

ARTICLE

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

Jenny HDA van Beek<sup>\*,1,2</sup>, Gitta H Lubke<sup>1,3</sup>, Marleen HM de Moor<sup>1,2</sup>, Gonneke Willemsen<sup>1,2</sup>, Eco JC de Geus<sup>1,2,4</sup>, Jouke Jan Hottenga<sup>1,2,4</sup>, LivGen consortium<sup>6</sup>, Raymond K Walters<sup>3</sup>, Jan H Smit<sup>5</sup>, Brenda WJH Penninx<sup>5</sup> and Dorret I Boomsma<sup>1,2,4</sup>

Variation in the liver enzyme levels in humans is moderately heritable, as indicated by twin-family studies. At present, genome-wide association studies have traced <2% of the variance back to genome-wide significant single-nucleotide polymorphisms (SNPs). We estimated the SNP-based heritability of levels of three liver enzymes (gamma-glutamyl transferase (GGT); alanine aminotransferase (ALT); and aspartate aminotransferase (AST)) using genome-wide SNP data in a sample of 5421 unrelated Dutch individuals. Two estimation methods for SNP-based heritability were compared, one based on the distant genetic relatedness among all subjects as summarized in a Genetic Relatedness Matrix (GRM), and the other one based on density estimation (DE). The DE method was also applied to meta-analysis results on GGT and ALT. GRM-derived SNP-based heritability estimates were significant for GGT (16%) and AST (11%), but not for ALT (6%). DE estimates in the same sample varied as a function of pruning and were around 23% for all liver enzymes. Application of the DE approach to meta-analysis results for GGT and ALT gave SNP-based heritability estimates of 6 and 3%. The significant results in the Dutch sample indicate that genome-wide SNP platforms contain substantial information regarding the underlying genetic variation in the liver enzyme levels. A major part of this genetic variation remains however undetected. SNP-based heritability estimates, based on meta-analysis results, may point at substantial heterogeneity among cohorts contributing to the meta-analysis. This type of analysis may provide useful information to guide future gene searches.

*European Journal of Human Genetics* (2015) 23, 1223–1228; doi:10.1038/ejhg.2014.259; published online 26 November 2014

## INTRODUCTION

Concentrations of the liver enzymes gamma-glutamyl transferase (GGT), alanine, and aspartate aminotransferase (ALT, AST) predict liver disease and all-cause mortality.<sup>1,2</sup> Clinically, these enzymes are used as markers for liver injury.<sup>3</sup>

Variation in the liver enzyme levels can be partly explained by genetic differences among individuals. Broad-sense heritability estimates for liver enzyme levels from twin-family studies range from 22–60%.<sup>4</sup> Although genetic influences on liver enzyme levels are substantial, most of the genes underlying the variation are still unidentified. For GGT, adding the effects of all genome-wide significant single-nucleotide polymorphisms (SNPs with  $P$ -values  $<5 \times 10^{-8}$ ) explains 2% of the variation, for ALT and AST this is  $<1\%$ .<sup>5,6</sup>

Several explanations for this so-called ‘missing heritability’<sup>7</sup> have been put forward. The sample sizes of current genome-wide association (GWA) studies might be too small to detect the effects of individual SNPs under the stringent significance thresholds that are used to correct for multiple testing.<sup>8</sup> Alternatively, genetic variation may be due to effects other than those captured by SNPs on current genotyping platforms (eg, rare variants or copy number variants).<sup>9</sup> These causes of missing heritability may well differ between phenotypes.

To gain insight in the genetic architecture of the liver enzyme levels and optimize the success of future gene-finding studies, it is important to know to what extent missing heritability is due to inadequate power to find small SNP effects, and to what extent it is due to SNP platforms not containing relevant information. By examining the proportion of the variation in liver enzyme concentrations that can be explained by the joint effect of all measured and imputed genome-wide SNPs, it can be tested to what extent the heritability is hidden among existing SNP platforms instead of missing. The degree to which this estimate is higher than the proportion of variance that is currently explained by genome-wide significant SNPs most probably reflects associations that have not yet been detected because of the multiple testing burden.

