SNP and Haplotype Regional Heritability Mapping (SNHap-RHM): joint mapping of common and rare variation affecting complex traits

Richard F. Oppong[1], Pau Navarro[2], Chris S. Haley[2,3] and Sara Knott[1]

[1]IEB, SBS, University of Edinburgh, Edinburgh, United Kingdom,

[2]MRC IGMM, University of Edinburgh, Edinburgh, United Kingdom,

[3]The Roslin Institute and Royal (Dick) School of Veterinary Studies, Edinburgh, United Kingdom.

Correspondence: s.knott@ed.ac.uk

Abstract

We describe a genome-wide analytical approach, SNP and Haplotype Regional Heritability Mapping (SNHap-RHM), that provides regional estimates of the heritability across locally defined regions in the genome. This approach utilises relationship matrices that are based on sharing of SNP and haplotype alleles at local haplotype blocks delimited by recombination boundaries in the genome. We implemented the approach on simulated data and show that the haplotype-based regional GRMs capture variation that is complementary to that captured by SNP-based regional GRMs, and thus justifying the fitting of the two GRMs jointly in a single analysis (SNHap-RHM). SNHap-RHM captures regions in the genome contributing to the phenotypic variation that existing genome-wide analysis methods may fail to capture. We further demonstrate that there are real benefits to be gained from this approach by applying it to real data from about 20,000 individuals from the Generation Scotland: Scottish Family Health Study. We analysed height and major depressive disorder (MDD). We identified seven genomic regions that are genome-wide significant for height, and three regions significant at a suggestive threshold (p-value $< 1 \times 10^{-5}$) for MDD. These significant regions have genes mapped to within 400kb of them. The genes mapped for height have been reported to be associated with height in humans, whiles those mapped for MDD have been reported to be associated with major depressive disorder and other psychiatry phenotypes. The results show that SNHap-RHM presents an exciting new opportunity to analyse complex traits by allowing the joint mapping of novel genomic regions tagged by either SNPs or haplotypes, potentially leading to the recovery of some of the "missing" heritability.

Keywords: MDD; height; haplotypes; regional heritability mapping; missing heritability; rare variation; genome-wide analysis

Introduction

Estimates of the genetic component of complex trait variation using genotyped SNPs led to the conclusion that a proportion of the heritability of complex traits is still unexplained or "missing" (Maher, 2008; Manolio et al., 2009). Full sequence data will contain all the variants that account for all the heritability of complex traits (Wainschtein et al., 2019). Moreover, some of these true causal variants may be rare (Pritchard, 2001) and therefore may be in incomplete linkage disequilibrium (LD) with genotyped SNPs (Yang et al., 2010). Thus, some of the "missing" heritability may be "hidden" in rare variants whose effects are difficult to capture because of lack of statistical power. There is, therefore, some benefit to be gained in terms of improving the heritability estimates and uncovering gene variants involved in the control of traits by fitting genome-wide analytical models that adequately capture the combined effects of rare genetic variants (Cirulli and Goldstein, 2010; Gonzalez-Recio et al., 2015).

In light of this, we proposed a genome-wide analytical approach that draws its theoretical basis from the genome-based restricted maximum likelihood (GREML) approach (Clarke and Cooper, 2010; Maher, 2008; Manolio et al., 2009; Speed et al., 2012; Yang et al., 2011) which utilises both local and genome-wide relationship matrices to provide regional estimates of the heritability across locally defined regions in the genome (Nagamine et al., 2012). This regional heritability analysis can capture the combined effect of SNPs in a region, and thus small effect variants may be detectable. However, the analysis only captures effects associated with individual SNPs.

Haplotypes may provide a better strategy to capture genomic relationships amongst individuals in the presence of causal rare variants. Although rare variants are not in LD with

66 genotyped variants and thus are difficult to capture in conventional GWAS, these rare
67 variants, may be in LD with some haplotypes and thus can be captured using haplotype
68 methods. Compared with genotyped SNPs, capturing haplotype effects may offer an
69 advantage because haplotypes can be functional units (Vormfelde and Brockmöller, 2007).
70 Therefore, haplotype effects may reflect the combined effects of closely linked cis-acting
71 causal variants (Balding, 2006) and using haplotypes could provide real benefit over SNPs in
72 recovering some of the "missing" heritability and identifying novel trait-associated variants.
73 Therefore, we extended the SNP-based regional heritability analysis further by incorporating
74 haplotypes in addition to SNPs in the calculation of the regional GRMs used in the analysis
75 (Shirali et al., 2018). This approach includes two regional GRMs and divides the genome into
76 windows based on local haplotype blocks delimited by recombination boundaries.

77 This paper further explores the properties of both the SNP-based and the haplotype-
78 based regional heritability mapping (SNP-RHM and Hap-RHM respectively). We hypothesise
79 and show by simulation that the Hap-RHM complements existing SNP-RHM analytical
80 approaches by capturing regional effects in the genome that existing SNP-based methods fail
81 to capture. This leads us to propose a mapping strategy that jointly utilises SNP and haplotype
82 GRMs in a single analysis called SNHap-RHM. We then confirm the utility of this approach by
83 applying it to real data obtained from about 20,000 individuals from the Generation Scotland:
84 Scottish Family Health Study (GS: SFHS) (Smith et al., 2012). We analysed two phenotypes:
85 height and major depressive disorder (MDD). The aim was to uncover novel genetic loci that
86 may affect these traits and improve the estimates of the genetic components of the variation
87 in these traits.

88

Methods

The general statistical setting of a regional GREML analysis

Consider a vector $\boldsymbol{y}$ of phenotype values with length n, the linear mixed-effects model for fitting the effects of genomic region i and background polygenic markers is given as:

$$y = X\beta + W_i u_i + Z u_b + e$$

where $\boldsymbol{y}$ is a vector of phenotypes, $\boldsymbol{X}$ is a design matrix of fixed effects, and $\boldsymbol{\beta}$ is a vector of fixed effects, $\boldsymbol{W_i}$ is a design matrix relating phenotype measures to genetic markers in region i and $\boldsymbol{u_i}$ is a vector of random genetic effects due to region i assumed to be multivariate normal, $MVN\left(0, \sigma^2_{u_i}\boldsymbol{L_{u_i}}\right)$. $\boldsymbol{L_{u_i}}$ is a relationship matrix calculated using markers (SNPs or haplotypes) in region i: calculated in the subsequent sections as $\boldsymbol{G}$ for the SNP and $\boldsymbol{H}$ for the haplotype-based models. $\boldsymbol{Z}$ is a design matrix for background polygenic effects of markers outside the region i and $\boldsymbol{u_b}$ is a vector of random polygenic effect of genetic markers excluded from region i, assumed to be multivariate normal, $MVN\left(0, \sigma^2_{u_b}\boldsymbol{B_{u_b}}\right)$. $\boldsymbol{B_{u_b}}$ is a relationship matrix calculated using the markers outside the region i: calculated in the subsequent section in the same way as $\boldsymbol{G}$. And $\boldsymbol{e}$ is a vector of residual effects assumed to be multivariate normal, $MVN(0, \sigma^2_e\boldsymbol{I})$. $\boldsymbol{I}$ is an identity matrix.

Under the model, the vector of phenotypes $\boldsymbol{y}$ is assumed to be normally distributed, $N(\boldsymbol{X\beta}, \boldsymbol{V})$ where the variance is

$$V = \sigma^2_{u_i}\boldsymbol{L_{u_i}} + \sigma^2_{u_b}\boldsymbol{B_{u_b}} + \sigma^2_e\boldsymbol{I}$$

SNP-RHM: SNP-based regional GREML model

A SNP-based regional GREML analysis was first reported by Nagamine et al. (2012). The regional GREML analysis approach we employ here differs from the analysis done by Nagamine et al. (2012) in the way the regions are defined. That analysis defined local regions by breaking the genome into smaller user-defined windows of p SNPs, which overlapped by q SNPs. Here, however, we define regions based on recombination boundaries in the genome.

The regional GREML model fits two genetic relationship matrices (GRMs): one local GRM for the region and a whole-genome GRM for the remaining SNPs in the genome that are outside the region. The GRMs are genomic relatedness matrices calculated as the weighted proportion of the local or genome-wide autosomal SNPs shared identity by state (IBS) between pairs of individuals. The SNP IBS matrices are calculated as follows, following the second scaling factor proposed by VanRaden (2008)

$$G = \frac{MM'}{m}$$

where m is the total number of r local or b background autosomal SNPs, and $\boldsymbol{M}$ is a matrix of genotype codes for the sampled individuals centred by loci means and normalised by the standard deviation of each locus. $\boldsymbol{M}$ is calculated as follows for individual i at locus j

$$M_{ij} = \frac{(x_{ij} - 2p_j)}{\sqrt{2p_j(1-p_j)}}$$

where x_{ij} is the genotype code at locus j for individual i and takes the values 0, 1 and 2 for AA, Aa and aa genotypes respectively, p_j is the frequency of allele 'a' at locus j**.** The SNP-based relationship for individuals i and k is therefore calculated as follows

$$G_{ik} = \frac{1}{m} \times \sum_{j=1}^{m} \frac{(x_{ij} - 2p_j)(x_{kj} - 2p_j)}{2p_j(1 - p_j)}$$

Hap-RHM: Haplotype-based regional GREML model

The haplotype-based regional GREML model follows theoretically from the SNP-based analysis and utilises haplotypes instead of SNPs as the genetic markers for the regional analysis. The analysis fits two GRMs, a haplotype-based regional GRM and a SNP-based background genome-wide GRM. The haplotype-based GRM is similar to the SNP-based GRM defined in the previous section. For a locally defined region (haplotype block) containing h haplotype variants, the haplotype-based kinship for individuals i and k is calculated as follows

$$H_{ik} = \frac{1}{h} \times \sum_{j=1}^{h} \frac{(d_{ij} - 2p_j)(d_{kj} - 2p_j)}{2p_j(1 - p_j)}$$

where d_{ij} is the diplotype code (coded as the number of copies of haplotype j) for individual i and takes the values 0, 1 and 2 for the $h_t h_t$, $h_t h_j$, $h_j h_j$ diplotypes respectively where haplotype t is any haplotype other than haplotype j, i.e. $t \neq j$, p_j is the haplotype frequency for haplotype j.

Phenotype Simulations

Five phenotypes were simulated using available genotypic information of 20,032 individuals from the Generation Scotland: Scottish Family Health Study (Smith et al., 2012). A total of 593,932 genotyped SNPs were used, and missing genotypes were filled in by imputation. A total of 555,091 SNPs remained after a QC that removed SNPs of MAF < 0.01 and SNPs that were out of Hardy-Weinberg equilibrium at p-value < 0.000001.

The five phenotypes were simulated to have a total variance of 1. This total is composed of 0.6 environmental (residual) variance and genetic variance of 0.4. The genetic

151 variance was partitioned into two components, a polygenic variance of 0.3 and a total QTL
152 variance of 0.1 (20 QTLs, each explaining a variance of 0.005). A common polygenic variance
153 was simulated for all five phenotypes from 20,000 markers randomly selected across the
154 genome. The polygenic variance was simulated to be normally distributed with zero mean
155 and variance of 0.3.

156 For each phenotype, 20 regions (haplotype blocks) were randomly selected, one on
157 each chromosome (except chromosomes 6 and 8 because of the unusually high LD in the MHC
158 regions on chromosome 6 and a large inversion on chromosome 8 (Amador et al., 2015)), to
159 simulate quantitative trait loci (QTL). This gave a total of 20 QTLs for each phenotype. The
160 regions were delimited by natural boundaries: recombination hotspots where the estimated
161 recombination frequency exceeds ten centiMorgans per Megabase (10cM/Mb) with the
162 estimated recombination frequency between boundaries being less than ten centiMorgans
163 per Megabase (10cM/Mb) based on the Genome Reference Consortium Human Build 37
164 (International Human Genome Sequencing Consortium, 2004). This recombination threshold
165 resulted in a total 48,772 regions across the genome. The number and type of marker used
166 to simulate the QTL are what defined the five phenotypes. The five phenotypes are, a 1-SNP
167 QTL within the haplotype block, a multiple-SNP (5 SNPs) QTL within the haplotype block, two
168 types of 1-haplotype QTL within the haplotype block (taking either a common or a rare
169 haplotype as causal) and multiple (5) haplotype QTL within the haplotype block. Details of
170 these phenotypes are described below.

171 For the haplotype QTL phenotypes, a haplotype block is treated as a single genetic
172 locus having multiple alleles. Each haplotype variant within a block is considered as an allele
173 of that locus. Each study individual will carry two alleles, or have a diplotype, for each locus

or haplotype block. The genotype data used to simulate the phenotypes were phased using SHAPEIT2 (Delaneau et al., 2013) to produce the haplotypes for study individuals. The multiple haplotype QTL phenotypes were simulated by randomly sampling two rare haplotypes and three common haplotypes within each haplotype block to give five haplotypes per block. The two types of 1-haplotype QTL phenotypes were simulated by randomly sampling a rare haplotype per haplotype block for one type and for the other type a common haplotype was randomly sampled within each haplotype block.

