on the *x* axis). Thresholds for suggestive and genome wide significance are $1x10^{-5}$ and $5x10^{-8}$ respectively. Annotated SNPs are located in or flanking the corresponding gene, unless for locations that are indicated by two gene names (those are intergenic).

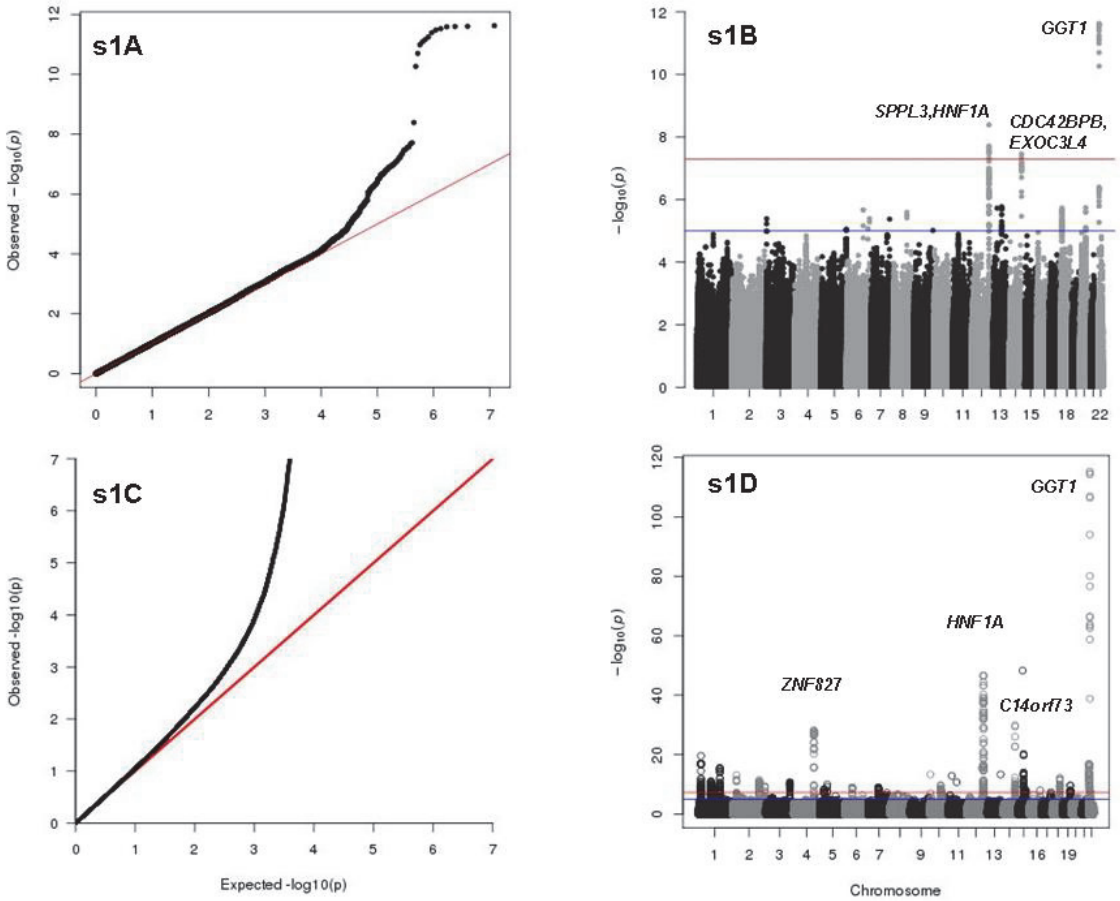


Figure s1A-s1D QQ and Manhattan plots for GGT

s1A-B: combined NTR/NESDA sample; s1C-D: meta-analysis sample [3]

