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

43 **Introduction**

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

56 In light of this, we proposed a genome-wide analytical approach that draws its
57 theoretical basis from the genome-based restricted maximum likelihood (GREML) approach
58 (Clarke and Cooper, 2010; Maher, 2008; Manolio et al., 2009; Speed et al., 2012; Yang et al.,
59 2011) which utilises both local and genome-wide relationship matrices to provide regional
60 estimates of the heritability across locally defined regions in the genome (Nagamine et al.,
61 2012). This regional heritability analysis can capture the combined effect of SNPs in a region,
62 and thus small effect variants may be detectable. However, the analysis only captures effects
63 associated with individual SNPs.

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

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

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

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

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

110 **SNP-RHM: SNP-based regional GREML model**

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

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

122 $$G = \frac{MM'}{m}$$

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

126 $$M_{ij} = \frac{(x_{ij} - 2p_j)}{\sqrt{2p_j(1 - p_j)}}$$

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

130 $$G_{ik} = \frac{1}{m} \times \sum_{j=1}^{m} \frac{(x_{ij} - 2p_j)(x_{kj} - 2p_j)}{2p_j(1 - p_j)}$$

131 **Hap-RHM: Haplotype-based regional GREML model**

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

138 $$H_{ik} = \frac{1}{h} \times \sum_{j=1}^{h} \frac{(d_{ij} - 2p_j)(d_{kj} - 2p_j)}{2p_j(1 - p_j)}$$

139 where d_{ij} is the diplotype code (coded as the number of copies of haplotype j) for individual
140 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
141 haplotype t is any haplotype other than haplotype j, i.e. $t \neq j$, p_j is the haplotype frequency
142 for haplotype j.

143 **Phenotype Simulations**

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

149 The five phenotypes were simulated to have a total variance of 1. This total is
150 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

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

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

183 $$\sigma_j^2 = 2p_j(1 - p_j)g_j^2$$

184 $$g_j = \sqrt{\frac{\sigma_j^2}{2p_j(1 - p_j)}}$$

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

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

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

197 $$y(single\ markers\ per\ QTL\ region)_i = \sum_{j=1}^{20000} x_{ij}g_j + \sum_{j=1}^{20} x_{ij}g_j + e_i,$$

198 $$y(multiple\ markers\ per\ QTL\ region)_i = \sum_{j=1}^{20000} x_{ij}g_j + \sum_{l=1}^{20}\sum_{j=1}^{5} x_{ij}g_j + e_i,$$

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

203 **Analysis of simulated data**

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

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

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

227 We estimated the regional genetic variance and polygenic variance using restricted
228 maximum likelihood (REML). For each simulated phenotype, we analysed 220 regions in total
229 to map the 20 simulated QTLs. This involved analysing the region containing the QTL and ten
230 adjacent regions (five in either direction). In this way, we limit the analysis to the regions in
231 the genome with simulated effects, thereby reducing computation time considerably. Also,
232 by analysing neighbouring regions, we are able to explore the precision of estimates of the
233 location of regional effects. We assessed the significance of a region using the Likelihood Ratio
234 Test (LRT). The genome-wide significance threshold was calculated to be LRT = 23.9 (p-value
235 $< 1.02 \times 10^{-6}$) using a Bonferroni correction for testing 48,772 regions.

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

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

240 **GS: SFHS Data**

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

243 The data from the Generation Scotland: Scottish Family Health Study comprised
244 23,960 participants recruited from Scotland (Smith et al., 2006). The DNA from about 20,032
245 of the participants had been genotyped using the Illumina HumanOmniExpressExome8v1-2_A
246 chip (~700K genome-wide SNP chip) (Smith et al., 2012).

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

255 ***Phenotype definition***

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

260 answered yes to one or both of the screening questions were further interviewed by the
261 structured clinical interview to diagnose mood disorders (SCID) (First et al., 2002). A total of
262 18,725 participants (2,603 MDD cases and 16,122 controls) were retained for analysis for
263 MDD. A total of 19,944 participants from the GS: SFHS were analysed for height.