The individual marker contribution to the polygenic effect and the QTL effects were calculated as follows

$$\sigma_j^2 = 2p_j(1-p_j)g_j^2$$

$$g_j = \sqrt{\frac{\sigma_j^2}{2p_j(1-p_j)}}$$

where σ_j^2 is the contribution of a marker to the QTL or polygenic variance, g_j is the effect of a SNP j or haplotype j randomly sampled to have polygenic or QTL effect, p_j is the frequency of haplotype j or the effect allele of the SNP j. For the single marker QTL phenotypes, each QTL explained a variance of 0.005. For the multiple marker QTL phenotypes, each causal variant explained the same variance, with the effects scaled to account for LD in the region so each QTL locus had a variance of 0.005. For the multiple haplotype QTL effects, the haplotype effects were scaled relative to the inverse of their frequency to give a total variance explained by the region of 0.005.

Common environmental effects were randomly sampled for the five phenotypes from a normal distribution $N(0, \sigma_e^2)$ where σ_e^2 is 0.6. This, together with a genetic variance of 0.4,

gave a total variance of 1 for each phenotype. The final simulated phenotype for an individual i was then calculated as follows

$$y(single\ markers\ per\ QTL\ region)_i = \sum_{j=1}^{20000} x_{ij}g_j + \sum_{j=1}^{20} x_{ij}g_j + e_i,$$

$$y(multiple\ markers\ per\ QTL\ region)_i = \sum_{j=1}^{20000} x_{ij}g_j + \sum_{l=1}^{20}\sum_{j=1}^{5} x_{ij}g_j + e_i,$$

where x_{ij} is the number of copies of the effect allele of SNP j for individual i (for haplotypes, this is defined as d_{ij}; the number of copies of haplotype j for individual i) and g_j is the effect of haplotype j or SNP j. Twenty replicates were analysed for each of the five phenotypes with a different set of QTL markers sampled for each replicate.

Analysis of simulated data

We have shown previously that regional GREML analysis (Regional Heritability Mapping or RHM) using fixed region sizes in the genome is a suitable mapping method for finding local genetic effects (Nagamine et al., 2012). The conventional RHM model fits two genomic relationship matrices (GRMs) in the analyses to map genetic loci that affect trait variation: a local GRM (rGRM) calculated using SNPs located in the region and a genome-wide GRM (gwGRM) calculated from SNPs outside the region. We have since extended this conventional regional GREML analysis to incorporate haplotypes in the calculation of the local GRM and have successfully implemented this in a simulation study (Shirali et al., 2018). This study by Shirali et al. (2018) utilises a regional GREML model that breaks the genome into naturally defined regions by delimiting them by recombination hotspots. Two types of regional GREML models are then fitted in turn to the phenotypes. One model (SNP-RHM) uses

SNPs to estimate local genetic relationships between study individuals, and the other model (Hap-RHM) estimates local genetic relationships amongst individuals using haplotypes.

In this simulation study, the five simulated phenotypes were analysed using the two models, the SNP-based regional GREML model (SNP-RHM for the SNP QTL phenotypes) and the haplotype-based regional GREML model (Hap-RHM for the haplotype QTL phenotypes). To test the analytical models' specificity, we applied Hap-RHM to SNP QTL phenotypes and SNP-RHM to the haplotype QTL phenotypes. We also performed a Hap-RHM analysis in which the units of analysis in the haplotype blocks were restricted to regions of 20 or fewer SNPs per haplotype block. This was because we observed that longer haplotype blocks had many SNPs (and hence many, many haplotypes), and this impacted the estimation of the simulated regional effect. The hybrid Hap-RHM, therefore, investigates whether the regional effect is well captured by the haplotype-based model when shorter haplotypes are used.

We estimated the regional genetic variance and polygenic variance using restricted maximum likelihood (REML). For each simulated phenotype, we analysed 220 regions in total to map the 20 simulated QTLs. This involved analysing the region containing the QTL and ten adjacent regions (five in either direction). In this way, we limit the analysis to the regions in the genome with simulated effects, thereby reducing computation time considerably. Also, by analysing neighbouring regions, we are able to explore the precision of estimates of the location of regional effects. We assessed the significance of a region using the Likelihood Ratio Test (LRT). The genome-wide significance threshold was calculated to be LRT = 23.9 (p-value $< 1.02 \times 10^{-6}$) using a Bonferroni correction for testing 48,772 regions.

Also, we selected one replicate for each simulated phenotype and performed a regional heritability analysis that jointly fitted the SNP and the haplotype GRM in an approach

that we termed SNP and Haplotype Regional Heritability Mapping (SNHap-RHM). An overview of SNHap-RHM is shown in Figure 1.

GS: SFHS Data

Genotyping, quality control and phasing of Generation Scotland: Scottish Family Health Study dataset

The data from the Generation Scotland: Scottish Family Health Study comprised 23,960 participants recruited from Scotland (Smith et al., 2006). The DNA from about 20,032 of the participants had been genotyped using the Illumina HumanOmniExpressExome8v1-2_A chip (~700K genome-wide SNP chip) (Smith et al., 2012).

Quality control excluded SNPs and individuals with a call rate less than 98%, SNPs with minor allele frequency (MAF) less than 1% and SNPs that were out of Hardy-Weinberg equilibrium (p-value < 0.000001). A total of 555,091 autosomal SNPs passed quality control for downstream analysis. Phasing of the GS: SFHS data was done using SHAPEIT2 (Delaneau et al., 2013). Haplotype blocks were defined using recombination hotspots with a recombination rate of 10cM/Mb inferred from the Reference Consortium Human Build 37 (International Human Genome Sequencing Consortium, 2004). Haplotypes variants within blocks were determined using the phased data.

Phenotype definition

MDD status for GS: SFHS participants was assigned following an initial mental health screening questionnaire with the questions: "Have you ever seen anybody for emotional or psychiatric problems?" or "Was there ever a time when you, or someone else, thought you should see someone because of the way you were feeling or acting?" Participants who

answered yes to one or both of the screening questions were further interviewed by the structured clinical interview to diagnose mood disorders (SCID) (First et al., 2002). A total of 18,725 participants (2,603 MDD cases and 16,122 controls) were retained for analysis for MDD. A total of 19,944 participants from the GS: SFHS were analysed for height.

SNHap-RHM of MDD and Height

SNHap-RHM fits jointly, the two types of regional GRMs, SNP-based and haplotype-based, in the analysis of phenotypes (Figure 1). We pre-corrected the phenotypes with the whole-genome GRM before performing SNHap-RHM to speed up the GREML analysis of each block. This pre-correction has previously been shown to speed the regional heritability analysis by Shirali et al. (2018). This step involved 22 separate GREML analyses each fitting a whole-genome GRM that excluded SNPs from one chromosome. The residuals from the pre-correction step were then used in the SNHap-RHM analysis. The models adjusted for sex, age, age^2, and the first 20 principal components calculated from the study participants' genomic relationship matrix (calculated using 555,091 autosomal SNPs).

The significance of a region was tested with a likelihood ratio test (LRT) with two degrees of freedom which compared a model with three variance components fitted (the two regional variances together with the residual variance) against a model with only the residual variance component fitted. The individual regional variance components were subsequently tested with an LRT with one degree of freedom which compared a model with three variance components fitted against a model with two variance components fitted (one regional variance component dropped from the model).

The p-values obtained from the LRTs were used to generate genome-wide association plots for each phenotype (equivalent to GWAS Manhattan plots). The genome-wide

significance threshold was calculated to be LRT = 23.9 (p-value $< 1.02 \times 10^{-6}$) using a Bonferroni correction for testing 48,772 regions. The suggestive significance threshold of a region was set at an LRT = 19.5 (p-value $< 1 \times 10^{-5}$).

Results

Simulation: SNP-RHM, Hap-RHM and SNHap-RHM

SNP-RHM and Hap-RHM: We performed a regional GREML analysis that fits two GRMs (one for the region and one for the rest of the genome) per region across multiple genomic regions delimited by recombination hotspots. We tested two types of regional GREML models, SNP-RHM and Hap-RHM, on 20 replicates of five simulated phenotypes. In SNP-RHM, the regional matrix is derived from SNP genotypes whereas in Hap-RHM the regional matrix is derived from haplotypes. The phenotypes were simulated to be determined by 20 regional QTL effects and genome-wide polygenic effects. The regional QTL effects of the five phenotypes were simulated using SNPs as causal variants for two of them and haplotypes for the remaining three as described in the methods section.

The likelihood ratio test (LRT) was used to test the null hypothesis, H_0: that the genetic variance explained by the region is not significant, against the alternative hypothesis, H_1: that the region accounts for a significant proportion of the phenotypic variance. A large LRT statistic is evidence against the null hypothesis, and therefore means the region explains a significant proportion of the phenotypic variance.

The LRTs averaged over the 20 replicates of the five phenotypes are shown in Figure 2. The figure shows plots of average LRT for the QTL regions and ten adjacent regions (5 to each side). The results show that both models detected the simulated regional effects at the

genome-wide significance level (LRT = 23.9) and can capture true causal loci in traits with different genetic architectures. The LRTs were higher on average for the SNP-based model (SNP-RHM) than the haplotype-based model (Hap-RHM). This could be because for Hap-RHM, the genome-wide GRM which is a SNP-based GRM does not tag any of the background haplotype effects that are outside any one particular region being analysed, and thus the residual variance may be inflated by the other haplotype QTLs which downwardly impact the LRTs.

We provide further investigation of the results from the simulation in the supplementary data. For both analysis models, we have presented detailed results of the relationships between the LRT statistics, region size, variance estimates and allele frequencies (Supplementary Figures 3 – 10). We observed that the longer haplotype blocks had many SNPs (and hence many, many haplotypes), and this impacted the estimation of the simulated regional variance (Supplementary Figure 8). We, therefore, performed a hybrid-Hap-RHM analysis that restricted the natural haplotype block sizes to 20 or fewer SNPs per haplotype block. This hybrid-Hap-RHM was to investigate whether the regional variance is well captured by Hap-RHM when shorter haplotypes are used. The hybrid-Hap-RHM underestimated the regional variance for larger regions but did not offer any discernible improvement in the LRT statistics (Supplementary Figure 9). The relationship between region size and estimated variance was different between the Hap-RHM and hybrid-Hap-RHM, while we observed a similar relationship between LRTs and the region size.

Both SNP-RHM and Hap-RHM fail to capture the simulated regional effects when the simulated phenotype has a genetic architecture that does not match the analysis model, i.e., SNP or haplotype (Figure 3 and Supplementary Figure 1). These figures show the results for

the situation where the SNP QTL phenotypes were analysed with the haplotype-based model (Hap-RHM) and the haplotype QTL phenotypes were analysed with the SNP-based model (SNP-RHM). Both models fail to detect the simulated effects in such situations, therefore, showing that the models complement each other since they capture effects due to different types of genetic variants (i.e., tagged by SNPs or haplotypes).

To confirm that two models are complementary and thus independent of each other, we implemented SNHap-RHM that fits the regional SNP and haplotype GRMs jointly, on a replicate of each of the five simulated phenotypes. The significance of regional effects was tested with an LRT with two degrees of freedom. The results are shown in Figure 4 and confirm that the two models are complementary since even when fitted jointly, we can still capture the simulated effects independently.

SNHap-RHM analysis of Height and MDD

The heritability estimates for height and MDD, calculated using the whole-genome GRM, were 81.4% and 13.8% respectively. There were no overlaps between regions identified as significant (tested with an LRT with one degree of freedom) by the haplotype and SNP-based models for either of the two traits (Supplementary Figure 2). This reaffirms our hypothesis shown by simulation that the Hap-RHM is complementary to SNP-RHM in mapping associated genomic loci.

The regional heritability results for height and MDD are presented as plots of minus-Log10 of the LRT p-values (Figures 5 and 6). The plots for the SNHap-RHM, SNP-RHM and Hap-RHM analyses are shown.

The results for height show that nine regions passed the Bonferroni-corrected genome-wide significance threshold in the analysis using SNP-RHM. No region was genome-wide significant for height when analysed with Hap-RHM. Furthermore, these associations still come up when SNPs and haplotypes in those regions are analysed jointly using SNHap-RHM. There are GWAS reported genes that lie in or are within 400kb of these regions (Supplementary Table 1).

For MDD, no region passed the Bonferroni-corrected genome-wide significance threshold for the analysis done with the SNP-based and haplotype-based regional GREML models (Figure 6). Three regions passed the suggestive significance threshold at p-value $< 1 \times 10^{-5}$ for Hap-RHM analysis of MDD. A further nine regions were significant at p-value $< 5 \times 10^{-5}$ for the haplotype-based analysis, and one region for the SNP-based analysis (Supplementary Table 2). Figure 6 shows that when the two local GRMs are fitted jointly using SNHap-RHM, the genomic regions associated with MDD can still be mapped. The associated regions mapped by the haplotype-based model for MDD contains genes reported by GWAS to be associated with several psychiatry phenotypes (Figure 6 and Supplementary Table 2). The top associated region was within 400kb of the *DCC* gene. This gene is part of the NETRIN1 pathway, which has been reported to be associated with major depressive disorder in two GWAS samples (Generation Scotland: Scottish Family Health Study and Psychiatric Genomics Consortium) (Zeng et al., 2017). Zeng et al. (2017a) used a SNP-RHM guided by pathway analysis (first uncover pathway association and then localise *DCC* within the pathway) to show the *DCC* association with major depressive disorder.