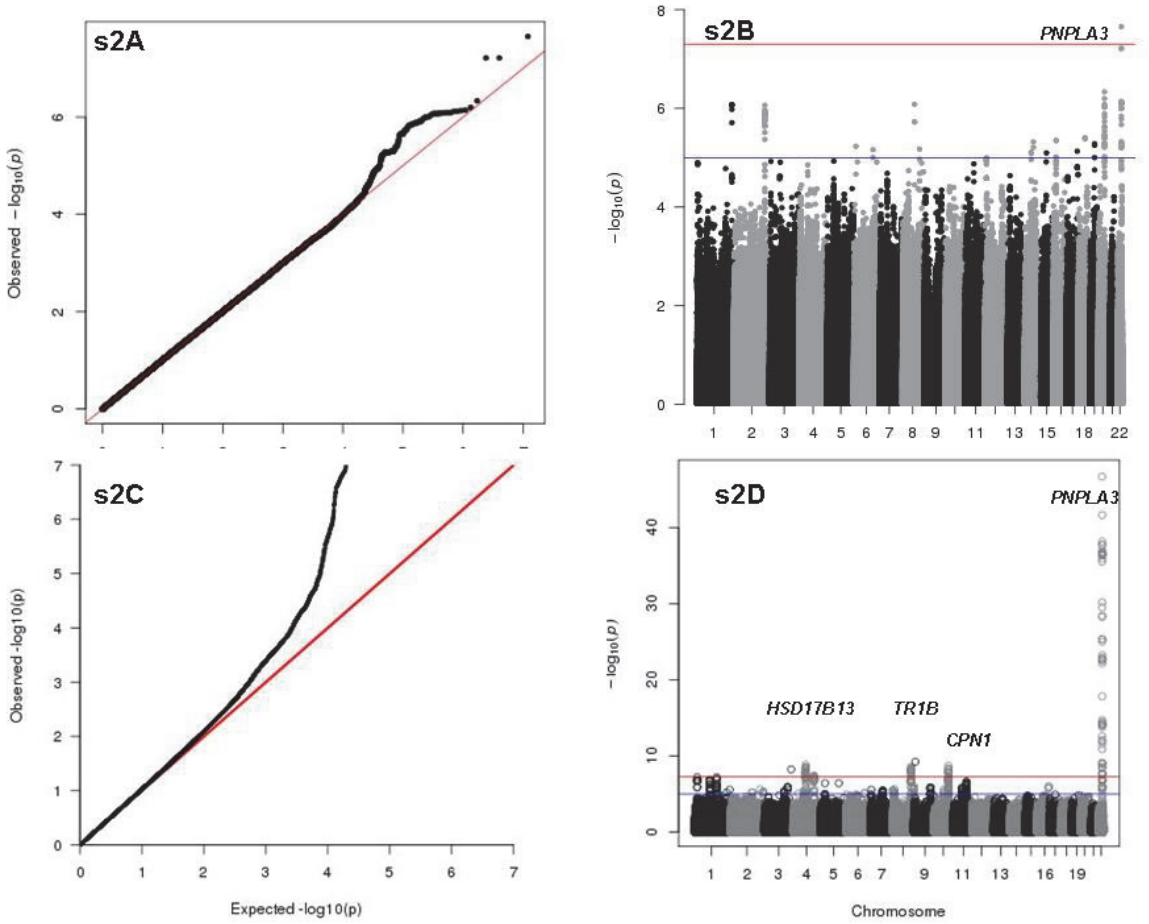


Figure s2A-s2D QQ and Manhattan plots for ALT

s2A-B: combined NTR/NESDA sample; s2C-D: meta-analysis sample [3]