264 ***SNHap-RHM of MDD and Height***

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

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

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

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

286 **Results**

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

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

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

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

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

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

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

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

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

339 **SNHap-RHM analysis of Height and MDD**

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

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

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

355 For MDD, no region passed the Bonferroni-corrected genome-wide significance
356 threshold for the analysis done with the SNP-based and haplotype-based regional GREML
357 models (Figure 6). Three regions passed the suggestive significance threshold at p-value $<$
358 1×10^{-5} for Hap-RHM analysis of MDD. A further nine regions were significant at p-value $<$
359 5×10^{-5} for the haplotype-based analysis, and one region for the SNP-based analysis
360 (Supplementary Table 2). Figure 6 shows that when the two local GRMs are fitted jointly using
361 SNHap-RHM, the genomic regions associated with MDD can still be mapped. The associated
362 regions mapped by the haplotype-based model for MDD contains genes reported by GWAS
363 to be associated with several psychiatry phenotypes (Figure 6 and Supplementary Table 2).
364 The top associated region was within 400kb of the *DCC* gene. This gene is part of the NETRIN1
365 pathway, which has been reported to be associated with major depressive disorder in two
366 GWAS samples (Generation Scotland: Scottish Family Health Study and Psychiatric Genomics
367 Consortium) (Zeng et al., 2017). Zeng et al. (2017a) used a SNP-RHM guided by pathway
368 analysis (first uncover pathway association and then localise *DCC* within the pathway) to show
369 the *DCC* association with major depressive disorder.

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

372 MDD. The model tested for association of SNPs by fitting them individually in the model and
373 fitting a GRM to account for relatedness of individuals. The region on chromosome 3 was
374 chosen because there is a psychiatric phenotype associated gene, *MYRIP* (Luciano et al.,
375 2011), mapped to it, unlike the *DCC* region which has the gene outside the region. The results
376 are shown in Table 1. Five SNPs within this region are nominally significant at p-value < 0.05.
377 Four out of these five SNPs confer about 2% increased risk of the disease each. These four
378 SNPs lie within the *MYRIP* gene sequence. The *MYRIP* gene is expressed in the brain (Ganat et
379 al., 2012). A SNP in this gene is reported to be associated with brain processing speed in the
380 Lothian birth cohort (Luciano et al., 2011). Brain processing speed is an important cognitive
381 function that is compromised in psychiatric illness such as schizophrenia and depression, and
382 old age. Also, a SNP in the *MYRIP* gene region is associated with sleep duration (Gottlieb et
383 al., 2007). Sleep durations outside the normal range (both short sleep and long sleep) is
384 significantly associated with increased risk of depression (Mohan et al., 2017; Roberts and
385 Duong, 2014; Watson et al., 2014; Zhai et al., 2015). The *MYRIP* gene is also reported to have
386 a role in insulin secretion (Waselle et al., 2003) and low insulin levels have been linked to
387 depression (Greenwood et al., 2015; Pearson et al., 2010; Webb et al., 2017).

388 **Comparison with published GWAS SNPs**

389 For both traits, the SNPs in the regions that were significant at p-value $< 5 \times 10^{-5}$
390 were compared to SNPs reported in the GWAS catalogue (MacArthur et al., 2017) to be
391 significant for the two traits. The GWAS catalogue was accessed on the 15th of January 2021.
392 The results are presented in Table 2. The SNP-based and haplotype-based models identified
393 1,380 and 45 SNPs respectively for height, and 78 and 495 SNPs respectively for MDD taking
394 all SNPs within haplotype blocks significant at p-value $< 5 \times 10^{-5}$. Out of the 1,380 SNPs