The first aim of the current study is to apply two alternative methods to study the aggregate effect of all SNPs on phenotypic variability in GGT, ALT, and AST levels. The first method is a two-step procedure where the first step consists of estimating the genetic relatedness matrix (GRM) between all pairs of subjects. This pair-wise genetic relatedness is similar to a correlation between two individuals using all SNPs. In the second step, the pair-wise genetic relatedness is used as a random effect in a linear mixed model to estimate the proportion of variance attributable to additive genetic effects. This method, denoted as the GRM method in this paper, is

<sup>1</sup>VU University Amsterdam, Department of Biological Psychology, Amsterdam, The Netherlands; <sup>2</sup>VU University Medical Center, EMGO Institute for Health and Care Research, Amsterdam, The Netherlands; <sup>3</sup>University of Notre Dame, Department of Psychology, Notre Dame, IN, USA; <sup>4</sup>Neuroscience Campus Amsterdam, Amsterdam, The Netherlands; <sup>5</sup>VU University Amsterdam, Department of Psychiatry, Amsterdam, The Netherlands

\*Correspondence: Dr JHDA van Beek, VU University Amsterdam, Department of Biological Psychology, Van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands. Tel: +31 20 598 8787; Fax: +31 20 598 8832; E-mail: jennyvbeek@gmail.com

<sup>6</sup>A full list of members is given in the Supplementary Materials.

Received 21 December 2013; revised 18 October 2014; accepted 22 October 2014; published online 26 November 2014

implemented in the software package Genome-wide Complex Trait Analysis (GCTA).<sup>10,11</sup>

An alternative approach to estimate the proportion of variance that can be ascribed to SNP effects is the density estimation (DE) method proposed by So *et al.*<sup>12</sup> As the DE method uses summary statistics from a GWA analysis, it does not require raw SNP data. Therefore, it can also be applied to regression coefficients or *P*-values obtained in meta-analyses. The DE method compares the distribution of observed effect sizes of SNPs that resulted from a GWA study (or meta-analysis of GWA studies) to the distribution expected under the null hypothesis of zero effects. The extent to which the distribution of observed effect sizes has thicker tails than the distribution under the null reflects the proportion of phenotypic variance that is captured by SNPs. Specifically, the proportion of phenotypic variance explained is estimated from 'true' effect sizes computed using a correction for sampling variation suggested by Efron.<sup>3</sup> To avoid inflated estimates due to SNPs with non-zero effects that are in linkage disequilibrium (LD), the SNP data need to be pruned to obtain independent SNP signals. The phenotypic variance of continuous phenotypes due to SNPs is calculated using a sums of squares approach similar to ANOVA.<sup>12</sup>

Application of the two methods to a range of phenotypes shows that about 30–50% of the classic heritability estimated from twin-family data is recovered.<sup>14,15</sup> Less well-known is what proportion of variation is recovered if the DE approach is applied to meta-analysis results of GWAs. Therefore, the second aim of the study is to compare DE estimates on the SNP-based heritability for a single sample (data from the Netherlands Twin Register (NTR) and Netherlands Study on Depression and Anxiety (NESDA)) with DE estimates based on GWA meta-analysis results. Note that the GRM method can only be used for meta-analyses if raw SNP data are available for all cohorts, which is rarely the case. The GCTA package provides meta-analysis methods related to DE, but at the moment does not include Efron's correction for sampling variation and assumes that LD among SNPs can be accurately estimated.<sup>13</sup>

Data for this study originate from (a) participants of the NTR study ( $N=3309$  unrelated subjects),<sup>16</sup> (b) participants of the NESDA study ( $N=2212$  unrelated subjects),<sup>17</sup> and (c) available summary statistics from a meta-analysis on GGT and ALT ( $N=61\,089$ ) by an international consortium.<sup>5</sup> To compare the performance of the DE method, SNP-based heritability estimates were also estimated for BMI. BMI served as a bench-mark trait as its additive genetic variance explained by SNPs has been studied before.<sup>18</sup>

## MATERIALS AND METHODS

### Participants

Data came from 5421 unrelated individuals from European descent who participated in the NTR biobank<sup>19</sup> or NESDA<sup>17</sup> study and for whom valid genotype data and data on one or more liver enzyme concentrations were available (NTR:  $N=3309$ ; 60.6% females; year of birth 1914–1987; NESDA:  $N=2112$ ; 66.6% females; 1939–1988). See the Supplementary Materials for a full description of inclusion and exclusion criteria. 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.<sup>17,19</sup>

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 1721 NTR and 1724 NESDA participants; and 8739 participants with an Indian-Asian background.<sup>5</sup>

For NTR/NESDA participants, data on BMI ( $N=5406$ ) were assessed at the same time as their liver enzyme data. Meta-analysis summary statistics (*P*-values) from large GWA studies on BMI<sup>20</sup> ( $N=249\,796$ ) were downloaded

from [http://www.broadinstitute.org/collaboration/giant/index.php/GIANT\\_consortium\\_data\\_files](http://www.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files). Supplementary Table S1 gives a summary of all data available for each phenotype.