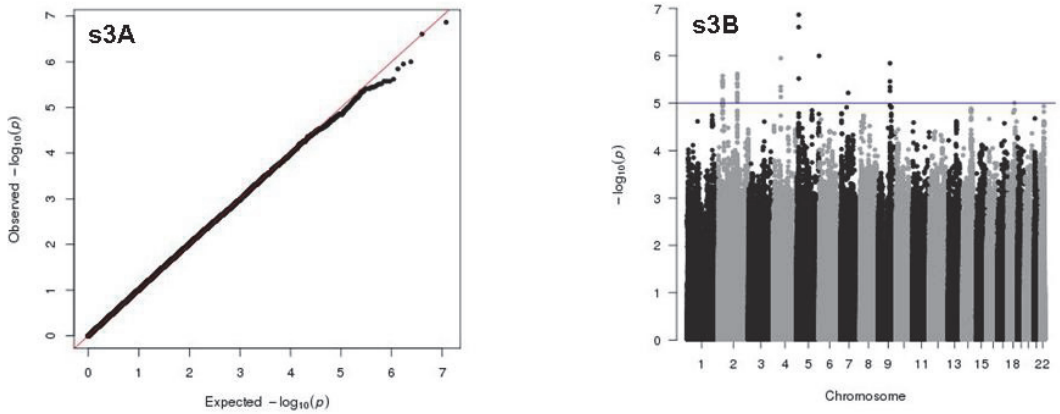
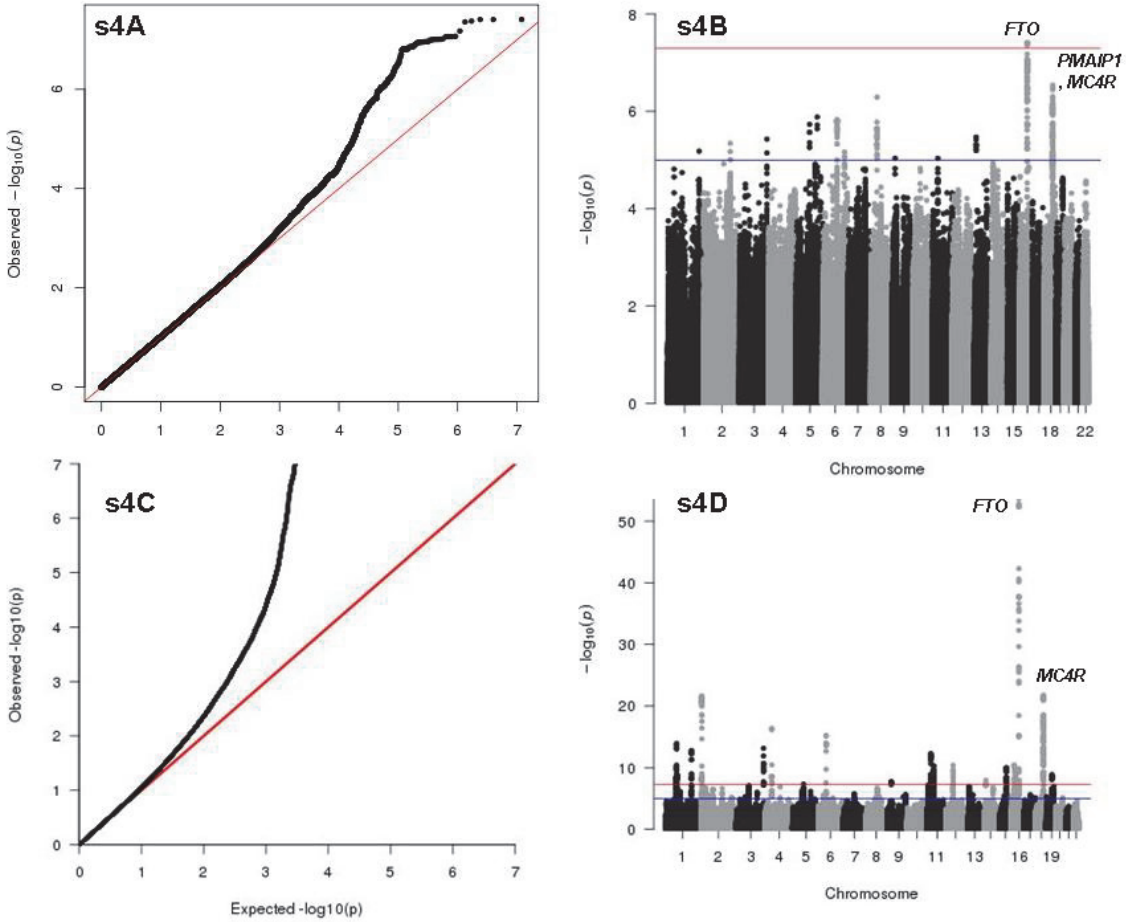