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

397 **Discussion**

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

410 We hypothesised that the Hap-RHM would complement the SNP-RHM. We
411 investigated this hypothesis in a simulation study in which we simulated 20 replicates each of
412 two types of SNP QTL phenotypes and three types of haplotype QTL phenotypes. The results
413 show that the two GREML models can capture the effects of causal variants within genomic
414 loci associated with the phenotype analysed. The results also show that the two models are
415 specific about the type of causal effect they can capture, therefore, providing support for the
416 hypothesis that haplotype-based regional GREML models will complement SNP-based
417 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

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

442 In both traits, the top significant regions we mapped at p-value $< 5 \times 10^{-5}$ had genes
443 mapped to those regions or within 400kb of those regions. For height, these genes have been
444 reported to be associated with height in humans (Gudbjartsson et al., 2008; Kichaev et al.,
445 2019; Lango Allen et al., 2010; Nagy et al., 2017; Tachmazidou et al., 2017; Weedon et al.,
446 2008; Wood et al., 2014). For MDD, these genes have been reported to be associated with
447 major depressive disorder and other psychiatry phenotypes (Arnau-Soler et al., 2019; Howard
448 et al., 2019; Liu et al., 2019; Luciano et al., 2011; Wray et al., 2018; Zeng et al., 2017). In one
449 of such regions for MDD, five SNPs within the region are individually significantly associated
450 with MDD at the nominal level (p-value < 0.05). Four of these SNPs lie within the gene
451 sequence of *MYRIP*, and they each confer 2% disease risk. A conventional GWAS analysis
452 would have missed these nominally associated SNPs because they will not reach genome-
453 wide (GW) significance. However, analysing these SNPs within the region as haplotypes, gave
454 us the power to detect the combined effect of these SNPs in the region at a suggestive-
455 significance level even with our relatively small sample size compared to recent genome-wide
456 association studies of MDD: 322,580 (Howard et al., 2018), and 480,359 (Wray et al., 2018).