### Genotyping and quality control

Genotyping in the combined NTR/NESDA sample was performed on five platforms: Affymetrix 6.0, Affymetrix 5.0-Perlegen (Affymetrix, Santa Clara CA, USA), Illumina 660, Illumina Omni Express 1 M, and Illumina 370 (Illumina, San Diego, CA, USA). The final data set after SNP imputation and data quality control (described in the Supplementary Materials) consisted of 5 994 956 autosomal SNPs.

### Phenotypes

Liver enzymes were determined in heparin plasma (see Supplementary Materials) 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.<sup>21</sup>

### Statistical analyses

*Creating sample of unrelated individuals.* To create a sample of independent individuals, a GRM was first estimated (option `--make-grm`) for all NTR and NESDA individuals with valid liver enzyme level and genotype data, using the free software package GCTA (v1.24.2).<sup>11</sup> This GRM matrix was then pruned for relatedness at a level of 0.025 (option `--grm-cutoff 0.025`), resulting in a set of 5421 individuals with estimated pair-wise relatedness  $<0.025$ .

*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. On the basis of these analyses, regression analyses were carried out (see Supplementary Materials, for a list of predictors) in SPSS 19.0.<sup>22</sup> The residuals from these regression analyses were used in all subsequent analyses.

*GWA.* In the NESDA/NTR data set, SNP associations were tested in a linear model assuming additive SNP effects using Plink (v1.07).<sup>23</sup> GWA results are the input for the methods to estimate heritability and were inspected by QQ and Manhattan plots.

### SNP heritability based on the NTR/NESDA sample

*GRM method.* A linear mixed model was used to estimate the phenotypic variance that is due to the genetic relatedness captured by the GRM using GCTA (v1.24.2).<sup>11</sup> Estimation was performed using restricted maximum likelihood (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.

*DE method.* Analyses with the DE method were performed in R3.0.2<sup>24</sup> with the script for continuous traits obtained from the developer's website: <https://sites.google.com/site/honcheongso/software/total-vg>. See the Supplementary Materials for a detailed description of the DE method. To obtain independent SNP signals, the data set was pruned at a level of  $r^2 0.25$  as suggested by So *et al.*<sup>12</sup> (`--indep-pairwise 100, 25, 0.25`), resulting in a set of 226 243 SNPs. As 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 data set was pruned 10 times using the same pruning parameters. 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 compared with meta-analysis results

*DE method.* To compare SNP 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 data sets. This pruned set consisted of SNPs that were present in the GGT and ALT meta-analyses as well as in the NTR/NESDA data set. Pruning was based on the LD pattern among SNPs in the NTR/NESDA data set, and was performed at a level of  $r^2 0.25$  as suggested by So *et al.*<sup>12</sup> (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, as here a data set of ~2.7 M SNPs was pruned at  $r^2$  0.25; whereas for the comparison with GRM-based estimates, the entire data set (containing ~6 M SNPs) was pruned at  $r^2$  0.25. SNP markers in the GWA meta-analyses were imputed against build 36 (HG18) of the Human reference genome,<sup>5,20</sup> and lifted over to build 37 (HG19). The latter was the reference for NTR/NESDA (see Supplementary Materials). To verify that the DE estimates did not depend on a specific pruned set of SNPs, the NTR/NESDA SNP data set was pruned 10 times, and DE estimates were averaged over these 10 pruned sets.

An overview of all analyses is included in Supplementary Table S1. SNP heritability estimates, obtained with the GRM method, were considered to be significant if  $P$ -values  $< 0.05$ . In the case that GRM- and DE-based estimates differed, we applied a conservative approach by focusing on the lower of the two estimates as Walters<sup>25</sup> has shown that DE-based heritability estimates could be overestimated when sample size is small.

## RESULTS

Table 1 summarizes the mean (with standard error) and median liver enzyme levels (with range) for NTR and NESDA, separately over sex. Mean liver enzyme levels were higher for men than women. For both sexes, GGT levels were higher in NTR than in NESDA participants, whereas for ALT and AST, the reverse was observed (see Table 1). Supplementary Table S2 shows the correlations among liver enzyme levels, by source and sex. There were positive correlations among the three liver enzyme levels ranging from 0.26 to 0.66. The correlation between GGT and ALT was higher for NESDA than NTR participants in males (0.53 vs 0.26) and females (0.53 vs 0.30). Correlations of AST with GGT and ALT were similar in NTR and NESDA (~0.33 and ~0.60 respectively, for both sexes).