A linear mixed effects model was used to test for association of the SNPs within the suggestive significant region identified by the haplotype-based model on chromosome 3 for

MDD. The model tested for association of SNPs by fitting them individually in the model and fitting a GRM to account for relatedness of individuals. The region on chromosome 3 was chosen because there is a psychiatric phenotype associated gene, *MYRIP* (Luciano et al., 2011), mapped to it, unlike the *DCC* region which has the gene outside the region. The results are shown in Table 1. Five SNPs within this region are nominally significant at p-value < 0.05. Four out of these five SNPs confer about 2% increased risk of the disease each. These four SNPs lie within the *MYRIP* gene sequence. The *MYRIP* gene is expressed in the brain (Ganat et al., 2012). A SNP in this gene is reported to be associated with brain processing speed in the Lothian birth cohort (Luciano et al., 2011). Brain processing speed is an important cognitive function that is compromised in psychiatric illness such as schizophrenia and depression, and old age. Also, a SNP in the *MYRIP* gene region is associated with sleep duration (Gottlieb et al., 2007). Sleep durations outside the normal range (both short sleep and long sleep) is significantly associated with increased risk of depression (Mohan et al., 2017; Roberts and Duong, 2014; Watson et al., 2014; Zhai et al., 2015). The *MYRIP* gene is also reported to have a role in insulin secretion (Waselle et al., 2003) and low insulin levels have been linked to depression (Greenwood et al., 2015; Pearson et al., 2010; Webb et al., 2017).

Comparison with published GWAS SNPs

For both traits, the SNPs in the regions that were significant at p-value $< 5 \times 10^{-5}$ were compared to SNPs reported in the GWAS catalogue (MacArthur et al., 2017) to be significant for the two traits. The GWAS catalogue was accessed on the 15th of January 2021. The results are presented in Table 2. The SNP-based and haplotype-based models identified 1,380 and 45 SNPs respectively for height, and 78 and 495 SNPs respectively for MDD taking all SNPs within haplotype blocks significant at p-value $< 5 \times 10^{-5}$. Out of the 1,380 SNPs

identified for height by the SNP-based model, 57 SNPs spanning 20 haplotype regions were in common with published GWAS results for height.

Discussion

We have proposed and implemented a genome-wide analytical method that analyses genomic regions using a regional GREML model (Nagamine et al., 2012). We have since extended this method to include haplotypes by fitting a regional haplotype-based GRM (Hap-RHM) and also redefined genomic regions in our analysis to be delimited by recombination hotspots generated using HapMap Phase II (Shirali et al., 2018). In this study, we build on our previous regional GREML methods by exploring the properties of the SNP and haplotype-based regional heritability mapping models by simulation and demonstrate that the two variance components fitted are largely independent of each other (Supplementary Figure 2). The novelty in this study shows the two models capture two different kinds of effects in terms of genetic architecture and thus the two variance components can be fitted jointly (by fitting the SNP and haplotype regional matrices together) in a joint marker regional heritability mapping procedure called SNHap-RHM.

We hypothesised that the Hap-RHM would complement the SNP-RHM. We investigated this hypothesis in a simulation study in which we simulated 20 replicates each of two types of SNP QTL phenotypes and three types of haplotype QTL phenotypes. The results show that the two GREML models can capture the effects of causal variants within genomic loci associated with the phenotype analysed. The results also show that the two models are specific about the type of causal effect they can capture, therefore, providing support for the hypothesis that haplotype-based regional GREML models will complement SNP-based regional GREML models. We provide further support for this hypothesis by fitting the two

418 GRMs jointly and showing (using a LRT with two degrees of freedom) that we can still capture
419 the simulated effects and real effects from real data.

420 We applied SNHap-RHM to height and MDD phenotypes from the Generation
421 Scotland: Scottish Family Health Study. Again, we draw comparisons between the effects
422 captured by the SNP-RHM and the Hap-RHM. The SNP-RHM identified more genome-wide
423 (GW) significant regions (p-value $< 1.02 \times 10^{-6}$) for height compared to MDD. Fifty-seven of
424 the SNPs identified for height by the SNP-RHM have been reported by other studies to be
425 associated with height. These SNPs spanned 20 genomic regions in the GS: SFHS cohort.
426 Height is a highly polygenic trait with many common genetic variants accounting for a majority
427 of the additive genetic variation (Yang et al., 2015). These common genetic variants may be
428 in LD with genotyped SNPs on SNP chips (these chips are disproportionately enriched for
429 common SNPs). Therefore, the SNP-based regional GREML model is better suited to capture
430 SNP loci in height compared to MDD. MDD is a very heterogeneous phenotype, and thus every
431 MDD case will have a set of genetic and non-genetic risk factors exclusive to them (Levinson
432 et al., 2014). These unique genetic risk factors will mean that at the population level, a lot of
433 the genetic variants driving the disease will be rare.

434 Three genomic regions were identified for MDD by the haplotype-based regional
435 GREML model with p-value $< 1 \times 10^{-5}$. The Hap-RHM works well for MDD because MDD is
436 believed to be driven by rare genetic variants, and the model can capture rare genetic
437 variants. The haplotype model can capture rare variants because of the LD between rare
438 variants (both typed and untyped) and the flanking variants that aggregate to form the
439 haplotypes within the genomic regions. There were no overlaps between regions identified

by the Hap-RHM and SNP-RHM for each of the two traits, which again support the hypothesis that the two classes of models complement each other in mapping associated loci.

In both traits, the top significant regions we mapped at p-value $< 5 \times 10^{-5}$ had genes mapped to those regions or within 400kb of those regions. For height, these genes have been reported to be associated with height in humans (Gudbjartsson et al., 2008; Kichaev et al., 2019; Lango Allen et al., 2010; Nagy et al., 2017; Tachmazidou et al., 2017; Weedon et al., 2008; Wood et al., 2014). For MDD, these genes have been reported to be associated with major depressive disorder and other psychiatry phenotypes (Arnau-Soler et al., 2019; Howard et al., 2019; Liu et al., 2019; Luciano et al., 2011; Wray et al., 2018; Zeng et al., 2017). In one of such regions for MDD, five SNPs within the region are individually significantly associated with MDD at the nominal level (p-value < 0.05). Four of these SNPs lie within the gene sequence of *MYRIP*, and they each confer 2% disease risk. A conventional GWAS analysis would have missed these nominally associated SNPs because they will not reach genome-wide (GW) significance. However, analysing these SNPs within the region as haplotypes, gave us the power to detect the combined effect of these SNPs in the region at a suggestive-significance level even with our relatively small sample size compared to recent genome-wide association studies of MDD: 322,580 (Howard et al., 2018), and 480,359 (Wray et al., 2018).

The current study's primary strength is the ability of SNHap-RHM to incorporate SNP and haplotype information jointly to map genomic regions that affect complex traits. This gives SNHap-RHM a uniquely useful role to play in the future of complex traits analysis. The plummeting costs of whole-genome resequencing (Caulfield et al., 2013) has shifted research focus in GWA studies towards sequence data analysis (Höglund et al., 2019). Although whole-genome sequence data analysis allows incorporating all the genetic variants that drive the

phenotypic variation, there may still be some variants whose individual effects may be too small to be picked up in a conventional GWA analysis. However, regionally analysing sequence information can help overcome this because multiple small-effect variants in a region can add up to a substantial regional effect that can be captured by a regional SNP GRM or tagged by a haplotype GRM. Moreover, by defining haplotype blocks using recombination hotspots, whole-genome information can be summarised naturally without setting arbitrary blocks. More so, regional heritability analysis of sequence data would be an efficient way to deal with the burden of multiple testing which has long been a problem of conventional GWAS.

One limitation of the current study is the computation burden of the analyses which necessitates the pre-correction of the phenotypes with the whole-genome GRM before performing SNHap-RHM. This step involved 22 separate GREML analyses each fitting a whole-genome GRM that excluded SNPs from one chromosome. Although this was done to speed up the analysis, this step was used as an approximation to account for the background polygenic effects of genetic markers outside each region; this would have been about 48,772 separate GREMLs to account for each region. Also, although this study thoroughly evaluates the robustness of SNP and Haplotype RHM using simulation and demonstrates the utility of SNHap-RHM in real phenotype analysis, seeking replication in other cohorts could have improved our understanding and more importantly demonstrate that the analysis is portable across studies and genotyping platforms.

Conclusions

We have implemented a regional GREML analysis and undertaken analyses of regions in the genome delimited by recombination boundaries and shown by simulation that haplotype-based GRMs can capture genetic variance that may be missed by conventional

SNP-based GRMs. We then applied this method in the analysis of real phenotype data from GS: SFHS. Again, we show that the haplotype-based regional GREML model uncovers associations in regions of the genome that explain genetic variance missed by the SNP-based GREML model. In light of this, we went further to show that regional effects can still be captured when the two regional GRMs (SNP and haplotype-based) are fitted jointly: an analytical procedure we termed as SNHap-RHM. This SNHap-RHM presents an exciting new opportunity to analyse complex traits by allowing the joint mapping of novel genomic regions tagged by either SNPs or haplotypes, potentially leading to the recovery of some of the "missing" heritability.

Acknowledgements

Generation Scotland received core support from the Chief Scientist Office of the Scottish Government Health Directorates [CZD/16/6] and the Scottish Funding Council [HR03006]. Genotyping of the GS: SFHS samples was carried out by the Genetics Core Laboratory at the Wellcome Trust Clinical Research Facility, Edinburgh, Scotland and was funded by the Medical Research Council UK and the Wellcome Trust (Wellcome Trust Strategic Award "STratifying Resilience and Depression Longitudinally" (STRADL) Reference 104036/Z/14/Z). The research was conducted using the UK Biobank Resource under Application Number 19655. We acknowledge the Medical Research Council (MRC) UK for funding (grants MC_UU_00007/10, MC_PC_U127592696 and MC_PC_U127561128). We are grateful to all the families who took part, the general practitioners and the Scottish School of Primary Care for their help in recruiting them, and the whole Generation Scotland team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, healthcare assistants and nurses.

Conflict of interest

The authors declare no conflict of interest.

References

Amador, C., Huffman, J., Trochet, H., Campbell, A., Porteous, D., Wilson, J.F., Hastie, N., Vitart, V., Hayward, C., Navarro, P., Haley, C.S., 2015. Recent genomic heritage in Scotland. BMC Genomics 16, 1–17. https://doi.org/10.1186/s12864-015-1605-2

Arnau-Soler, A., Macdonald-Dunlop, E., Adams, M.J., Clarke, T.-K., MacIntyre, D.J., Milburn, K., Navrady, L., Generation Scotland, Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, Hayward, C., McIntosh, A.M., Thomson, P.A., 2019. Genome-wide by environment interaction studies of depressive symptoms and psychosocial stress in UK Biobank and Generation Scotland. Transl Psychiatry 9, 14. https://doi.org/10.1038/s41398-018-0360-y

Balding, D.J., 2006. A tutorial on statistical methods for population association studies. Nature Reviews Genetics 7, 781–791. https://doi.org/10.1038/nrg1916

Caulfield, T., Evans, J., McGuire, A., McCabe, C., Bubela, T., Cook-Deegan, R., Fishman, J., Hogarth, S., Miller, F.A., Ravitsky, V., Biesecker, B., Borry, P., Cho, M.K., Carroll, J.C., Etchegary, H., Joly, Y., Kato, K., Lee, S.S.-J., Rothenberg, K., Sankar, P., Szego, M.J., Ossorio, P., Pullman, D., Rousseau, F., Ungar, W.J., Wilson, B., 2013. Reflections on the Cost of "Low-Cost" Whole Genome Sequencing: Framing the Health Policy Debate. PLOS Biology 11, e1001699. https://doi.org/10.1371/journal.pbio.1001699

Cirulli, E.T., Goldstein, D.B., 2010. Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nat Rev Genet 11, 415–425. https://doi.org/10.1038/nrg2779

Clarke, A.J., Cooper, D.N., 2010. GWAS: heritability missing in action? Eur J Hum Genet 18, 859–861. https://doi.org/10.1038/ejhg.2010.35

Delaneau, O., Zagury, J.-F., Marchini, J., 2013. Improved whole-chromosome phasing for disease and population genetic studies. Nature Methods 10, 5–6. https://doi.org/10.1038/nmeth.2307

First, M.B., Spitzer, R.L., Gibbon, M., Williams, J.B.W., 2002. Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Non-patient Edition.