Figure s3A-s3B QQ and Manhattan plots for AST

s3A-B: combined NTR/NESDA sample



Figures s4A-s4D QQ and Manhattan plots for BMI

s4A-B: combined NTR/NESDA sample; S4C-D: meta-analysis sample [5]

Acknowledgments

This work was supported by grants from the Netherlands Organization for Scientific Research [ZonMW Addiction 31160008; ZonMW 940-37-024; NWO/SPI 56-464-14192; NWO-400-05-717; NWO-MW 904-61-19; NWO-MagW 480-04-004; NWO-Veni 016-115-035], the European Research Council [Genetics of Mental Illness: ERC-230374], Geestkracht [10-000-1002], the National Institute on Drug Abuse [DA018673] and Royal Netherlands Academy of Arts and Science [ISK/6827/VPPl], funds from universities and mental health care institutes involved in NESDA, the Centre for Medical Systems Biology (NWO Genomics), Netherlands Bioinformatics Center/BioAssist/RK/2008.024, Biobanking and Biomolecular Resources Research Infrastructure [NL-184.021.007], European Science Council [230374], Rutgers University Cell and DNA Repository cooperative agreement [National Institute of Mental Health U24 MH068457-06], as well as the EMGO travel Grant and VU-FPP Regeling Talentgeld.

We thank the consortium for providing the GWA meta-analysis results on GGT and ALT, professor P.C. Sham for useful discussions on the DE method and the Center for Research Computing at the University of Notre Dame (Indiana, USA) and SURFsara (www.surfsara.nl) for using the computing facilities. The work has been executed in the Mental Health research program of the EMGO Institute for Health and Care research and was partly executed during a stay at the University of Notre Dame (Indiana, USA).

Full list of members of the LivGen consortium

John C Chambers,[1,2,3,113] Weihua Zhang,[1,3,113] Joban Sehmi,[3,4,113] Xinzhong Li,[5,113] Mark N Wass,[6,113] Pim Van der Harst,[7,113] Hilma Holm,[8,113] Serena Sanna,[9,113] Maryam Kavousi,[10,11,113] Sebastian E Baumeister,[12] Lachlan J Coin,[1] Guohong Deng,[13] Christian Gieger,[14] Nancy L Heard-Costa,[15] Jouke-Jan Hottenga,[16] Brigitte Kühnel,[14] Vinod Kumar,[17] Vasiliki Lagou,[18,19,20] Liming Liang,[21,22] Jian'an Luan,[23] Pedro Marques Vidal,[24] Irene Mateo Leach,[7] Paul F O'Reilly,[1] John F Peden,[25] Nilufer Rahmioglu,[19] Pasi Soininen,[26,27] Elizabeth K Speliotes,[28,29] Xin Yuan,[30] Gudmar Thorleifsson,[8] Behrooz Z Alizadeh,[18] Larry D Atwood,[31] Ingrid B Borecki,[32] Morris J Brown,[33] Pimphen Charoen,[1,34] Francesco Cucca,[9] Debashish Das,[3] Eco J C de Geus,[16,35] Anna L Dixon,[36] Angela Döring,[37] Georg Ehret,[38,39,40] Gudmundur I Eyjolfsson,[41] Martin Farrall,[25,42] Nita G Forouhi,[23] Nele Friedrich,[43] Wolfram Goessling,[44,45,46] Daniel F Gudbjartsson,[8] Tamara B Harris,[47] Anna-Liisa Hartikainen,[48] Simon Heath,[49] Gideon M Hirschfield,[50,51,52] Albert Hofman,[10,11] Georg Homuth,[53] Elina Hyppönen,[54] Harry L A Janssen,[10,55] Toby Johnson,[56] Antti J Kangas,[26] Ido P Kema,[57] Jens P Kühn,[58] Sandra Lai,[9] Mark Lathrop,[49,59] Markus M Lerch,[60] Yun Li,[61] T Jake Liang,[62] Jing-Ping Lin,[63] Ruth J F Loos,[23] Nicholas G Martin,[64] Miriam F Moffatt,[36] Grant W Montgomery,[64] Patricia B Munroe,[56] Kiran Musunuru,[31,65,68] Yusuke Nakamura,[17] Christopher J O'Donnell,[69] Isleifur Olafsson,[70] Brenda W Penninx,[71,72,73] Anneli Pouta,[48,74] Bram P Prins,[18] Inga Prokopenko,[19,20] Ralf Puls,[58] Aimo Ruokonen,[75] Markku J Savolainen,[26,76] David Schlessinger,[77] Jeoffrey N L Schouten,[55] Udo Seedorf,[78] Srijita Sen-Chowdhry,[1] Katherine A Siminovitch,[50,79,80,81,82] Johannes H Smit,[71] Timothy D Spector,[83] Wenting Tan,[13] Tanya M Teslovich,[84] Taru Tukiainen,[1,26] Andre G Uitterlinden,[10,11,85] Melanie M Van der Klauw,[86,87] Ramachandran S Vasan,[88,89] Chris Wallace,[33] Henri