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

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

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

482 **Conclusions**

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

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

495 **Acknowledgements**

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

509 **Conflict of interest**

510 The authors declare no conflict of interest.

511 **References**

512 Amador, C., Huffman, J., Trochet, H., Campbell, A., Porteous, D., Wilson, J.F., Hastie, N.,
513 Vitart, V., Hayward, C., Navarro, P., Haley, C.S., 2015. Recent genomic heritage in
514 Scotland. BMC Genomics 16, 1–17. https://doi.org/10.1186/s12864-015-1605-2
515 Arnau-Soler, A., Macdonald-Dunlop, E., Adams, M.J., Clarke, T.-K., MacIntyre, D.J., Milburn,
516 K., Navrady, L., Generation Scotland, Major Depressive Disorder Working Group of
517 the Psychiatric Genomics Consortium, Hayward, C., McIntosh, A.M., Thomson, P.A.,
518 2019. Genome-wide by environment interaction studies of depressive symptoms and
519 psychosocial stress in UK Biobank and Generation Scotland. Transl Psychiatry 9, 14.
520 https://doi.org/10.1038/s41398-018-0360-y
521 Balding, D.J., 2006. A tutorial on statistical methods for population association studies.
522 Nature Reviews Genetics 7, 781–791. https://doi.org/10.1038/nrg1916
523 Caulfield, T., Evans, J., McGuire, A., McCabe, C., Bubela, T., Cook-Deegan, R., Fishman, J.,
524 Hogarth, S., Miller, F.A., Ravitsky, V., Biesecker, B., Borry, P., Cho, M.K., Carroll, J.C.,
525 Etchegary, H., Joly, Y., Kato, K., Lee, S.S.-J., Rothenberg, K., Sankar, P., Szego, M.J.,
526 Ossorio, P., Pullman, D., Rousseau, F., Ungar, W.J., Wilson, B., 2013. Reflections on
527 the Cost of "Low-Cost" Whole Genome Sequencing: Framing the Health Policy
528 Debate. PLOS Biology 11, e1001699. https://doi.org/10.1371/journal.pbio.1001699
529 Cirulli, E.T., Goldstein, D.B., 2010. Uncovering the roles of rare variants in common disease
530 through whole-genome sequencing. Nat Rev Genet 11, 415–425.
531 https://doi.org/10.1038/nrg2779
532 Clarke, A.J., Cooper, D.N., 2010. GWAS: heritability missing in action? Eur J Hum Genet 18,
533 859–861. https://doi.org/10.1038/ejhg.2010.35
534 Delaneau, O., Zagury, J.-F., Marchini, J., 2013. Improved whole-chromosome phasing for
535 disease and population genetic studies. Nature Methods 10, 5–6.
536 https://doi.org/10.1038/nmeth.2307
537 First, M.B., Spitzer, R.L., Gibbon, M., Williams, J.B.W., 2002. Structured Clinical Interview for
538 DSM-IV-TR Axis I Disorders, Research Version, Non-patient Edition.
539 Ganat, Y.M., Calder, E.L., Kriks, S., Nelander, J., Tu, E.Y., Jia, F., Battista, D., Harrison, N.,
540 Parmar, M., Tomishima, M.J., Rutishauser, U., Studer, L., 2012. Identification of
541 embryonic stem cell-derived midbrain dopaminergic neurons for engraftment. J.
542 Clin. Invest. 122, 2928–2939. https://doi.org/10.1172/JCI58767
543 Gonzalez-Recio, O., Daetwyler, H.D., MacLeod, I.M., Pryce, J.E., Bowman, P.J., Hayes, B.J.,
544 Goddard, M.E., 2015. Rare Variants in Transcript and Potential Regulatory Regions
545 Explain a Small Percentage of the Missing Heritability of Complex Traits in Cattle.
546 PLoS ONE 10, e0143945. https://doi.org/10.1371/journal.pone.0143945
547 Gottlieb, D.J., O'Connor, G.T., Wilk, J.B., 2007. Genome-wide association of sleep and
548 circadian phenotypes. BMC Med Genet 8, S9. https://doi.org/10.1186/1471-2350-8-
549 S1-S9