Ganat, Y.M., Calder, E.L., Kriks, S., Nelander, J., Tu, E.Y., Jia, F., Battista, D., Harrison, N., Parmar, M., Tomishima, M.J., Rutishauser, U., Studer, L., 2012. Identification of embryonic stem cell-derived midbrain dopaminergic neurons for engraftment. J. Clin. Invest. 122, 2928–2939. https://doi.org/10.1172/JCI58767

Gonzalez-Recio, O., Daetwyler, H.D., MacLeod, I.M., Pryce, J.E., Bowman, P.J., Hayes, B.J., Goddard, M.E., 2015. Rare Variants in Transcript and Potential Regulatory Regions Explain a Small Percentage of the Missing Heritability of Complex Traits in Cattle. PLoS ONE 10, e0143945. https://doi.org/10.1371/journal.pone.0143945

Gottlieb, D.J., O'Connor, G.T., Wilk, J.B., 2007. Genome-wide association of sleep and circadian phenotypes. BMC Med Genet 8, S9. https://doi.org/10.1186/1471-2350-8-S1-S9

Greenwood, E.A., Pasch, L.A., Shinkai, K., Cedars, M.I., Huddleston, H.G., 2015. Putative role for insulin resistance in depression risk in polycystic ovary syndrome. Fertility and Sterility 104, 707-714.e1. https://doi.org/10.1016/j.fertnstert.2015.05.019

Gudbjartsson, D.F., Walters, G.B., Thorleifsson, G., Stefansson, H., Halldorsson, B.V., Zusmanovich, P., Sulem, P., Thorlacius, S., Gylfason, A., Steinberg, S., Helgadottir, A., Ingason, A., Steinthorsdottir, V., Olafsdottir, E.J., Olafsdottir, G.H., Jonsson, T., Borch-Johnsen, K., Hansen, T., Andersen, G., Jorgensen, T., Pedersen, O., Aben, K.K., Witjes, J.A., Swinkels, D.W., den Heijer, M., Franke, B., Verbeek, A.L.M., Becker, D.M., Yanek, L.R., Becker, L.C., Tryggvadottir, L., Rafnar, T., Gulcher, J., Kiemeney, L.A., Kong, A., Thorsteinsdottir, U., Stefansson, K., 2008. Many sequence variants affecting diversity of adult human height. Nat Genet 40, 609–615. https://doi.org/10.1038/ng.122

Höglund, J., Rafati, N., Rask-Andersen, M., Enroth, S., Karlsson, T., Ek, W.E., Johansson, Å., 2019. Improved power and precision with whole genome sequencing data in genome-wide association studies of inflammatory biomarkers. Scientific Reports 9, 16844. https://doi.org/10.1038/s41598-019-53111-7

Howard, D.M., Adams, M.J., Clarke, T.-K., Hafferty, J.D., Gibson, J., Shirali, M., Coleman, J.R.I., Hagenaars, S.P., Ward, J., Wigmore, E.M., Alloza, C., Shen, X., Barbu, M.C., Xu, E.Y., Whalley, H.C., Marioni, R.E., Porteous, D.J., Davies, G., Deary, I.J., Hemani, G., Berger, K., Teismann, H., Rawal, R., Arolt, V., Baune, B.T., Dannlowski, U., Domschke, K., Tian, C., Hinds, D.A., 23andMe Research Team, Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, Trzaskowski, M., Byrne, E.M., Ripke, S., Smith, D.J., Sullivan, P.F., Wray, N.R., Breen, G., Lewis, C.M., McIntosh, A.M., 2019. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. Nat Neurosci 22, 343–352. https://doi.org/10.1038/s41593-018-0326-7

Howard, D.M., Adams, M.J., Shirali, M., Clarke, T.-K., Marioni, R.E., Davies, G., Coleman, J.R.I., Alloza, C., Shen, X., Barbu, M.C., Wigmore, E.M., Gibson, J., Hagenaars, S.P., Lewis, C.M., Ward, J., Smith, D.J., Sullivan, P.F., Haley, C.S., Breen, G., Deary, I.J., McIntosh, A.M., 2018. Genome-wide association study of depression phenotypes in UK Biobank identifies variants in excitatory synaptic pathways. Nature Communications 9, 1470. https://doi.org/10.1038/s41467-018-03819-3

International Human Genome Sequencing Consortium, 2004. Finishing the euchromatic sequence of the human genome. Nature 431, 931–945. https://doi.org/10.1038/nature03001

Kichaev, G., Bhatia, G., Loh, P.-R., Gazal, S., Burch, K., Freund, M.K., Schoech, A., Pasaniuc, B., Price, A.L., 2019. Leveraging Polygenic Functional Enrichment to Improve GWAS Power. Am J Hum Genet 104, 65–75. https://doi.org/10.1016/j.ajhg.2018.11.008

Lango Allen, H., Estrada, K., Lettre, G., Berndt, S.I., Weedon, M.N., Rivadeneira, F., Willer, C.J., Jackson, A.U., Vedantam, S., Raychaudhuri, S., Ferreira, T., Wood, A.R., Weyant, R.J., Segrè, A.V., Speliotes, E.K., Wheeler, E., Soranzo, N., Park, J.-H., Yang, J., Gudbjartsson, D., Heard-Costa, N.L., Randall, J.C., Qi, L., Vernon Smith, A., Mägi, R., Pastinen, T., Liang, L., Heid, I.M., Luan, J., Thorleifsson, G., Winkler, T.W., Goddard, M.E., Sin Lo, K., Palmer, C., Workalemahu, T., Aulchenko, Y.S., Johansson, A., Zillikens, M.C., Feitosa, M.F., Esko, T., Johnson, T., Ketkar, S., Kraft, P., Mangino, M., Prokopenko, I., Absher, D., Albrecht, E., Ernst, F., Glazer, N.L., Hayward, C., Hottenga, J.-J., Jacobs, K.B., Knowles, J.W., Kutalik, Z., Monda, K.L., Polasek, O., Preuss, M., Rayner, N.W., Robertson, N.R., Steinthorsdottir, V., Tyrer, J.P., Voight, B.F., Wiklund,

F., Xu, J., Zhao, J.H., Nyholt, D.R., Pellikka, N., Perola, M., Perry, J.R.B., Surakka, I., Tammesoo, M.-L., Altmaier, E.L., Amin, N., Aspelund, T., Bhangale, T., Boucher, G., Chasman, D.I., Chen, C., Coin, L., Cooper, M.N., Dixon, A.L., Gibson, Q., Grundberg, E., Hao, K., Juhani Junttila, M., Kaplan, L.M., Kettunen, J., König, I.R., Kwan, T., Lawrence, R.W., Levinson, D.F., Lorentzon, M., McKnight, B., Morris, A.P., Müller, M., Suh Ngwa, J., Purcell, S., Rafelt, S., Salem, R.M., Salvi, E., Sanna, S., Shi, J., Sovio, U., Thompson, J.R., Turchin, M.C., Vandenput, L., Verlaan, D.J., Vitart, V., White, C.C., Ziegler, A., Almgren, P., Balmforth, A.J., Campbell, H., Citterio, L., De Grandi, A., Dominiczak, A., Duan, J., Elliott, P., Elosua, R., Eriksson, J.G., Freimer, N.B., Geus, E.J.C., Glorioso, N., Haiqing, S., Hartikainen, A.-L., Havulinna, A.S., Hicks, A.A., Hui, J., Igl, W., Illig, T., Jula, A., Kajantie, E., Kilpeläinen, T.O., Koiranen, M., Kolcic, I., Koskinen, S., Kovacs, P., Laitinen, J., Liu, J., Lokki, M.-L., Marusic, A., Maschio, A., Meitinger, T., Mulas, A., Paré, G., Parker, A.N., Peden, J.F., Petersmann, A., Pichler, I., Pietiläinen, K.H., Pouta, A., Ridderstråle, M., Rotter, J.I., Sambrook, J.G., Sanders, A.R., Schmidt, C.O., Sinisalo, J., Smit, J.H., Stringham, H.M., Bragi Walters, G., Widen, E., Wild, S.H., Willemsen, G., Zagato, L., Zgaga, L., Zitting, P., Alavere, H., Farrall, M., McArdle, W.L., Nelis, M., Peters, M.J., Ripatti, S., van Meurs, J.B.J., Aben, K.K., Ardlie, K.G., Beckmann, J.S., Beilby, J.P., Bergman, R.N., Bergmann, S., Collins, F.S., Cusi, D., den Heijer, M., Eiriksdottir, G., Gejman, P.V., Hall, A.S., Hamsten, A., Huikuri, H.V., Iribarren, C., Kähönen, M., Kaprio, J., Kathiresan, S., Kiemeney, L., Kocher, T., Launer, L.J., Lehtimäki, T., Melander, O., Mosley, T.H., Musk, A.W., Nieminen, M.S., O'Donnell, C.J., Ohlsson, C., Oostra, B., Palmer, L.J., Raitakari, O., Ridker, P.M., Rioux, J.D., Rissanen, A., Rivolta, C., Schunkert, H., Shuldiner, A.R., Siscovick, D.S., Stumvoll, M., Tönjes, A., Tuomilehto, J., van Ommen, G.-J., Viikari, J., Heath, A.C., Martin, N.G., Montgomery, G.W., Province, M.A., Kayser, M., Arnold, A.M., Atwood, L.D., Boerwinkle, E., Chanock, S.J., Deloukas, P., Gieger, C., Grönberg, H., Hall, P., Hattersley, A.T., Hengstenberg, C., Hoffman, W., Lathrop, G.M., Salomaa, V., Schreiber, S., Uda, M., Waterworth, D., Wright, A.F., Assimes, T.L., Barroso, I., Hofman, A., Mohlke, K.L., Boomsma, D.I., Caulfield, M.J., Cupples, L.A., Erdmann, J., Fox, C.S., Gudnason, V., Gyllensten, U., Harris, T.B., Hayes, R.B., Jarvelin, M.-R., Mooser, V., Munroe, P.B., Ouwehand, W.H., Penninx, B.W., Pramstaller, P.P., Quertermous, T., Rudan, I., Samani, N.J., Spector, T.D., Völzke, H., Watkins, H., Wilson, J.F., Groop, L.C., Haritunians, T., Hu, F.B., Kaplan, R.C., Metspalu, A., North, K.E., Schlessinger, D., Wareham, N.J., Hunter, D.J., O'Connell, J.R., Strachan, D.P., Wichmann, H.-E., Borecki, I.B., van Duijn, C.M., Schadt, E.E., Thorsteinsdottir, U., Peltonen, L., Uitterlinden, A.G., Visscher, P.M., Chatterjee, N., Loos, R.J.F., Boehnke, M., McCarthy, M.I., Ingelsson, E., Lindgren, C.M., Abecasis, G.R., Stefansson, K., Frayling, T.M., Hirschhorn, J.N., 2010. Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature 467, 832–838. https://doi.org/10.1038/nature09410

Levinson, D.F., Mostafavi, S., Milaneschi, Y., Rivera, M., Ripke, S., Wray, N.R., Sullivan, P.F., 2014. Genetic studies of major depressive disorder: Why are there no GWAS findings, and what can we do about it? Biol Psychiatry 76, 510–512. https://doi.org/10.1016/j.biopsych.2014.07.029

Liu, M., Jiang, Y., Wedow, R., Li, Y., Brazel, D.M., Chen, F., Datta, G., Davila-Velderrain, J., McGuire, D., Tian, C., Zhan, X., Choquet, H., Docherty, A.R., Faul, J.D., Foerster, J.R., Fritsche, L.G., Gabrielsen, M.E., Gordon, S.D., Haessler, J., Hottenga, J.-J., Huang, H.,

644 Jang, S.-K., Jansen, P.R., Ling, Y., Mägi, R., Matoba, N., McMahon, G., Mulas, A., Orrù,
645 V., Palviainen, T., Pandit, A., Reginsson, G.W., Skogholt, A.H., Smith, J.A., Taylor, A.E.,
646 Turman, C., Willemsen, G., Young, H., Young, K.A., Zajac, G.J.M., Zhao, W., Zhou, W.,
647 Bjornsdottir, G., Boardman, J.D., Boehnke, M., Boomsma, D.I., Chen, C., Cucca, F.,
648 Davies, G.E., Eaton, C.B., Ehringer, M.A., Esko, T., Fiorillo, E., Gillespie, N.A.,
649 Gudbjartsson, D.F., Haller, T., Harris, K.M., Heath, A.C., Hewitt, J.K., Hickie, I.B.,
650 Hokanson, J.E., Hopfer, C.J., Hunter, D.J., Iacono, W.G., Johnson, E.O., Kamatani, Y.,
651 Kardia, S.L.R., Keller, M.C., Kellis, M., Kooperberg, C., Kraft, P., Krauter, K.S., Laakso,
652 M., Lind, P.A., Loukola, A., Lutz, S.M., Madden, P.A.F., Martin, N.G., McGue, M.,
653 McQueen, M.B., Medland, S.E., Metspalu, A., Mohlke, K.L., Nielsen, J.B., Okada, Y.,
654 Peters, U., Polderman, T.J.C., Posthuma, D., Reiner, A.P., Rice, J.P., Rimm, E., Rose,
655 R.J., Runarsdottir, V., Stallings, M.C., Stančáková, A., Stefansson, H., Thai, K.K., Tindle,
656 H.A., Tyrfingsson, T., Wall, T.L., Weir, D.R., Weisner, C., Whitfield, J.B., Winsvold, B.S.,
657 Yin, J., Zuccolo, L., Bierut, L.J., Hveem, K., Lee, J.J., Munafo, M.R., Saccone, N.L.,
658 Willer, C.J., Cornelis, M.C., David, S.P., Hinds, D., Jorgenson, E., Kaprio, J., Stitzel, J.A.,
659 Stefansson, K., Thorgeirsson, T.E., Abecasis, G., Liu, D.J., Vrieze, S., 2019. Association
660 studies of up to 1.2 million individuals yield new insights into the genetic etiology of
661 tobacco and alcohol use. Nat Genet 51, 237–244. https://doi.org/10.1038/s41588-
662 018-0307-5
663 Luciano, M., Hansell, N.K., Lahti, J., Davies, G., Medland, S.E., Räikkönen, K., Tenesa, A.,
664 Widen, E., McGhee, K.A., Palotie, A., Liewald, D., Porteous, D.J., Starr, J.M.,
665 Montgomery, G.W., Martin, N.G., Eriksson, J.G., Wright, M.J., Deary, I.J., 2011. Whole
666 genome association scan for genetic polymorphisms influencing information
667 processing speed. Biol Psychol 86, 193–202.
668 https://doi.org/10.1016/j.biopsycho.2010.11.008
669 MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., Junkins, H., McMahon, A.,
670 Milano, A., Morales, J., Pendlington, Z.M., Welter, D., Burdett, T., Hindorff, L., Flicek,
671 P., Cunningham, F., Parkinson, H., 2017. The new NHGRI-EBI Catalog of published
672 genome-wide association studies (GWAS Catalog). Nucleic Acids Res 45, D896–D901.
673 https://doi.org/10.1093/nar/gkw1133
674 Maher, B., 2008. Personal genomes: The case of the missing heritability. Nature News 456,
675 18–21. https://doi.org/10.1038/456018a
676 Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy,
677 M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., Cho, J.H., Guttmacher, A.E., Kong, A.,
678 Kruglyak, L., Mardis, E., Rotimi, C.N., Slatkin, M., Valle, D., Whittemore, A.S.,
679 Boehnke, M., Clark, A.G., Eichler, E.E., Gibson, G., Haines, J.L., Mackay, T.F.C.,
680 McCarroll, S.A., Visscher, P.M., 2009. Finding the missing heritability of complex
681 diseases. Nature 461, 747–753. https://doi.org/10.1038/nature08494
682 Mohan, J., Xiaofan, G., Yingxian, S., 2017. Association between sleep time and depression: a
683 cross-sectional study from countries in rural Northeastern China. J Int Med Res 45,
684 984–992. https://doi.org/10.1177/0300060517701034
685 Nagamine, Y., Pong-Wong, R., Navarro, P., Vitart, V., Hayward, C., Rudan, I., Campbell, H.,
686 Wilson, J., Wild, S., Hicks, A.A., Pramstaller, P.P., Hastie, N., Wright, A.F., Haley, C.S.,
687 2012. Localising Loci underlying Complex Trait Variation Using Regional Genomic
688 Relationship Mapping. PLoS ONE 7, e46501.
689 https://doi.org/10.1371/journal.pone.0046501