Wallaschofski,[43] H-Erich Wichmann,[37,90,91] Gonneke Willemsen,[16,92] Peter Würtz,[1,26] Chun Xu,[93] Laura M Yerges-Armstrong,[94] Alcohol Genome-wide Association (AlcGen) Consortium,[95] Diabetes Genetics Replication and Meta-analyses (DIAGRAM+) Study,[95] Genetic Investigation of Anthropometric Traits (GIANT) Consortium,[95] Global Lipids Genetics Consortium,[95] Genetics of Liver Disease (GOLD) Consortium,[95] International Consortium for Blood Pressure (ICBP-GWAS),[95] Meta-analyses of Glucose and Insulin-Related Traits Consortium (MAGIC),[95] Goncalo R Abecasis,[84] Kourosh R Ahmadi,[83] Dorret I Boomsma,[16,92] Mark Caulfield,[56] William O Cookson,[36] Cornelia M van Duijn,[10,11,96] Philippe Froguel,[97] Koichi Matsuda,[17] Mark I McCarthy,[19,20,98] Christa Meisinger,[99] Vincent Mooser,[30] Kirsi H Pietiläinen,[100,101,102] Gunter Schumann,[103] Harold Snieder,[18] Michael J E Sternberg,[6,87] Ronald P Stolk,[104] Howard C Thomas,[2,105] Unnur Thorsteinsdottir,[8,106] Manuela Uda,[9] Gérard Waeber,[107] Nicholas J Wareham,[23] Dawn M Waterworth,[30] Hugh Watkins,[25,42] John B Whitfield,[64] Jacqueline C M Witteman,[10,11] Bruce H R Wolffenbuttel,[86,87] Caroline S Fox,[69,108] Mika Ala-Korpela,[26,27,76,113] Kari Stefansson,[8,106,113] Peter Vollenweider,[107,113] Henry Völzke,[12,113] Eric E Schadt,[109,113] James Scott,[4,113] Marjo-Riitta Järvelin,[1,74,110,111,112,113] Paul Elliott,[1,112,113] and Jaspal S Kooner[2,3,4,113]

[1]Epidemiology and Biostatistics, Imperial College London, Norfolk Place, London, UK

[2]Imperial College Healthcare National Health Service (NHS) Trust, London, UK

[3]Ealing Hospital NHS Trust, Middlesex, UK

[4]National Heart and Lung Institute, Imperial College London, Hammersmith Hospital, London, UK