550 Greenwood, E.A., Pasch, L.A., Shinkai, K., Cedars, M.I., Huddleston, H.G., 2015. Putative role
551 for insulin resistance in depression risk in polycystic ovary syndrome. Fertility and
552 Sterility 104, 707-714.e1. https://doi.org/10.1016/j.fertnstert.2015.05.019
553 Gudbjartsson, D.F., Walters, G.B., Thorleifsson, G., Stefansson, H., Halldorsson, B.V.,
554 Zusmanovich, P., Sulem, P., Thorlacius, S., Gylfason, A., Steinberg, S., Helgadottir, A.,
555 Ingason, A., Steinthorsdottir, V., Olafsdottir, E.J., Olafsdottir, G.H., Jonsson, T., Borch-
556 Johnsen, K., Hansen, T., Andersen, G., Jorgensen, T., Pedersen, O., Aben, K.K., Witjes,
557 J.A., Swinkels, D.W., den Heijer, M., Franke, B., Verbeek, A.L.M., Becker, D.M., Yanek,
558 L.R., Becker, L.C., Tryggvadottir, L., Rafnar, T., Gulcher, J., Kiemeney, L.A., Kong, A.,
559 Thorsteinsdottir, U., Stefansson, K., 2008. Many sequence variants affecting diversity
560 of adult human height. Nat Genet 40, 609–615. https://doi.org/10.1038/ng.122
561 Höglund, J., Rafati, N., Rask-Andersen, M., Enroth, S., Karlsson, T., Ek, W.E., Johansson, Å.,
562 2019. Improved power and precision with whole genome sequencing data in
563 genome-wide association studies of inflammatory biomarkers. Scientific Reports 9,
564 16844. https://doi.org/10.1038/s41598-019-53111-7
565 Howard, D.M., Adams, M.J., Clarke, T.-K., Hafferty, J.D., Gibson, J., Shirali, M., Coleman,
566 J.R.I., Hagenaars, S.P., Ward, J., Wigmore, E.M., Alloza, C., Shen, X., Barbu, M.C., Xu,
567 E.Y., Whalley, H.C., Marioni, R.E., Porteous, D.J., Davies, G., Deary, I.J., Hemani, G.,
568 Berger, K., Teismann, H., Rawal, R., Arolt, V., Baune, B.T., Dannlowski, U., Domschke,
569 K., Tian, C., Hinds, D.A., 23andMe Research Team, Major Depressive Disorder
570 Working Group of the Psychiatric Genomics Consortium, Trzaskowski, M., Byrne,
571 E.M., Ripke, S., Smith, D.J., Sullivan, P.F., Wray, N.R., Breen, G., Lewis, C.M.,
572 McIntosh, A.M., 2019. Genome-wide meta-analysis of depression identifies 102
573 independent variants and highlights the importance of the prefrontal brain regions.
574 Nat Neurosci 22, 343–352. https://doi.org/10.1038/s41593-018-0326-7
575 Howard, D.M., Adams, M.J., Shirali, M., Clarke, T.-K., Marioni, R.E., Davies, G., Coleman,
576 J.R.I., Alloza, C., Shen, X., Barbu, M.C., Wigmore, E.M., Gibson, J., Hagenaars, S.P.,
577 Lewis, C.M., Ward, J., Smith, D.J., Sullivan, P.F., Haley, C.S., Breen, G., Deary, I.J.,
578 McIntosh, A.M., 2018. Genome-wide association study of depression phenotypes in
579 UK Biobank identifies variants in excitatory synaptic pathways. Nature
580 Communications 9, 1470. https://doi.org/10.1038/s41467-018-03819-3
581 International Human Genome Sequencing Consortium, 2004. Finishing the euchromatic
582 sequence of the human genome. Nature 431, 931–945.
583 https://doi.org/10.1038/nature03001
584 Kichaev, G., Bhatia, G., Loh, P.-R., Gazal, S., Burch, K., Freund, M.K., Schoech, A., Pasaniuc,
585 B., Price, A.L., 2019. Leveraging Polygenic Functional Enrichment to Improve GWAS
586 Power. Am J Hum Genet 104, 65–75. https://doi.org/10.1016/j.ajhg.2018.11.008
587 Lango Allen, H., Estrada, K., Lettre, G., Berndt, S.I., Weedon, M.N., Rivadeneira, F., Willer,
588 C.J., Jackson, A.U., Vedantam, S., Raychaudhuri, S., Ferreira, T., Wood, A.R., Weyant,
589 R.J., Segrè, A.V., Speliotes, E.K., Wheeler, E., Soranzo, N., Park, J.-H., Yang, J.,
590 Gudbjartsson, D., Heard-Costa, N.L., Randall, J.C., Qi, L., Vernon Smith, A., Mägi, R.,
591 Pastinen, T., Liang, L., Heid, I.M., Luan, J., Thorleifsson, G., Winkler, T.W., Goddard,
592 M.E., Sin Lo, K., Palmer, C., Workalemahu, T., Aulchenko, Y.S., Johansson, A.,
593 Zillikens, M.C., Feitosa, M.F., Esko, T., Johnson, T., Ketkar, S., Kraft, P., Mangino, M.,
594 Prokopenko, I., Absher, D., Albrecht, E., Ernst, F., Glazer, N.L., Hayward, C., Hottenga,
595 J.-J., Jacobs, K.B., Knowles, J.W., Kutalik, Z., Monda, K.L., Polasek, O., Preuss, M.,
596 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.,