690 Nagy, R., Boutin, T.S., Marten, J., Huffman, J.E., Kerr, S.M., Campbell, A., Evenden, L., Gibson,
691 J., Amador, C., Howard, D.M., Navarro, P., Morris, A., Deary, I.J., Hocking, L.J.,
692 Padmanabhan, S., Smith, B.H., Joshi, P., Wilson, J.F., Hastie, N.D., Wright, A.F.,
693 McIntosh, A.M., Porteous, D.J., Haley, C.S., Vitart, V., Hayward, C., 2017. Exploration
694 of haplotype research consortium imputation for genome-wide association studies in
695 20,032 Generation Scotland participants. Genome Med 9, 23.
696 https://doi.org/10.1186/s13073-017-0414-4
697 Pearson, S., Schmidt, M., Patton, G., Dwyer, T., Blizzard, L., Otahal, P., Venn, A., 2010.
698 Depression and Insulin Resistance. Diabetes Care 33, 1128–1133.
699 https://doi.org/10.2337/dc09-1940
700 Pritchard, J.K., 2001. Are rare variants responsible for susceptibility to complex diseases?
701 Am. J. Hum. Genet. 69, 124–137. https://doi.org/10.1086/321272
702 Roberts, R.E., Duong, H.T., 2014. The Prospective Association between Sleep Deprivation
703 and Depression among Adolescents. Sleep 37, 239–244.
704 https://doi.org/10.5665/sleep.3388
705 Shirali, M., Knott, S.A., Pong-Wong, R., Navarro, P., Haley, C.S., 2018. Haplotype Heritability
706 Mapping Method Uncovers Missing Heritability of Complex Traits. Scientific Reports
707 8, 4982. https://doi.org/10.1038/s41598-018-23307-4
708 Smith, B.H., Campbell, A., Linksted, P., Fitzpatrick, B., Jackson, C., Kerr, S.M., 2012. Cohort
709 profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its
710 participants and their potential for genetic research on health and illness. Int J
711 Epidemiol 42. https://doi.org/10.1093/ije/dys084
712 Smith, B.H., Campbell, H., Blackwood, D., Connell, J., Connor, M., Deary, I.J., 2006.
713 Generation Scotland: the Scottish Family Health Study; a new resource for
714 researching genes and heritability. BMC Med Genet 7.
715 https://doi.org/10.1186/1471-2350-7-74
716 Speed, D., Hemani, G., Johnson, M.R., Balding, D.J., 2012. Improved Heritability Estimation
717 from Genome-wide SNPs. The American Journal of Human Genetics 91, 1011–1021.
718 https://doi.org/10.1016/j.ajhg.2012.10.010
719 Tachmazidou, I., Süveges, D., Min, J.L., Ritchie, G.R.S., Steinberg, J., Walter, K., Iotchkova, V.,
720 Schwartzentruber, J., Huang, J., Memari, Y., McCarthy, S., Crawford, A.A., Bombieri,
721 C., Cocca, M., Farmaki, A.-E., Gaunt, T.R., Jousilahti, P., Kooijman, M.N., Lehne, B.,
722 Malerba, G., Männistö, S., Matchan, A., Medina-Gomez, C., Metrustry, S.J., Nag, A.,
723 Ntalla, I., Paternoster, L., Rayner, N.W., Sala, C., Scott, W.R., Shihab, H.A., Southam,
724 L., St Pourcain, B., Traglia, M., Trajanoska, K., Zaza, G., Zhang, W., Artigas, M.S.,
725 Bansal, N., Benn, M., Chen, Z., Danecek, P., Lin, W.-Y., Locke, A., Luan, J., Manning,
726 A.K., Mulas, A., Sidore, C., Tybjaerg-Hansen, A., Varbo, A., Zoledziewska, M., Finan,
727 C., Hatzikotoulas, K., Hendricks, A.E., Kemp, J.P., Moayyeri, A., Panoutsopoulou, K.,
728 Szpak, M., Wilson, S.G., Boehnke, M., Cucca, F., Di Angelantonio, E., Langenberg, C.,
729 Lindgren, C., McCarthy, M.I., Morris, A.P., Nordestgaard, B.G., Scott, R.A., Tobin,
730 M.D., Wareham, N.J., SpiroMeta Consortium, GoT2D Consortium, Burton, P.,
731 Chambers, J.C., Smith, G.D., Dedoussis, G., Felix, J.F., Franco, O.H., Gambaro, G.,
732 Gasparini, P., Hammond, C.J., Hofman, A., Jaddoe, V.W.V., Kleber, M., Kooner, J.S.,
733 Perola, M., Relton, C., Ring, S.M., Rivadeneira, F., Salomaa, V., Spector, T.D., Stegle,
734 O., Toniolo, D., Uitterlinden, A.G., arcOGEN Consortium, Understanding Society
735 Scientific Group, UK10K Consortium, Barroso, I., Greenwood, C.M.T., Perry, J.R.B.,
736 Walker, B.R., Butterworth, A.S., Xue, Y., Durbin, R., Small, K.S., Soranzo, N., Timpson,

737 N.J., Zeggini, E., 2017. Whole-Genome Sequencing Coupled to Imputation Discovers
738 Genetic Signals for Anthropometric Traits. Am J Hum Genet 100, 865–884.
739 https://doi.org/10.1016/j.ajhg.2017.04.014
740 Vormfelde, S.V., Brockmöller, J., 2007. On the value of haplotype-based genotype-
741 phenotype analysis and on data transformation in pharmacogenetics and -genomics.
742 Nat. Rev. Genet. 8. https://doi.org/10.1038/nrg1916-c1
743 Wainschtein, P., Jain, D.P., Yengo, L., Zheng, Z., TOPMed Anthropometry Working Group, T.-
744 O. for P.M.C., Cupples, L.A., Shadyab, A.H., McKnight, B., Shoemaker, B.M., Mitchell,
745 B.D., Psaty, B.M., Kooperberg, C., Roden, D., Darbar, D., Arnett, D.K., Regan, E.A.,
746 Boerwinkle, E., Rotter, J.I., Allison, M.A., McDonald, M.-L.N., Chung, M.K., Smith,
747 N.L., Ellinor, P.T., Vasan, R.S., Mathias, R.A., Rich, S.S., Heckbert, S.R., Redline, S.,
748 Guo, X., Chen, Y.-D.I., Liu, C.-T., Andrade, M. de, Yanek, L.R., Albert, C.M., Hernandez,
749 R.D., McGarvey, S.T., North, K.E., Lange, L.A., Weir, B.S., Laurie, C.C., Yang, J.,
750 Visscher, P.M., 2019. Recovery of trait heritability from whole genome sequence
751 data. bioRxiv 588020. https://doi.org/10.1101/588020
752 Waselle, L., Coppola, T., Fukuda, M., Iezzi, M., El-Amraoui, A., Petit, C., Regazzi, R., 2003.
753 Involvement of the Rab27 Binding Protein Slac2c/MyRIP in Insulin Exocytosis. Mol
754 Biol Cell 14, 4103–4113. https://doi.org/10.1091/mbc.E03-01-0022
755 Watson, N.F., Harden, K.P., Buchwald, D., Vitiello, M.V., Pack, A.I., Strachan, E., Goldberg, J.,
756 2014. Sleep Duration and Depressive Symptoms: A Gene-Environment Interaction.
757 Sleep 37, 351–358. https://doi.org/10.5665/sleep.3412
758 Webb, M., Davies, M., Ashra, N., Bodicoat, D., Brady, E., Webb, D., Moulton, C., Ismail, K.,
759 Khunti, K., 2017. The association between depressive symptoms and insulin
760 resistance, inflammation and adiposity in men and women. PLOS ONE 12, e0187448.
761 https://doi.org/10.1371/journal.pone.0187448
762 Weedon, M.N., Lango, H., Lindgren, C.M., Wallace, C., Evans, D.M., Mangino, M., Freathy,
763 R.M., Perry, J.R.B., Stevens, S., Hall, A.S., Samani, N.J., Shields, B., Prokopenko, I.,
764 Farrall, M., Dominiczak, A., Johnson, T., Bergmann, S., Beckmann, J.S., Vollenweider,
765 P., Waterworth, D.M., Mooser, V., Palmer, C.N.A., Morris, A.D., Ouwehand, W.H.,
766 Zhao, J.H., Li, S., Loos, R.J.F., Barroso, I., Deloukas, P., Sandhu, M.S., Wheeler, E.,
767 Soranzo, N., Inouye, M., Wareham, N.J., Caulfield, M., Munroe, P.B., Hattersley, A.T.,
768 McCarthy, M.I., Frayling, T.M., 2008. Genome-wide association analysis identifies 20
769 loci that influence adult height. Nat Genet 40, 575–583.
770 https://doi.org/10.1038/ng.121
771 Wood, A.R., Esko, T., Yang, J., Vedantam, S., Pers, T.H., Gustafsson, S., Chu, A.Y., Estrada, K.,
772 Luan, J., Kutalik, Z., Amin, N., Buchkovich, M.L., Croteau-Chonka, D.C., Day, F.R.,
773 Duan, Y., Fall, T., Fehrmann, R., Ferreira, T., Jackson, A.U., Karjalainen, J., Lo, K.S.,
774 Locke, A.E., Mägi, R., Mihailov, E., Porcu, E., Randall, J.C., Scherag, A., Vinkhuyzen,
775 A.A.E., Westra, H.-J., Winkler, T.W., Workalemahu, T., Zhao, J.H., Absher, D.,
776 Albrecht, E., Anderson, D., Baron, J., Beekman, M., Demirkan, A., Ehret, G.B.,
777 Feenstra, B., Feitosa, M.F., Fischer, K., Fraser, R.M., Goel, A., Gong, J., Justice, A.E.,
778 Kanoni, S., Kleber, M.E., Kristiansson, K., Lim, U., Lotay, V., Lui, J.C., Mangino, M.,
779 Leach, I.M., Medina-Gomez, C., Nalls, M.A., Nyholt, D.R., Palmer, C.D., Pasko, D.,
780 Pechlivanis, S., Prokopenko, I., Ried, J.S., Ripke, S., Shungin, D., Stancáková, A.,
781 Strawbridge, R.J., Sung, Y.J., Tanaka, T., Teumer, A., Trompet, S., Laan, S.W. van der,
782 Setten, J. van, Vliet-Ostaptchouk, J.V.V., Wang, Z., Yengo, L., Zhang, W., Afzal, U.,
783 Ärnlöv, J., Arscott, G.M., Bandinelli, S., Barrett, A., Bellis, C., Bennett, A.J., Berne, C.,