[5]Institute of Clinical Science, Imperial College London, Royal Brompton Hospital, London, UK

[6]Structural Bioinformatics Group, Division of Molecular Biosciences, Imperial College London, South Kensington, London, UK

[7]Department of Cardiology, University Medical Center Groningen, University of Groningen, The Netherlands

[8]deCODE genetics, Reykjavik, Iceland

[9]Istituto di Ricerca Genetica e Biomedica del Consiglio Nazionale delle Ricerche, Monserrato, Cagliari, Italy

[10]Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands

[11]Netherlands Genomics Initiative-Sponsored Netherlands Consortium for Health Aging, Rotterdam, The Netherlands

[12]Institute for Community Medicine, University of Greifswald, Germany

[13]Institute of Infectious Diseases, Southwest Hospital, Third Military Medical University, Chongqing, China

[14]Institute of Genetic Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany

[15]Department of Neurology, Boston University School of Medicine, Boston Massachusetts, USA

[16]Department of Biological Psychology, VU University Amsterdam (VUA), Amsterdam, The Netherlands

[17]Laboratory of Molecular Medicine, Institute of Medical Science, The University of Tokyo, Tokyo, Japan

[18]Unit of Genetic Epidemiology and Bioinformatics, Department of Epidemiology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

[19]Wellcome Trust Center for Human Genetics, University of Oxford, Oxford, UK

[20]Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Churchill Hospital, Oxford, UK

[21]Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA

[22]Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, USA

[23]Medical Research Council (MRC) Epidemiology Unit, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge UK

[24]Institute of Social and Preventive Medicine (IUMSP), University Hospital and University of Lausanne, Lausanne, Switzerland

[25]Department of Cardiovascular Medicine, The Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK

[26]Computational Medicine Research Group, Institute of Clinical Medicine, University of Oulu and Biocenter Oulu, Oulu, Finland

[27]Nuclear Magnetic Resonance (NMR) Metabonomics Laboratory, Department of Biosciences, University of Eastern Finland, Kuopio, Finland

[28]Department of Internal Medicine, Division of Gastroenterology, University of Michigan, Ann Arbor, Michigan, USA

[29]Center for Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan, USA

[30]Genetics, GlaxoSmithKline, King of Prussia, Pennsylvania, USA

[31]Boston University School of Medicine, Boston, Massachusetts, USA

[32]Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, Saint Louis, Missouri, USA

[33]The Diabetes Inflammation Laboratory, Cambridge Institute of Medical Research, University of Cambridge, Cambridge, UK

[34]Department of Tropical Hygiene, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand

[35]Neuroscience Campus Amsterdam, VUA and VUA Medical Center, Amsterdam, The Netherlands

[36]National Heart and Lung Institute, Imperial College London, London, UK

[37]Institute of Epidemiology I, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany

[38]Center for Complex Disease Genomics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

[39]IUMSP, Centre Hospitalier Universitaire Vaudois and University of Lausanne, Lausanne, Switzerland

[40]Cardiology, Department of Medicine, Geneva University Hospital, Geneva, Switzerland

[41]The Laboratory in Mjodd, Reykjavik, Iceland

[42]Department of Cardiovascular Medicine, University of Oxford, John Radcliffe Hospital, Oxford, UK

[43]Institute of Clinical Chemistry and Laboratory Medicine, University of Greifswald, Germany

[44]Genetics and Gastroenterology Divisions, Brigham and Women's Hospital, Gastrointestinal Cancer Center, Dana-Farber Cancer Institute, Boston, Massachusetts, USA

[45]Harvard Medical School, Boston, Massachusetts, USA

[46]Harvard Stem Cell Institute, Cambridge, Massachusetts, USA

[47]Laboratory of Epidemiology, Demography, and Biometry, National Institute on Aging, US National Institutes of Health (NIH), Bethesda, Maryland, USA