Jang, S.-K., Jansen, P.R., Ling, Y., Mägi, R., Matoba, N., McMahon, G., Mulas, A., Orrù, V., Palviainen, T., Pandit, A., Reginsson, G.W., Skogholt, A.H., Smith, J.A., Taylor, A.E., Turman, C., Willemsen, G., Young, H., Young, K.A., Zajac, G.J.M., Zhao, W., Zhou, W., Bjornsdottir, G., Boardman, J.D., Boehnke, M., Boomsma, D.I., Chen, C., Cucca, F., Davies, G.E., Eaton, C.B., Ehringer, M.A., Esko, T., Fiorillo, E., Gillespie, N.A., Gudbjartsson, D.F., Haller, T., Harris, K.M., Heath, A.C., Hewitt, J.K., Hickie, I.B., Hokanson, J.E., Hopfer, C.J., Hunter, D.J., Iacono, W.G., Johnson, E.O., Kamatani, Y., Kardia, S.L.R., Keller, M.C., Kellis, M., Kooperberg, C., Kraft, P., Krauter, K.S., Laakso, M., Lind, P.A., Loukola, A., Lutz, S.M., Madden, P.A.F., Martin, N.G., McGue, M., McQueen, M.B., Medland, S.E., Metspalu, A., Mohlke, K.L., Nielsen, J.B., Okada, Y., Peters, U., Polderman, T.J.C., Posthuma, D., Reiner, A.P., Rice, J.P., Rimm, E., Rose, R.J., Runarsdottir, V., Stallings, M.C., Stančáková, A., Stefansson, H., Thai, K.K., Tindle, H.A., Tyrfingsson, T., Wall, T.L., Weir, D.R., Weisner, C., Whitfield, J.B., Winsvold, B.S., Yin, J., Zuccolo, L., Bierut, L.J., Hveem, K., Lee, J.J., Munafo, M.R., Saccone, N.L., Willer, C.J., Cornelis, M.C., David, S.P., Hinds, D., Jorgenson, E., Kaprio, J., Stitzel, J.A., Stefansson, K., Thorgeirsson, T.E., Abecasis, G., Liu, D.J., Vrieze, S., 2019. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. Nat Genet 51, 237–244. https://doi.org/10.1038/s41588-018-0307-5

Luciano, M., Hansell, N.K., Lahti, J., Davies, G., Medland, S.E., Räikkönen, K., Tenesa, A., Widen, E., McGhee, K.A., Palotie, A., Liewald, D., Porteous, D.J., Starr, J.M., Montgomery, G.W., Martin, N.G., Eriksson, J.G., Wright, M.J., Deary, I.J., 2011. Whole genome association scan for genetic polymorphisms influencing information processing speed. Biol Psychol 86, 193–202. https://doi.org/10.1016/j.biopsycho.2010.11.008

MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., Junkins, H., McMahon, A., Milano, A., Morales, J., Pendlington, Z.M., Welter, D., Burdett, T., Hindorff, L., Flicek, P., Cunningham, F., Parkinson, H., 2017. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res 45, D896–D901. https://doi.org/10.1093/nar/gkw1133

Maher, B., 2008. Personal genomes: The case of the missing heritability. Nature News 456, 18–21. https://doi.org/10.1038/456018a

Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., Cho, J.H., Guttmacher, A.E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C.N., Slatkin, M., Valle, D., Whittemore, A.S., Boehnke, M., Clark, A.G., Eichler, E.E., Gibson, G., Haines, J.L., Mackay, T.F.C., McCarroll, S.A., Visscher, P.M., 2009. Finding the missing heritability of complex diseases. Nature 461, 747–753. https://doi.org/10.1038/nature08494

Mohan, J., Xiaofan, G., Yingxian, S., 2017. Association between sleep time and depression: a cross-sectional study from countries in rural Northeastern China. J Int Med Res 45, 984–992. https://doi.org/10.1177/0300060517701034

Nagamine, Y., Pong-Wong, R., Navarro, P., Vitart, V., Hayward, C., Rudan, I., Campbell, H., Wilson, J., Wild, S., Hicks, A.A., Pramstaller, P.P., Hastie, N., Wright, A.F., Haley, C.S., 2012. Localising Loci underlying Complex Trait Variation Using Regional Genomic Relationship Mapping. PLoS ONE 7, e46501. 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,