Supplementary Figures S1A–4A show the QQ plots with  $P$ -values for the SNP associations for liver enzyme levels and BMI that resulted from GWA analyses performed on the NTR/NESDA data set. Supplementary Figures S1C–4C show the QQ plots for the downloaded meta-analysis results for GGT, ALT, and BMI. In line with the published results, these show that the observed  $P$ -values strongly deviated from the line with expected  $P$ -values, indicating large polygenic variation for GGT, ALT, and BMI (Supplementary Figures S1C, S2C, and S4C, respectively). For the NTR/NESDA data, observed  $P$ -values for GGT and BMI also show a strong deviation from the line with expected  $P$ -values (Supplementary Figures S1A and S4A) whereas this deviation was much weaker for ALT and AST (Supplementary Figures S2A and S3A). Manhattan plots for the NTR/NESDA data set and the GWA meta-analysis data for these phenotypes are shown in Supplementary Figures S1–4B and S1–4D.

### SNP heritability based on the NTR/NESDA sample

*GRM and DE method.* Table 2 shows the GRM- and DE-based estimates for the variance explained by SNPs for liver enzyme levels.

BMI is included for comparison. A significant proportion of GGT (16%;  $P=0.002$ ), AST (11%;  $P=0.018$ ), and BMI (15%;  $P=0.003$ ) was explained by SNPs according to the GRM method. For ALT, this was 6% (NS). Results obtained with the DE method were significantly higher (38%, 38%, 34%, and 38% for GGT, ALT, AST, and BMI, respectively; falling outside the two standard error range of the GRM-based estimates using GRM standard errors). These estimates were somewhat higher than the narrow-sense heritability estimates from twin-family studies for these phenotypes<sup>4,26</sup> (see Table 2). Noting the potential bias in DE-based estimates at small sample sizes,<sup>25</sup> we conservatively focus on the lower GRM-derived estimates. These GRM-derived estimates of SNP heritability were lower than twin-family-based estimates of narrow-sense heritability. The difference is at least in part due to imperfect LD between causal SNPs and the SNPs included in the analysis.<sup>10</sup> Additional analyses with the GRM method showed that for GGT, chromosomes 2, 3, 10, 20, and 22; and for AST, chromosomes 2 and 6 explained a significant part of the variance of the variance at  $P < 0.05$  (see Supplementary Table S3). After correction for multiple testing ( $P < 0.002$ ; 0.05 divided by 22 chromosomes), none of the chromosomes separately explained a significant part of the variance of liver enzyme levels. After correction for multiple testing ( $P < 0.002$ ; 0.05 divided by 22 chromosomes), none of the chromosomes separately explained a significant part of the variance of liver enzyme levels.

One explanation for the high DE estimates is that the level of SNP pruning necessary to obtain independent SNP signals, is dependent on SNP density in the genotype set. Then, the level of pruning recommended by So *et al*<sup>12</sup> ( $r^2$  0.25), which was based on data sets containing  $\leq 2.7$  M SNPs, would not be appropriate for the NTR/NESDA data set (containing ~6 M SNPs). Additional analyses indicated that DE estimates were lower when the number of SNPs in the data set was in line with those studied by So *et al*<sup>12</sup> and/or when the pruning threshold was lower. When pruning a data set of 2.7 M SNPs at  $r^2$  0.25, resulting in a pruned set of 111 995 SNPs, DE estimates were ~23% (see Table 3). When pruning the entire data set (containing ~6 M SNPs) at a very stringent level of SNP pruning ( $r^2$  0.001; resulting in a set of nearly independent SNPs), DE estimates were 13, 11, and 15% for GGT, ALT, and AST, respectively. These more conservative estimates agree rather well with the GRM-based estimates given above (estimates fell within two standard error range around the GRM estimates) giving further support for the GRM-based estimates.

### SNP heritability in the single sample compared with meta-analysis results

For both GGT and ALT, DE-derived SNP-based heritability was 23% in the NTR/NESDA sample, when using a set of 2.7 M SNPs pruned at  $r^2$  0.25. Estimates based on the consortium data were 6% for GGT

**Table 1** Descriptive statistics of liver enzyme levels<sup>a</sup> BMI, and age, split over source (NTR and NESDA) and sex