[48]Institute of Clinical Medicine, University of Oulu, Oulu, Finland

[49]CEA-IG Centre National de Genotypage, Evry Cedex, France

[50]Department of Medicine, University of Toronto, Toronto, Ontario, Canada

[51]Liver Center, Toronto Western Hospital, Toronto, Ontario, Canada

[52]Centre for Liver Research, University of Birmingham, Birmingham, UK

[53]Interfaculty Institute for Genetics and Functional Genomics, University of Greifswald, Greifswald, Germany

[54]Centre for Paediatric Epidemiology and Biostatistics, Institute of Child Health, London, UK

[55]Department of Gastroenterology and Hepatology, Erasmus Medical Center, Rotterdam, The Netherlands

[56]Clinical Pharmacology and The Genome Center, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK

[57]Department of Laboratory Medicine, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

[58]Institute of Diagnostic Radiology and Neuroradiology, University of Greifswald, Greifswald, Germany

[59]Fondation Jean Dausset Ceph, Paris, France

[60]Department of Medicine A, University Medicine Greifswald, Greifswald, Germany

[61]Department of Genetics, Department of Biostatistics, University of North Carolina, Chapel Hill, North Carolina, USA

[62]Liver Diseases Branch, National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), NIH, Bethesda, Maryland, USA

[63]Office of Biostatistics Research, Division of Cardiovascular Sciences, National Heart, Lung and Blood Institute (NHLBI), NIH, Bethesda, Maryland, USA

[64]Queensland Institute of Medical Research, Brisbane, Queensland, Australia

[65]Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, USA

[66]Cardiovascular Research Center, Massachusetts General Hospital, Boston, Massachusetts, USA

[67]Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA

[68]Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

[69]NHLBI Framingham Heart Study, Framingham, Massachusetts, USA

[70]Department of Clinical Biochemistry, Landspitali University Hospital, Reykjavik, Iceland

[71]Department of Psychiatry and EMGO Institute for Health and Care Research, VUA Medical Centre, Amsterdam, The Netherlands

[72]Department of Psychiatry, Leiden University Medical Centre, Leiden, The Netherlands

[73]Department of Psychiatry, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

[74]Department of Lifecourse and Services, National Institute for Health and Welfare, Oulu, Finland

[75]Institute of Diagnostics, Clinical Chemistry, University of Oulu, Oulu, Finland

[76]Department of Internal Medicine and Biocenter Oulu, Clinical Research Center, University of Oulu, Oulu, Finland

[77]Laboratory of Genetics, National Institute on Aging, Baltimore, Maryland, USA

[78]Gesellschaft für Arteriosklerosеforschung, Leibniz-Institut für Arteriosklerosеforschung an der Universität Münster, Münster, Germany

[79]Department of Immunology, University of Toronto, Toronto, Ontario, Canada

[80]Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada

[81]Mount Sinai Hospital Samuel Lunenfeld Research Institute, Toronto, Ontario, Canada

[82]Toronto General Research Institute, Toronto, Ontario, Canada

[83]Department of Twin Research and Genetic Epidemiology, King's College London, London, UK

[84]Center for Statistical Genetics, Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, Michigan, USA

[85]Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands

[86]Department of Endocrinology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

[87]LifeLines Cohort Study and Biobank, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

[88]Section of Preventive Medicine and Epidemiology, Boston University School of Medicine, Boston, Massachusetts, USA

[89]Cardiology, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, USA

[90]Institute of Medical Informatics, Biometry and Epidemiology, Chair of Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany

[91]Klinikum Grosshadern, Munich, Germany

[92]EMGO^{+}Institute, VUA Medical Center, Amsterdam, The Netherlands

[93]Samuel Lunenfeld and Toronto General Research Institutes, Toronto, Ontario, Canada

[94]Division of Endocrinology, Diabetes and Nutrition, Department of Medicine, University of Maryland School of Medicine, Baltimore, Maryland, USA