784 Blüher, M., Bolton, J.L., Böttcher, Y., Boyd, H.A., Bruinenberg, M., Buckley, B.M.,
785 Buyske, S., Caspersen, I.H., Chines, P.S., Clarke, R., Claudi-Boehm, S., Cooper, M.,
786 Daw, E.W., Jong, P.A.D., Deelen, J., Delgado, G., Denny, J.C., Dhonukshe-Rutten, R.,
787 Dimitriou, M., Doney, A.S.F., Dörr, M., Eklund, N., Eury, E., Folkersen, L., Garcia, M.E.,
788 Geller, F., Giedraitis, V., Go, A.S., Grallert, H., Grammer, T.B., Gräßler, J., Grönberg,
789 H., Groot, L.C.P.G.M. de, Groves, C.J., Haessler, J., Hall, P., Haller, T., Hallmans, G.,
790 Hannemann, A., Hartman, C.A., Hassinen, M., Hayward, C., Heard-Costa, N.L.,
791 Helmer, Q., Hemani, G., Henders, A.K., Hillege, H.L., Hlatky, M.A., Hoffmann, W.,
792 Hoffmann, P., Holmen, O., Houwing-Duistermaat, J.J., Illig, T., Isaacs, A., James, A.L.,
793 Jeff, J., Johansen, B., Johansson, Å., Jolley, J., Juliusdottir, T., Junttila, J., Kho, A.N.,
794 Kinnunen, L., Klopp, N., Kocher, T., Kratzer, W., Lichtner, P., Lind, L., Lindström, J.,
795 Lobbens, S., Lorentzon, M., Lu, Y., Lyssenko, V., Magnusson, P.K.E., Mahajan, A.,
796 Maillard, M., McArdle, W.L., McKenzie, C.A., McLachlan, S., McLaren, P.J., Menni, C.,
797 Merger, S., Milani, L., Moayyeri, A., Monda, K.L., Morken, M.A., Müller, G., Müller-
798 Nurasyid, M., Musk, A.W., Narisu, N., Nauck, M., Nolte, I.M., Nöthen, M.M.,
799 Oozageer, L., Pilz, S., Rayner, N.W., Renstrom, F., Robertson, N.R., Rose, L.M.,
800 Roussel, R., Sanna, S., Scharnagl, H., Scholtens, S., Schumacher, F.R., Schunkert, H.,
801 Scott, R.A., Sehmi, J., Seufferlein, T., Shi, J., Silventoinen, K., Smit, J.H., Smith, A.V.,
802 Smolonska, J., Stanton, A.V., Stirrups, K., Stott, D.J., Stringham, H.M., Sundström, J.,
803 Swertz, M.A., Syvänen, A.-C., Tayo, B.O., Thorleifsson, G., Tyrer, J.P., Dijk, S. van,
804 Schoor, N.M. van, Velde, N. van der, Heemst, D. van, Oort, F.V.A. van, Vermeulen,
805 S.H., Verweij, N., Vonk, J.M., Waite, L.L., Waldenberger, M., Wennauer, R., Wilkens,
806 L.R., Willenborg, C., Wilsgaard, T., Wojczynski, M.K., Wong, A., Wright, A.F., Zhang,
807 Q., Arveiler, D., Bakker, S.J.L., Beilby, J., Bergman, R.N., Bergmann, S., Biffar, R.,
808 Blangero, J., Boomsma, D.I., Bornstein, S.R., Bovet, P., Brambilla, P., Brown, M.J.,
809 Campbell, H., Caulfield, M.J., Chakravarti, A., Collins, R., Collins, F.S., Crawford, D.C.,
810 Cupples, L.A., Danesh, J., Faire, U. de, Ruijter, H.M. den, Erbel, R., Erdmann, J.,
811 Eriksson, J.G., Farrall, M., Ferrannini, E., Ferrières, J., Ford, I., Forouhi, N.G., Forrester,
812 T., Gansevoort, R.T., Gejman, P.V., Gieger, C., Golay, A., Gottesman, O., Gudnason,
813 V., Gyllensten, U., Haas, D.W., Hall, A.S., Harris, T.B., Hattersley, A.T., Heath, A.C.,
814 Hengstenberg, C., Hicks, A.A., Hindorff, L.A., Hingorani, A.D., Hofman, A., Hovingh,
815 G.K., Humphries, S.E., Hunt, S.C., Hypponen, E., Jacobs, K.B., Jarvelin, M.-R.,
816 Jousilahti, P., Jula, A.M., Kaprio, J., Kastelein, J.J.P., Kayser, M., Kee, F., Keinanen-
817 Kiukaanniemi, S.M., Kiemeney, L.A., Kooner, J.S., Kooperberg, C., Koskinen, S.,
818 Kovacs, P., Kraja, A.T., Kumari, M., Kuusisto, J., Lakka, T.A., Langenberg, C.,
819 Marchand, L.L., Lehtimäki, T., Lupoli, S., Madden, P.A.F., Männistö, S., Manunta, P.,
820 Marette, A., Matise, T.C., McKnight, B., Meitinger, T., Moll, F.L., Montgomery, G.W.,
821 Morris, A.D., Morris, A.P., Murray, J.C., Nelis, M., Ohlsson, C., Oldehinkel, A.J., Ong,
822 K.K., Ouwehand, W.H., Pasterkamp, G., Peters, A., Pramstaller, P.P., Price, J.F., Qi, L.,
823 Raitakari, O.T., Rankinen, T., Rao, D.C., Rice, T.K., Ritchie, M., Rudan, I., Salomaa, V.,
824 Samani, N.J., Saramies, J., Sarzynski, M.A., Schwarz, P.E.H., Sebert, S., Sever, P.,
825 Shuldiner, A.R., Sinisalo, J., Steinthorsdottir, V., Stolk, R.P., Tardif, J.-C., Tönjes, A.,
826 Tremblay, A., Tremoli, E., Virtamo, J., Vohl, M.-C., Consortium, T.E.M.R. and G.
827 (eMERGE), Consortium, T.Mig., Consortium, T.P., Study, T.L.C., Amouyel, P.,
828 Asselbergs, F.W., Assimes, T.L., Bochud, M., Boehm, B.O., Boerwinkle, E., Bottinger,
829 E.P., Bouchard, C., Cauchi, S., Chambers, J.C., Chanock, S.J., Cooper, R.S., Bakker,
830 P.I.W. de, Dedoussis, G., Ferrucci, L., Franks, P.W., Froguel, P., Groop, L.C., Haiman,

C.A., Hamsten, A., Hayes, M.G., Hui, J., Hunter, D.J., Hveem, K., Jukema, J.W., Kaplan, R.C., Kivimaki, M., Kuh, D., Laakso, M., Liu, Y., Martin, N.G., März, W., Melbye, M., Moebus, S., Munroe, P.B., Njølstad, I., Oostra, B.A., Palmer, C.N.A., Pedersen, N.L., Perola, M., Pérusse, L., Peters, U., Powell, J.E., Power, C., Quertermous, T., Rauramaa, R., Reinmaa, E., Ridker, P.M., Rivadeneira, F., Rotter, J.I., Saaristo, T.E., Saleheen, D., Schlessinger, D., Slagboom, P.E., Snieder, H., Spector, T.D., Strauch, K., Stumvoll, M., Tuomilehto, J., Uusitupa, M., Harst, P. van der, Völzke, H., Walker, M., Wareham, N.J., Watkins, H., Wichmann, H.-E., Wilson, J.F., Zanen, P., Deloukas, P., Heid, I.M., Lindgren, C.M., Mohlke, K.L., Speliotes, E.K., Thorsteinsdottir, U., Barroso, I., Fox, C.S., North, K.E., Strachan, D.P., Beckmann, J.S., Berndt, S.I., Boehnke, M., Borecki, I.B., McCarthy, M.I., Metspalu, A., Stefansson, K., Uitterlinden, A.G., Duijn, C.M. van, Franke, L., Willer, C.J., Price, A.L., Lettre, G., Loos, R.J.F., Weedon, M.N., Ingelsson, E., O'Connell, J.R., Abecasis, G.R., Chasman, D.I., Goddard, M.E., Visscher, P.M., Hirschhorn, J.N., Frayling, T.M., 2014. Defining the role of common variation in the genomic and biological architecture of adult human height. Nature Genetics 46, 1173–1186. https://doi.org/10.1038/ng.3097

Wray, N.R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E.M., Abdellaoui, A., Adams, M.J., Agerbo, E., Air, T.M., Andlauer, T.M.F., Bacanu, S.-A., Bækvad-Hansen, M., Beekman, A.F.T., Bigdeli, T.B., Binder, E.B., Blackwood, D.R.H., Bryois, J., Buttenschøn, H.N., Bybjerg-Grauholm, J., Cai, N., Castelao, E., Christensen, J.H., Clarke, T.-K., Coleman, J.I.R., Colodro-Conde, L., Couvy-Duchesne, B., Craddock, N., Crawford, G.E., Crowley, C.A., Dashti, H.S., Davies, G., Deary, I.J., Degenhardt, F., Derks, E.M., Direk, N., Dolan, C.V., Dunn, E.C., Eley, T.C., Eriksson, N., Escott-Price, V., Kiadeh, F.H.F., Finucane, H.K., Forstner, A.J., Frank, J., Gaspar, H.A., Gill, M., Giusti-Rodríguez, P., Goes, F.S., Gordon, S.D., Grove, J., Hall, L.S., Hannon, E., Hansen, C.S., Hansen, T.F., Herms, S., Hickie, I.B., Hoffmann, P., Homuth, G., Horn, C., Hottenga, J.-J., Hougaard, D.M., Hu, M., Hyde, C.L., Ising, M., Jansen, R., Jin, F., Jorgenson, E., Knowles, J.A., Kohane, I.S., Kraft, J., Kretzschmar, W.W., Krogh, J., Kutalik, Z., Lane, J.M., Li, Yihan, Li, Yun, Lind, P.A., Liu, X., Lu, L., MacIntyre, D.J., MacKinnon, D.F., Maier, R.M., Maier, W., Marchini, J., Mbarek, H., McGrath, P., McGuffin, P., Medland, S.E., Mehta, D., Middeldorp, C.M., Mihailov, E., Milaneschi, Y., Milani, L., Mill, J., Mondimore, F.M., Montgomery, G.W., Mostafavi, S., Mullins, N., Nauck, M., Ng, B., Nivard, M.G., Nyholt, D.R., O'Reilly, P.F., Oskarsson, H., Owen, M.J., Painter, J.N., Pedersen, C.B., Pedersen, M.G., Peterson, R.E., Pettersson, E., Peyrot, W.J., Pistis, G., Posthuma, D., Purcell, S.M., Quiroz, J.A., Qvist, P., Rice, J.P., Riley, B.P., Rivera, M., Mirza, S.S., Saxena, R., Schoevers, R., Schulte, E.C., Shen, L., Shi, J., Shyn, S.I., Sigurdsson, E., Sinnamon, G.B.C., Smit, J.H., Smith, D.J., Stefansson, H., Steinberg, S., Stockmeier, C.A., Streit, F., Strohmaier, J., Tansey, K.E., Teismann, H., Teumer, A., Thompson, W., Thomson, P.A., Thorgeirsson, T.E., Tian, C., Traylor, M., Treutlein, J., Trubetskoy, V., Uitterlinden, A.G., Umbricht, D., Auwera, S.V. der, Hemert, A.M. van, Viktorin, A., Visscher, P.M., Wang, Y., Webb, B.T., Weinsheimer, S.M., Wellmann, J., Willemsen, G., Witt, S.H., Wu, Y., Xi, H.S., Yang, J., Zhang, F., Arolt, V., Baune, B.T., Berger, K., Boomsma, D.I., Cichon, S., Dannlowski, U., Geus, E.C.J. de, DePaulo, J.R., Domenici, E., Domschke, K., Esko, T., Grabe, H.J., Hamilton, S.P., Hayward, C., Heath, A.C., Hinds, D.A., Kendler, K.S., Kloiber, S., Lewis, G., Li, Q.S., Lucae, S., Madden, P.F.A., Magnusson, P.K., Martin, N.G., McIntosh, A.M., Metspalu, A., Mors, O., Mortensen, P.B., Müller-Myhsok, B., Nordentoft, M., Nöthen, M.M.,

878 O'Donovan, M.C., Paciga, S.A., Pedersen, N.L., Penninx, B.W.J.H., Perlis, R.H.,
879 Porteous, D.J., Potash, J.B., Preisig, M., Rietschel, M., Schaefer, C., Schulze, T.G.,
880 Smoller, J.W., Stefansson, K., Tiemeier, H., Uher, R., Völzke, H., Weissman, M.M.,
881 Werge, T., Winslow, A.R., Lewis, C.M., Levinson, D.F., Breen, G., Børglum, A.D.,
882 Sullivan, P.F., 2018. Genome-wide association analyses identify 44 risk variants and
883 refine the genetic architecture of major depression. Nature Genetics 50, 668–681.
884 https://doi.org/10.1038/s41588-018-0090-3
885 Yang, J., Bakshi, A., Zhu, Z., Hemani, G., Vinkhuyzen, A.A.E., Lee, S.H., Robinson, M.R., Perry,
886 J.R.B., Nolte, I.M., van Vliet-Ostaptchouk, J.V., Snieder, H., Esko, T., Milani, L., Mägi,
887 R., Metspalu, A., Hamsten, A., Magnusson, P.K.E., Pedersen, N.L., Ingelsson, E.,
888 Soranzo, N., Keller, M.C., Wray, N.R., Goddard, M.E., Visscher, P.M., 2015. Genetic
889 variance estimation with imputed variants finds negligible missing heritability for
890 human height and body mass index. Nature Genetics 47, 1114–1120.
891 https://doi.org/10.1038/ng.3390
892 Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A.,
893 Heath, A.C., Martin, N.G., Montgomery, G.W., Goddard, M.E., Visscher, P.M., 2010.
894 Common SNPs explain a large proportion of the heritability for human height. Nat
895 Genet 42, 565–569. https://doi.org/10.1038/ng.608
896 Yang, J., Lee, S.H., Goddard, M.E., Visscher, P.M., 2011. GCTA: A Tool for Genome-wide
897 Complex Trait Analysis. The American Journal of Human Genetics 88, 76–82.
898 https://doi.org/10.1016/j.ajhg.2010.11.011
899 Zeng, Y., Navarro, P., Fernandez-Pujals, A.M., Hall, L.S., Clarke, T.-K., Thomson, P.A., Smith,
900 B.H., Hocking, L.J., Padmanabhan, S., Hayward, C., MacIntyre, D.J., Wray, N.R., Deary,
901 I.J., Porteous, D.J., Haley, C.S., McIntosh, A.M., 2017. A Combined Pathway and
902 Regional Heritability Analysis Indicates NETRIN1 Pathway Is Associated With Major
903 Depressive Disorder. Biol Psychiatry 81, 336–346.
904 https://doi.org/10.1016/j.biopsych.2016.04.017
905 Zhai, L., Zhang, H., Zhang, D., 2015. SLEEP DURATION AND DEPRESSION AMONG ADULTS: A
906 META-ANALYSIS OF PROSPECTIVE STUDIES. Depress Anxiety 32, 664–670.
907 https://doi.org/10.1002/da.22386
908

909

910

911

912

913

914

Table 1. SNP-based association test of MDD in the *MYRIP* gene region. The columns are the SNP ID, chromosome, genome position of SNP, minor allele frequency, odds ratio, log of odds ratio, standard error of log odds ratio and association p-value.