	NTR				NESDA			
	Males		Females		Males		Females	
	Mean (SE)	Median (range)	Mean (SE)	Median (range)	Mean (SE)	Median (range)	Mean (SE)	Median (range)
GGT	41.1 (1.20)	31 (10–917)	28.3 (0.73)	21 (9–867)	36.2 (1.42)	24.2 (7–285)	20.4 (0.65)	15 (2–563)
ALT	11.6 (0.20)	10 (3–107)	9.3 (0.14)	8 (3–100)	30.9 (0.72)	26 (1–218)	19.8 (0.32)	17 (4–248)
AST	23.4 (0.24)	22 (7–122)	20.1 (0.16)	19 (7–142)	29.4 (0.45)	26.8 (10–112)	23.5 (0.22)	22 (8–94)
BMI	26.0 (0.10)	25.7 (17.3–40.5)	25.5 (0.10)	24.6 (15.7–46.5)	26.3 (0.17)	25.9 (16.0–50.2)	25.2 (0.14)	24.1 (14.7–53.3)
Age	51.0 (0.40)	56 (18–86)	47.0 (0.31)	50.0 (20–90)	44.4 (0.47)	46 (18–64)	41.7 (0.35)	43 (18–65)

<sup>a</sup>This table shows untransformed liver enzyme levels; statistical analyses were performed on log-transformed levels of liver enzymes.

**Table 2 GRM-based estimates (with standard errors) and DE-based estimates<sup>a</sup> on the proportion variance explained by all SNPs for liver enzyme levels and BMI**

	GGT		ALT		AST		BMI	
	GRM	DE	GRM	DE	GRM	DE	GRM	DE
NTR+NESDA	0.155** (0.056)	0.376	0.055 (0.055)	0.377	0.111* (0.055)	0.337	0.149** (0.056)	0.379
% variance explained by GWAS	2% <sup>b</sup>		<1% <sup>b</sup>		<1% <sup>b</sup>		1.5% <sup>b</sup>	
Narrow-sense $h^2$ twin-family study <sup>c</sup>	0.30 (0.24–0.37)		0.29 (0.24–0.33)		0.28 (0.23–0.34)		0.40 (0.37–0.43)	
Broad-sense $h^2$ twin-family study <sup>c</sup>	0.30 (0.05) (males)		0.40 (0.05) (males)		0.43 (0.03)		0.85 (0.01) (males)	
	0.60 (0.03) (females)		0.22 (0.03) (females)		(males+females)		0.75 (0.01) (females)	

\*/\*\*GRM estimates that were significant at  $P < 0.05$  (\*) and  $P < 0.01$  (\*\*). Significance was not calculated for DE estimates as the DE method does not provide standard errors.

<sup>a</sup>DE-derived estimates were based on the NTR/NESDA data set containing ~6M SNPs that was pruned at an  $r^2$  level of .25 (recommended by So *et al*<sup>12</sup>). As pruning the dense NTR/NESDA data set (~6 M SNPs) at  $r^2$  0.25 resulted in an overestimation of the SNP-based heritability (when compared with narrow-sense heritability estimates), we also calculated the DE-based estimates on SNP heritability, by pruning the NTR/NESDA data set (containing ~6M SNPs) at  $r^2$  0.001, resulting in a set of nearly independent SNPs. These more conservative DE estimates are 0.129, 0.112, 0.148, and 0.126 for GGT, ALT, AST, and BMI, respectively. See Supplementary Materials and text for details.

<sup>b</sup>Estimates based on Chambers *et al*,<sup>5</sup> Kamatani *et al*,<sup>6</sup> and Speliotes *et al*,<sup>20</sup> respectively.

<sup>c</sup>Heritability of liver enzyme levels that can be ascribed to additive genetic effects (narrow-sense heritability) and additive+non-additive genetic effects (broad-sense heritability) as estimated in ACDE (GGT), AE (ALT), and ADE (AST, BMI) models in twin-family data on liver enzyme levels (Van Beek *et al*)<sup>4</sup> and BMI in the NTR biobank sample. For reasons of clarity, narrow-sense heritability estimates are constrained to be equal over sex in this table.

**Table 3 Comparison of DE-derived estimates (with estimates of variability<sup>a</sup>) of explained variance based on GWA results for a single sample (NTR/NESDA) vs GWA meta-analysis results based on multiple samples, for liver enzyme levels, and BMI**

	Selection SNPs from	No. of SNPs entire set	pruning level ( $r^2$ )	No. of SNPs pruned set	DE estimates explained variance			
					GGT	ALT	AST	BMI
NTR+NESDA	1000 Genomes b37	~6 M	0.25	226 243	0.376 (0.039)	0.377 (0.042)	0.337 (0.035)	0.379 (0.038)
NTR+NESDA	Hapmap b36	~2.7 M	0.25	111 995	0.234 (0.023)	0.229 (0.015)	0.234 (0.027)	0.277 (0.033)
Consortium meta-analysis <sup>b</sup>	Hapmap b36	~2.7 M	0.25	111 995	0.060 (0.003)	0.028 (0.002)		0.079 (0.001)