[96]Center for Medical Systems Biology, Rotterdam, The Netherlands

[97]Genomics of Common Diseases, School of Public Health, Imperial College London, Hammersmith Hospital, London, UK

[98]Oxford National Institute for Health Research Biomedical Research Centre, Churchill Hospital, Oxford, UK

[99]Institute of Epidemiology II, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany

[100]Obesity Research Unit, Department of Medicine, Division of Internal Medicine, Helsinki University Hospital, Helsinki, Finland

[101]The Institute for Molecular Medicine FIMM, Helsinki, Finland

[102]Hjelt Institute, Department of Public Health, University of Helsinki, Helsinki, Finland

[103]MRC-Social Genetic Developmental Psychiatry (SGDP) Centre, Institute of Psychiatry, King's College, London, UK

[104]Department of Epidemiology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

[105]Faculty of Medicine, Imperial College London, London, UK

[106]Faculty of Medicine, University of Iceland, Reykjavik, Iceland

[107]Department of Internal Medicine, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland

[108]Division of Endocrinology, Hypertension, and Metabolism, Brigham and Women's Hospital, Boston, Massachusetts, USA

[109]Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, New York, USA

[110]Institute of Health Sciences, University of Oulu, Oulu, Finland

[111]Biocenter Oulu, University of Oulu, Oulu, Finland

[112]MRC–Health Protection Agency (HPA) Centre for Environment and Health, Imperial College London, London, UK

[95]A full list of members is found in Chambers et al. [3].

[113]These authors contributed equally to this work.

References

1 Willemsen G, de Geus EJC, Bartels M *et al*: The Netherlands Twin Register biobank: a resource for genetic epidemiological studies. *Twin Res Hum Genet* 2010; **13:** 231-245.

2 Penninx BWJH, Beekman ATF, Smit JH *et al*: The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods. *Int J Methods Psychiatr Res* 2008; **17:** 121-140.

3 Chambers JC, Zhang W, Sehmi J *et al*: Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. *Nat Genet* 2011; **43:** 1131-1138.

4 Yang J, Manolio TA, Pasquale LR *et al*: Genome partitioning of genetic variation for complex traits using common SNPs. *Nat Genet* 2011; **43**: 519-525.

5 Speliotes EK, Willer CJ, Berndt SI *et al*: Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* 2010; **42:** 937-948.

6 Altshuler DM, Lander ES, Ambrogio L *et al*: A map of human genome variation from population scale sequencing. *Nature* 2010; **467:** 1061-1073.

7 Fugita PA, Rhead B, Zweig AS: The UCSC Genome Browser database: update 2011. *Nucleic Acids Res* 2011; **39**: D876-D882.

8 Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: Using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol* 2010; **34**: 816-834.

9 Boomsma DI, Willemsen G, Sullivan PF *et al*: Genome-wide association of major depression: description of samples for the GAIN Major Depressive Disorder study: NTR and NESDA biobank projects. *Eur J Hum Gen* 2008; **16:** 335-342.

10 Yang J, Lee SH, Goddard ME, Visscher PM: GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 2011; **88:** 76-82.

11 IBM SPSS Statistics for Windows [Version 19.0]. Armonk, NY, IBM Corp, 2010.

12 Purcell S, Neale BM, Todd-Brown K *et al*: PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81:** 559-575.

13 Turner SD: qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *BioRxiv* 2014; http://biorxiv.org/content/early/2014/05/14/005165

14 Yang J, Benyamin B, McEvoy BP *et al*: Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* 2010; **42:** 565-569.

15 R: a language and environment for statistical computing. Vienna, Austria, 2012.

16 So HC, Li M, Sham PC: Uncovering the total heritability explained by all true susceptibility variants in a genome-wide association study. *Genet Epidemiol* 2011; **35:** 447-456.

17 Walters RK: Estimating variance explained by all DNA loci in a genome-wide meta-analysis [dissertation]. South Bend, IN, University of Notre Dame, 2014.