SNP information				Depression association			
SNP ID	Chr	Pos	MAF	OR	Log (OR)	SE (logOR)	p
rs9842160	3	39844703	0.14	0.97	-0.030	0.013	0.02
rs9858242	3	39847606	0.19	1.02	0.025	0.011	0.03
rs1599902	3	39954674	0.41	1.02	0.019	0.009	0.04
rs7618607	3	39947936	0.41	1.02	0.019	0.009	0.04
rs9860916	3	39944942	0.41	1.02	0.019	0.009	0.04

Table 2. Comparison of SNPs within significant regions identified by both models and published GWAS results for height and MDD. The columns are the name of trait, number of SNPs in regions identified by SNP-RHM and Hap-RHM with p-value $< 5 \times 10^{-5}$ and SNPs in published GWAS (pubGWAS) for the traits, and the number of SNPs overlapping between the three.

	Number of SNPS			Number of overlapping SNPS		
Trait	SNP-RHM	Hap-RHM	pubGWAS	SNP-RHM & Hap-RHM	SNP-RHM & pubGWAS	Hap-RHM & pubGWAS
Height	1380	45	4960	0	57	0
MDD	78	495	1815	0	0	0

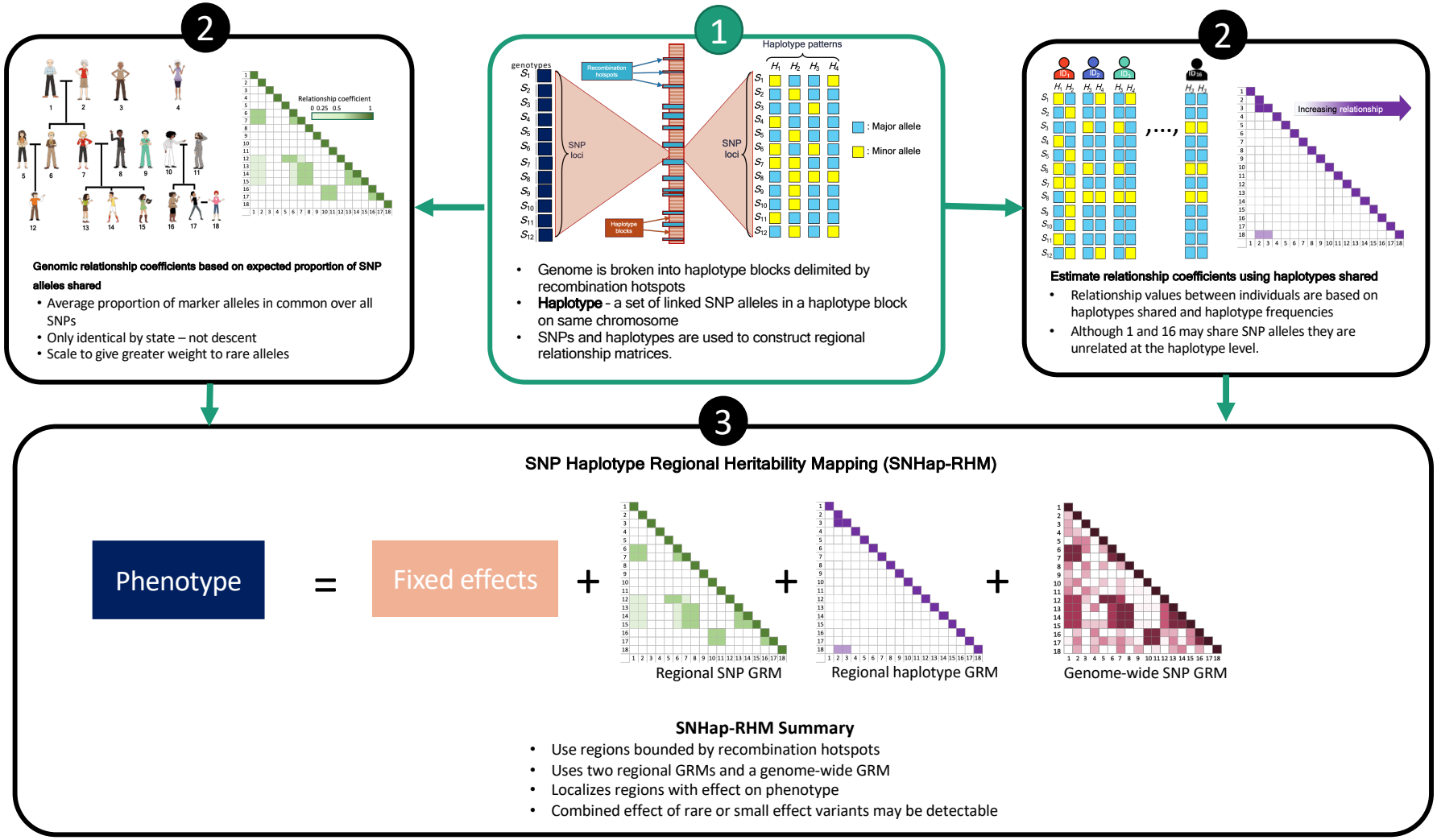

Figure 1. A Schema outlying SNHap-RHM

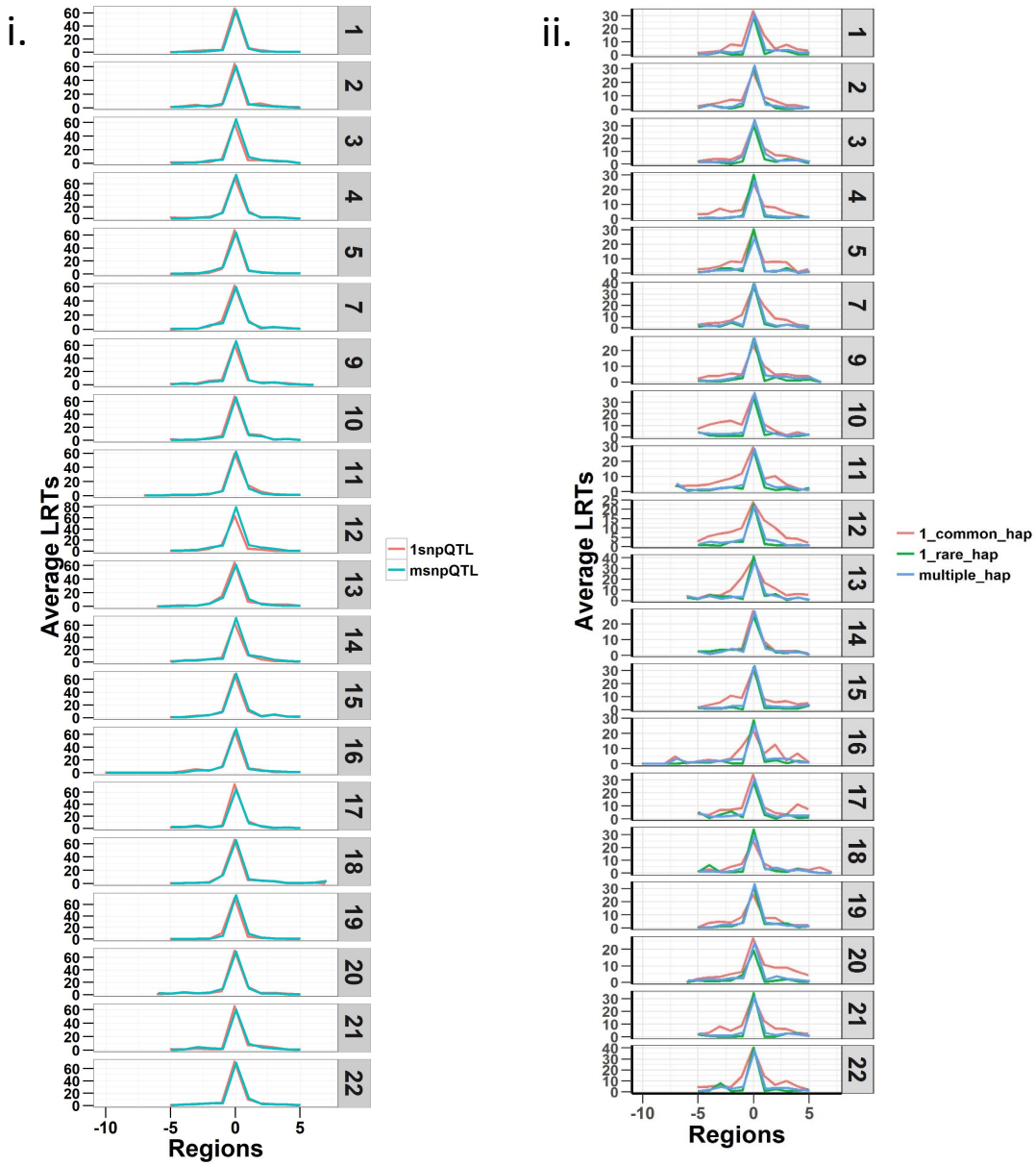

Figure 2. Plots of Likelihood ratio test (LRT) statistics at each QTL loci and 5 regions either side averaged for the 20 simulations of each of the five QTL phenotypes. Plot (i) is SNP QTL phenotypes analysed using the SNP-RHM and plot (ii) is the haplotype QTL phenotypes analysed using the Hap-RHM. Both models can capture the simulated QTL effects for their respective SNP and haplotype phenotypes.

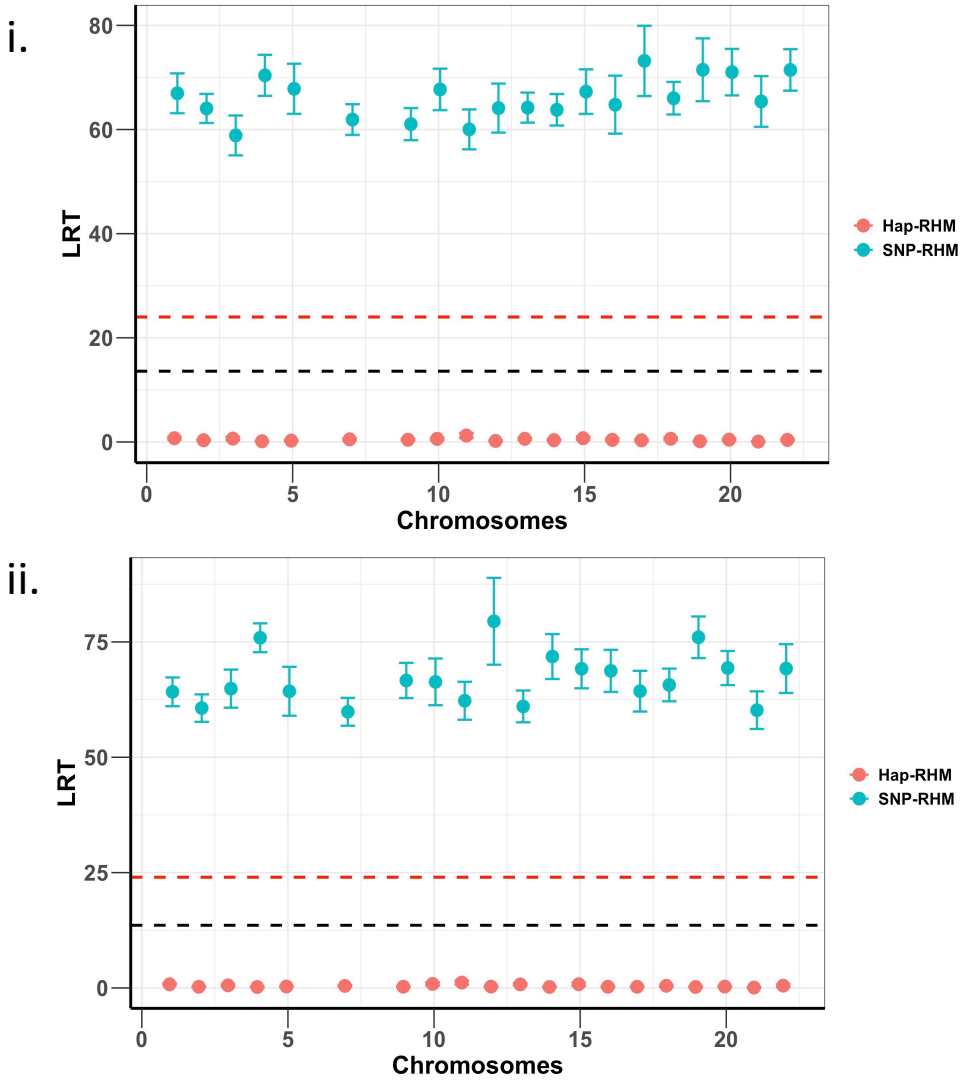

Figure 3. Plots of average LRT statistics over replicates of QTL loci across the chromosomes for the 20 simulations of each of the two SNP QTL phenotypes. The red dashed lines are genome-wide significance threshold (for 48,772 regions) and the black dashed lines are Bonferroni significance threshold (for 220 regions). The upper plot (i) is the 1-SNP QTL phenotype, and the lower plot (ii) is the multiple SNP QTL phenotype. The two phenotypes are analysed using both the SNP based model (SNP-RHM) (blue points) and the Haplotype based model (Hap-RHM) (red points). The Hap-RHM fails to capture the simulated effects for the SNP QTLs.