<sup>a</sup>Note that this estimate of variability should not be interpreted as a standard error (see Supplementary Materials for details). <sup>b</sup>The SNP-based heritability estimate (7.9%) for BMI was obtained by pruning the NTR/NESDA data set filtering on SNPs that were included in the GWA meta-analysis by Chambers *et al*<sup>5</sup> on GGT and ALT. When pruning the NTR/NESDA SNP data set after filtering on SNPs that were included in the GWA meta-analysis on BMI by Speliotes *et al*<sup>20</sup> (resulting in a pruned set of 109 120 SNPs), SNP-based heritability was 8.4%. DE-based estimates were slightly higher when using the (pruned) Hapmap b36 data set obtained from the Plink website <http://pngu.mgh.harvard.edu/~purcell/plink/res.shtml#hapmap>. For GGT and ALT, this was 7.9% and 4.7%, respectively (based on a pruned set of ~215 000 SNPs). For BMI, this was 9.9% (based on a pruned set of ~190 000 SNPs).

and 3% for ALT when using the same pruned set; see Table 3. The DE estimate based on the meta-analysis results for BMI was 8% vs 28% in the NTR/NESDA sample. Thus, DE-based estimates for the GWA meta-analysis results for GGT, ALT, and BMI were much lower than those on GWA results based on the single NTR/NESDA sample. A potential cause for the low estimates using meta-analysis effect sizes in the DE method is heterogeneity across the individual cohorts in the GWA meta-analysis. However, it should be noted that the meta-analysis estimates are within the confidence intervals of the single sample GRM estimates for GGT, ALT, and BMI.

## DISCUSSION

The current study aimed at estimating the proportion of variance of liver enzyme concentrations that can be explained by measured and imputed genome-wide SNPs in a single Dutch sample, and second, to compare this estimate to SNP-based heritability using GWA meta-analysis samples.

A significant proportion of the phenotypic variance of GGT (16%) and AST (11%) in the NTR/NESDA sample can be explained by additive SNP effects, based on the GRM method. For ALT, the GRM-based estimate on SNP heritability of 6% did not reach statistical significance. These GRM-based estimates of SNP heritability were lower than additive genetic variance estimated in twin and family studies. This was expected, and is at least partially due to imperfect LD and allelic frequency differences between causal SNPs and SNPs used

in the analyses.<sup>10</sup> The difference might also partially be due to current SNP platforms missing some of the relevant information. However, our significant findings underline the usefulness of SNP data in genetic analyses.

DE-based estimates for the NTR/NESDA sample (when pruning the entire ~6 M SNP set) were 38%, 38%, and 34% for GGT, ALT, and AST, respectively. These estimates were higher than GRM-based estimates, and also somewhat higher than narrow-sense heritability estimates based on twin-family studies.<sup>4</sup> Most likely, these high DE estimates can be explained by the fact that the appropriate level of SNP pruning is dependent on SNP density. On the one hand, the DE method requires a set of approximately independent SNPs. On the other hand, very conservative pruning increases the probability of removing tagging SNPs that are in LD with causal SNPs, thus resulting in a SNP density that is too low to obtain a correct estimate. To illustrate this trade-off, pruning the NTR/NESDA data set (containing ~6 M SNPs) at an  $r^2$  level of 0.001 (instead of  $r^2$  0.25) resulted in a set of 37 389 nearly independent SNPs. The resulting DE estimates in the NTR/NESDA sample were 13%, 11%, and 15% for GGT, ALT, and AST, respectively. These estimates agreed relatively well with the GRM-based estimates for the same phenotypes. Ongoing work with simulated data confirms the impact of the pruning threshold (Walters & Lubke, in preparation).

SNP-based heritability estimates using GWA meta-analysis statistics were higher than the amount of phenotypic variance of GGT and ALT



that is currently explained by genome-wide significant SNPs (<2%).<sup>5</sup> DE-derived estimates of SNP heritability based on GWA meta-analysis were lower than those for GWA results based on a single sample (NTR/NESDA) (GGT 6% and ALT 3% vs 23% and 23%, respectively). This underestimation when using meta-analysis data was also found for BMI (8% vs 28% in the NTR/NESDA sample and 16% in previous research<sup>18</sup>). It should be noted, however, that the DE estimates using meta-analysis data fall within the confidence intervals of the GRM estimates for GGT, and ALT in the NTR/NESDA sample.