N.J., Zeggini, E., 2017. Whole-Genome Sequencing Coupled to Imputation Discovers Genetic Signals for Anthropometric Traits. Am J Hum Genet 100, 865–884. https://doi.org/10.1016/j.ajhg.2017.04.014

Vormfelde, S.V., Brockmöller, J., 2007. On the value of haplotype-based genotype-phenotype analysis and on data transformation in pharmacogenetics and -genomics. Nat. Rev. Genet. 8. https://doi.org/10.1038/nrg1916-c1

Wainschtein, P., Jain, D.P., Yengo, L., Zheng, Z., TOPMed Anthropometry Working Group, T.-O. for P.M.C., Cupples, L.A., Shadyab, A.H., McKnight, B., Shoemaker, B.M., Mitchell, B.D., Psaty, B.M., Kooperberg, C., Roden, D., Darbar, D., Arnett, D.K., Regan, E.A., Boerwinkle, E., Rotter, J.I., Allison, M.A., McDonald, M.-L.N., Chung, M.K., Smith, N.L., Ellinor, P.T., Vasan, R.S., Mathias, R.A., Rich, S.S., Heckbert, S.R., Redline, S., Guo, X., Chen, Y.-D.I., Liu, C.-T., Andrade, M. de, Yanek, L.R., Albert, C.M., Hernandez, R.D., McGarvey, S.T., North, K.E., Lange, L.A., Weir, B.S., Laurie, C.C., Yang, J., Visscher, P.M., 2019. Recovery of trait heritability from whole genome sequence data. bioRxiv 588020. https://doi.org/10.1101/588020

Waselle, L., Coppola, T., Fukuda, M., Iezzi, M., El-Amraoui, A., Petit, C., Regazzi, R., 2003. Involvement of the Rab27 Binding Protein Slac2c/MyRIP in Insulin Exocytosis. Mol Biol Cell 14, 4103–4113. https://doi.org/10.1091/mbc.E03-01-0022

Watson, N.F., Harden, K.P., Buchwald, D., Vitiello, M.V., Pack, A.I., Strachan, E., Goldberg, J., 2014. Sleep Duration and Depressive Symptoms: A Gene-Environment Interaction. Sleep 37, 351–358. https://doi.org/10.5665/sleep.3412

Webb, M., Davies, M., Ashra, N., Bodicoat, D., Brady, E., Webb, D., Moulton, C., Ismail, K., Khunti, K., 2017. The association between depressive symptoms and insulin resistance, inflammation and adiposity in men and women. PLOS ONE 12, e0187448. https://doi.org/10.1371/journal.pone.0187448

Weedon, M.N., Lango, H., Lindgren, C.M., Wallace, C., Evans, D.M., Mangino, M., Freathy, R.M., Perry, J.R.B., Stevens, S., Hall, A.S., Samani, N.J., Shields, B., Prokopenko, I., Farrall, M., Dominiczak, A., Johnson, T., Bergmann, S., Beckmann, J.S., Vollenweider, P., Waterworth, D.M., Mooser, V., Palmer, C.N.A., Morris, A.D., Ouwehand, W.H., Zhao, J.H., Li, S., Loos, R.J.F., Barroso, I., Deloukas, P., Sandhu, M.S., Wheeler, E., Soranzo, N., Inouye, M., Wareham, N.J., Caulfield, M., Munroe, P.B., Hattersley, A.T., McCarthy, M.I., Frayling, T.M., 2008. Genome-wide association analysis identifies 20 loci that influence adult height. Nat Genet 40, 575–583. https://doi.org/10.1038/ng.121