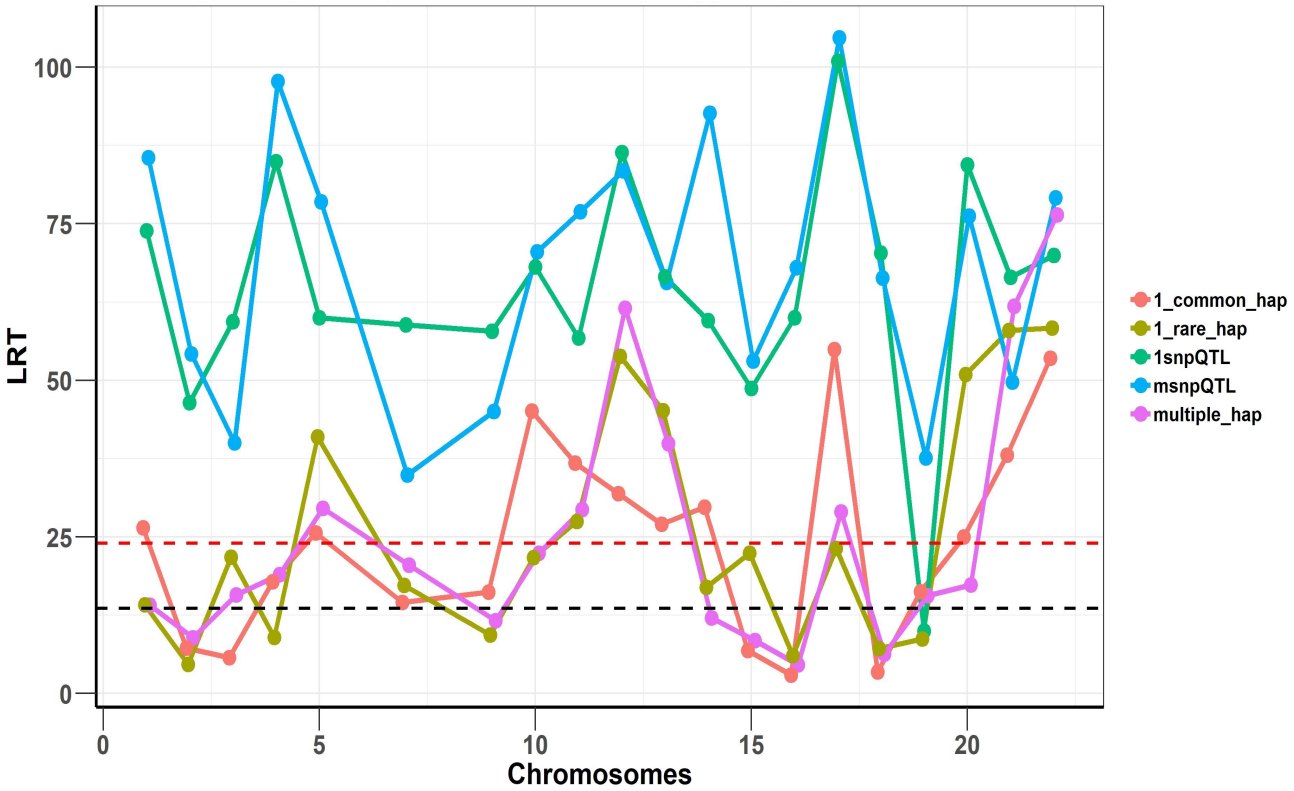

Figure 4. Joint analysis of the SNP and haplotype phenotypes using SNHap-RHM. The plot is an analysis of one replicate of each of the simulated phenotypes. The LRT statistics are plotted over QTL loci across the chromosomes. The red dashed lines are genome-wide significance threshold (for 48,772 regions) and the black dashed lines are Bonferroni significance threshold (for 220 regions).

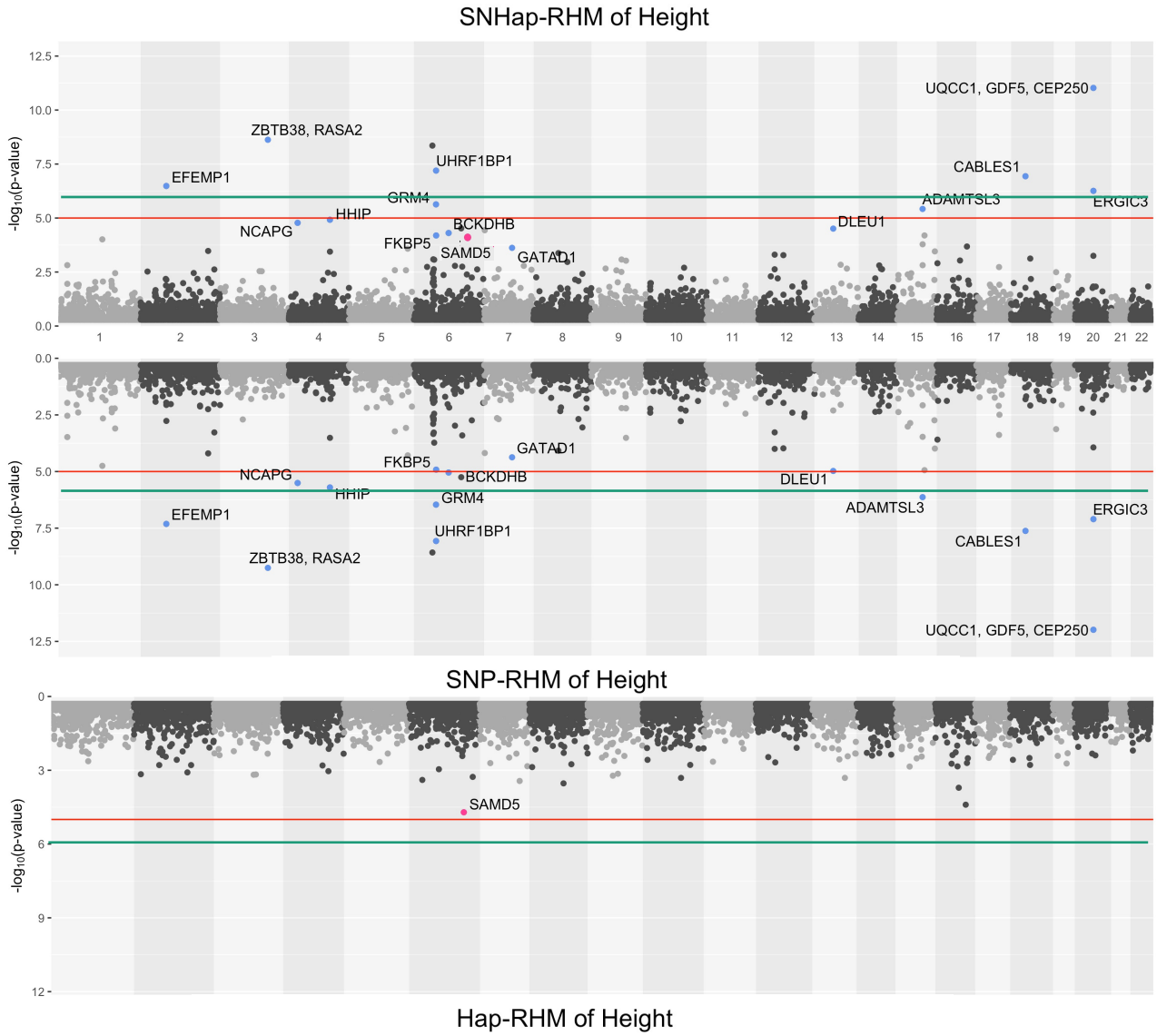

Figure 5. The genome-wide evidence of haplotype block association for height. Analysis done with SNHap-RHM, SNP-RHM and Hap-RHM. The points are plots of -log10 of the p-values of regions tested with the LRT for the regional GREML analyses. The green lines are the Bonferroni-corrected genome-wide significance threshold and the red lines are the suggestive significance threshold calculated to be p-value $< 1 \times 10^{-5}$. The top association hits at p-value $< 5 \times 10^{-5}$ with genes located within the region are highlighted in blue for SNP-RHM and red for the Hap-RHM.

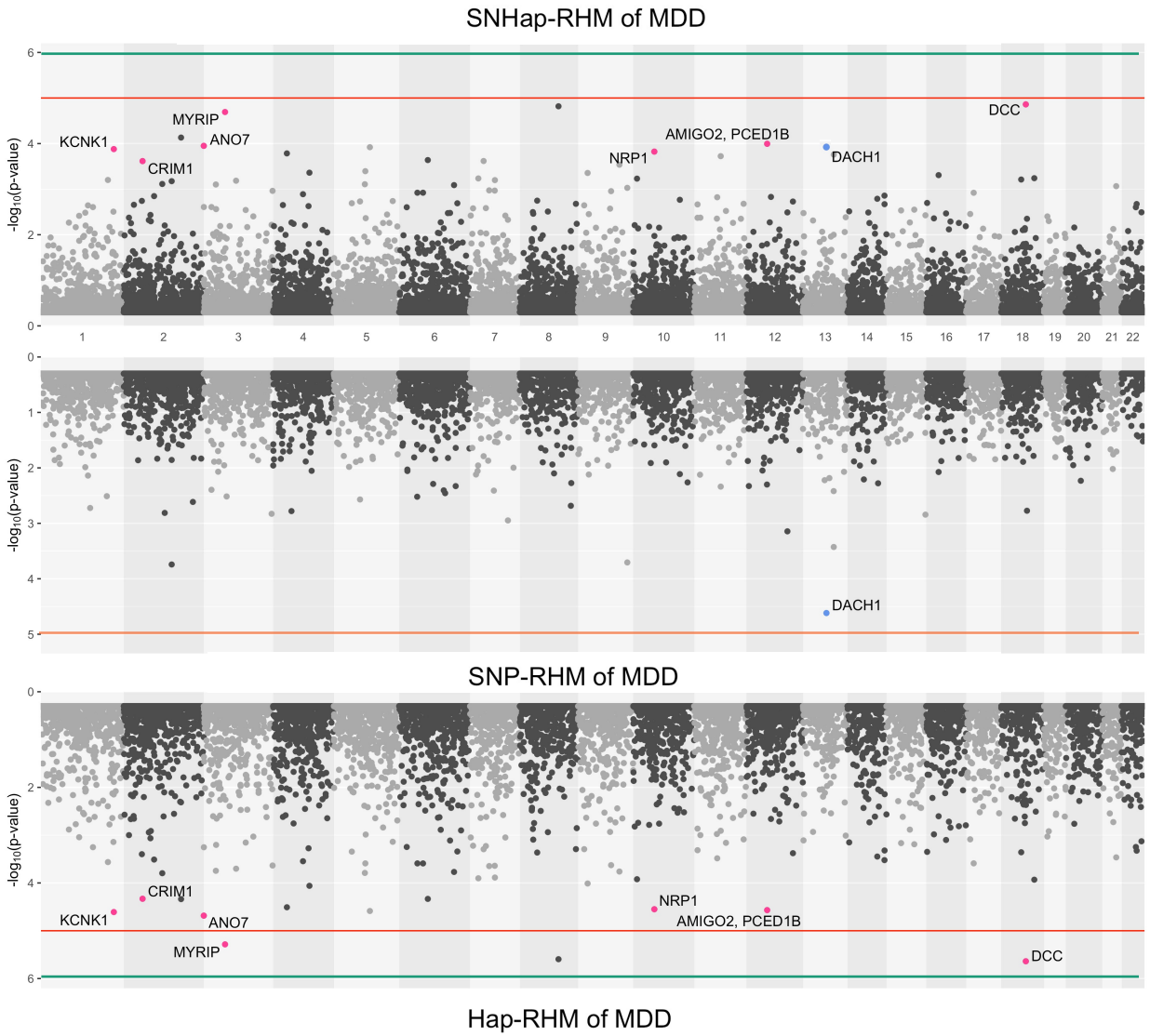

Figure 6. The genome-wide evidence of haplotype block association for Major Depressive Disorder. Analysis done with SNHap-RHM, SNP-RHM and Hap-RHM. The points are plots of -log10 of the p-values of regions tested with the LRT for the regional GREML analyses. The green lines are the Bonferroni-corrected genome-wide significance threshold and the red lines are the suggestive significance threshold calculated to be p-value $< 1 \times 10^{-5}$. The top association hits at p-value $< 5 \times 10^{-5}$ with genes located within the region are highlighted in blue for SNP-RHM and red for the Hap-RHM.