The large difference between meta-analysis DE estimates and single sample results remained when pruning was based on the LD pattern in the Hapmap 2 reference set (CEU sample; used for imputation in the individuals cohorts in the GWA meta-analyses; see footnote Table 3). Allelic differences between the NTR/NESDA data set and those in the GWA meta-analysis sets, thus, cannot explain the large difference between the single sample DE estimates and those based on the GWA meta-analysis.

A first explanation for the low SNP heritability estimates based on GWA meta-analysis results is heterogeneity among the samples included in the meta-analysis. If not taken into account, this will lead to a lower amount of variance that can be explained by SNPs.<sup>27</sup> In the case of genetic heterogeneity, if SNP  $x$  has an effect in sample 1 (eg, standardized beta,  $b=0.4$ ) but not in sample 2 (standardized  $b=0$ ), the meta-analysis (average) effect size of this SNP is halved (standardized  $b=0.2$ ). When the effect size of SNP  $x$  is halved, its explained variance is reduced to one quarter ( $0.4^2$  when  $b=0.4$ ;  $0.2^2$  when  $b=0.2$ ), as the standardized beta is equal to the square root of the total explained variance. Thus, 75% of the variance due to SNP  $x$  is lost in the case of heterogeneity between sample 1 and 2 (when the meta-analysis effect size is analyzed instead of that based on sample 1) (PC Sham, personal communication). Genetic heterogeneity will thus decrease the proportion of effects in the extreme upper and lower tails of the distribution. The distribution of effects (expressed in  $z$ -scores) is the input for the DE method and this will thus lead to lower DE estimates of explained variance.

It might be argued that the heterogeneity explanation of low heritability estimates when using meta-analysis data is not consistent with the large polygenic variation that is evident from the QQ plots for these meta-analysis samples (Supplementary Figures S1–4C). However, the deviation of observed  $P$ -values in these QQ plots is reflecting both effect size and sample size, meaning that large deviations can reflect small effect sizes if sample size is large enough. In contrast, DE estimates are based on observed effects that are corrected for sampling fluctuation to get 'true' effect sizes. Thus, the deviation of observed  $P$ -values that is evident in the QQ plots will only to some extent be picked up by the DE method.

Simulation studies suggested that the lower DE estimates for the meta-analysis samples could not be attributed to the DE method being dependent on sample size or sensitive to the distribution of effect sizes.<sup>25</sup> When the true population was simulated to consist of 30 000 individuals, drawing samples of 3000 individuals each did not result in overestimated proportions of variance explained by SNPs. Simulating data under the assumption that the distribution of SNP effect sizes was exponential with small effects for SNPs, which are relatively common and large effects for SNPs with low MAF, did not lead to distorted DE-based estimates.<sup>25</sup> Given the results from this simulation study, our low DE-based estimates for the GWA meta-analyses are strongly suggestive of effects of genetic/phenotypic heterogeneity.

An additional explanation that DE estimates based on GWA meta-analysis results are downward biased is that genomic control correction affected the SNP associations. In most meta-analyses,  $P$ -values are

corrected for the genomic control inflation factor ( $\lambda_{GC}$ ), and often double corrected (eg, see Speliotes *et al*<sup>20</sup>). As  $P$ -values are the direct input for the DE method, downwards adjustment of the  $P$ -values will result in lower DE estimates. For the current study, the DE estimates based on the meta-analysis results are based on  $P$ -values that were uncorrected for the overall genomic inflation factor correction (see Supplementary Materials). Nevertheless, to the extent that the first study-specific genomic control correction has affected the SNP associations, the DE estimates for meta-analysis data will be underestimated.

The GRM and DE methods to estimate SNP heritability are constantly being extended and improved.<sup>28</sup> One weakness of the DE method is the lack of standard errors. In this study, an indication of the stability/variability of the DE estimates across different sets of SNPs was obtained by repeating the DE method on 10 different pruned sets. However, this should not be regarded as an approximation of a standard error, but rather as an indication that the DE results do not depend much on which SNPs are used in the estimation of heritability. Ongoing research by some of the co-authors focuses on obtaining standard errors for the DE method.

Future research should explore how the results on the low SNP heritability estimates based on GWA meta-analysis results can inform future GWA studies. If the low DE estimates for GWA meta-analysis results can be accounted for by genetic heterogeneity, this calls for taking genetic heterogeneity into account when combining data from several studies.

Additional work is also needed to explore the merits and limitations of the GRM and DE methods. With regard to the DE method, the level of SNP pruning that was suggested by So *et al*<sup>12</sup> seems to be dependent on SNP density, and future research should explore whether the performance of the DE method can be further improved when the optimal level of pruning is considered to be a meta parameter whose value needs to be set through cross validation guided by the prediction error. The performance of the DE and GRM method can be compared with newly developed methods to estimate the amount of variance explained by SNPs, such as those that incorporate improvements on the GRM method,<sup>28,29</sup> other means to estimate and sum 'true' effect sizes for SNPs in pruned SNP sets,<sup>30</sup> and methods using a Bayesian approach.<sup>31</sup>

## CONCLUSION

To conclude, our results show that genome-wide SNP platforms contain substantial information regarding the underlying genetic variation in liver enzyme levels. Adequate sample sizes may therefore lead to the detection of new susceptibility loci, which in turn may elucidate new biological pathways underlying liver enzyme concentrations.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

- 1 Kazemi-Shirazi L, Ender G, Winkler S *et al*. Gamma glutamyltransferase and long-term survival: is it just the liver? *Clin Chem* 2007; **53**: 940–946.
- 2 Hyeon CK, Chung MN, Sun HJ *et al*. Normal serum aminotransferase concentration and risk of mortality from liver diseases: prospective cohort study. *BMJ* 2004; **328**: 983.
- 3 Pratt DS, Kaplan MM. Evaluation of abnormal liver-enzyme results in asymptomatic patients. *N Eng J Med* 2000; **342**: 1266–1271.
- 4 Van Beek JHDA, De Moor MHM, de Geus EJC *et al*. The genetic architecture of liver enzyme levels: GGT, ALT and AST. *Behav Genet* 2013; **43**: 329–339.

- 5 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.
- 6 Kamatani Y, Matsuda K, Okada Y *et al*: Genome-wide association study of hematological and biochemical traits in a Japanese population. *Nat Genet* 2010; **42**: 210–215.
- 7 Maher B: The case of the missing heritability. *Nature* 2008; **456**: 18–21.
- 8 Purcell SM, Wray NR, Stone JL *et al*: Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 2009; **460**: 748–752.
- 9 Stranger BE, Stahl EA, Raj T: Progress and promise of genome-wide association studies for human complex trait genetics. *Genetics* 2011; **187**: 367–383.
- 10 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.
- 11 Yang J, Lee SH, Goddard ME *et al*: GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 2011; **88**: 76–82.
- 12 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.
- 13 Efron B: Empirical Bayes estimates for large-scale prediction problems. *J Am Stat Assoc* 2009; **104**: 1015–1028.
- 14 Vinkhuyzen AA, Wray NR, Yang J *et al*: Estimation and partitioning of heritability in human populations using whole genome analysis methods. *Annu Rev Genet* 2013; **47**: 75–95.
- 15 Lubke GH, Hottenga JJ, Walters R *et al*: Estimating the genetic variance of major depressive disorder due to all single nucleotide polymorphisms. *Biol Psychiatry* 2012; **72**: 707–709.
- 16 Boomsma DI, Vink JM, Van Beijsterveldt CEM *et al*: Netherlands Twin Register: a focus on longitudinal research. *Twin Res Hum Genet* 2002; **5**: 401–406.
- 17 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.
- 18 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.
- 19 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.
- 20 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.
- 21 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.
- 22 IBM Corp. Released 2010: IBM SPSS statistics for Windows, Version 19.0. Armonk, NY: IBM Corp. 2010.
- 23 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.
- 24 R Core Team: R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. 2012.
- 25 Walters RK Estimating variance explained by all DNA loci in a genome-wide meta-analysis [dissertation]. South Bend, IN, University of Notre Dame 2014.
- 26 van Dongen J, Willemsen G, Chen W-M *et al*: Heritability of metabolic syndrome traits in a large population-based sample. *J Lipid Res* 2013; **54**: 2914–2923.
- 27 Manchia M, Cullis J, Turecki G *et al*: The impact of phenotypic and genetic heterogeneity on results of genome wide association studies of complex diseases. *PLoS One* 2013; **8**: e76295.
- 28 Zaitlen N, Kraft P, Patterson N *et al*: Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits. *PLoS Genet* 2013; **9**: e1003520.
- 29 Speed D, Hemani G, Johnson M, Balding DJ: Improved heritability estimation from genome-wide SNPs. *Am J Hum Genet* 2012; **91**: 1011–1021.
- 30 Kutalik Z, Whittaker J, Waterworth D, Beckmann JS, Bergmann S: Novel method to estimate the phenotypic variation explained by genome-wide association studies reveals large fraction of the missing heritability. *Genet Epidemiol* 2011; **35**: 341–349.
- 31 Stahl EA, Wegmann D, Trynka G *et al*: Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nat Genet* 2012; **44**: 483–489.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)