Wood, A.R., Esko, T., Yang, J., Vedantam, S., Pers, T.H., Gustafsson, S., Chu, A.Y., Estrada, K., Luan, J., Kutalik, Z., Amin, N., Buchkovich, M.L., Croteau-Chonka, D.C., Day, F.R., Duan, Y., Fall, T., Fehrmann, R., Ferreira, T., Jackson, A.U., Karjalainen, J., Lo, K.S., Locke, A.E., Mägi, R., Mihailov, E., Porcu, E., Randall, J.C., Scherag, A., Vinkhuyzen, A.A.E., Westra, H.-J., Winkler, T.W., Workalemahu, T., Zhao, J.H., Absher, D., Albrecht, E., Anderson, D., Baron, J., Beekman, M., Demirkan, A., Ehret, G.B., Feenstra, B., Feitosa, M.F., Fischer, K., Fraser, R.M., Goel, A., Gong, J., Justice, A.E., Kanoni, S., Kleber, M.E., Kristiansson, K., Lim, U., Lotay, V., Lui, J.C., Mangino, M., Leach, I.M., Medina-Gomez, C., Nalls, M.A., Nyholt, D.R., Palmer, C.D., Pasko, D., Pechlivanis, S., Prokopenko, I., Ried, J.S., Ripke, S., Shungin, D., Stancáková, A., Strawbridge, R.J., Sung, Y.J., Tanaka, T., Teumer, A., Trompet, S., Laan, S.W. van der, Setten, J. van, Vliet-Ostaptchouk, J.V.V., Wang, Z., Yengo, L., Zhang, W., Afzal, U., Ärnlöv, J., Arscott, G.M., Bandinelli, S., Barrett, A., Bellis, C., Bennett, A.J., Berne, C.,

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

O'Donovan, M.C., Paciga, S.A., Pedersen, N.L., Penninx, B.W.J.H., Perlis, R.H., Porteous, D.J., Potash, J.B., Preisig, M., Rietschel, M., Schaefer, C., Schulze, T.G., Smoller, J.W., Stefansson, K., Tiemeier, H., Uher, R., Völzke, H., Weissman, M.M., Werge, T., Winslow, A.R., Lewis, C.M., Levinson, D.F., Breen, G., Børglum, A.D., Sullivan, P.F., 2018. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nature Genetics 50, 668–681. https://doi.org/10.1038/s41588-018-0090-3

Yang, J., Bakshi, A., Zhu, Z., Hemani, G., Vinkhuyzen, A.A.E., Lee, S.H., Robinson, M.R., Perry, J.R.B., Nolte, I.M., van Vliet-Ostaptchouk, J.V., Snieder, H., Esko, T., Milani, L., Mägi, R., Metspalu, A., Hamsten, A., Magnusson, P.K.E., Pedersen, N.L., Ingelsson, E., Soranzo, N., Keller, M.C., Wray, N.R., Goddard, M.E., Visscher, P.M., 2015. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nature Genetics 47, 1114–1120. https://doi.org/10.1038/ng.3390

Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., Goddard, M.E., Visscher, P.M., 2010. Common SNPs explain a large proportion of the heritability for human height. Nat Genet 42, 565–569. https://doi.org/10.1038/ng.608

Yang, J., Lee, S.H., Goddard, M.E., Visscher, P.M., 2011. GCTA: A Tool for Genome-wide Complex Trait Analysis. The American Journal of Human Genetics 88, 76–82. https://doi.org/10.1016/j.ajhg.2010.11.011

Zeng, Y., Navarro, P., Fernandez-Pujals, A.M., Hall, L.S., Clarke, T.-K., Thomson, P.A., Smith, B.H., Hocking, L.J., Padmanabhan, S., Hayward, C., MacIntyre, D.J., Wray, N.R., Deary, I.J., Porteous, D.J., Haley, C.S., McIntosh, A.M., 2017. A Combined Pathway and Regional Heritability Analysis Indicates NETRIN1 Pathway Is Associated With Major Depressive Disorder. Biol Psychiatry 81, 336–346. https://doi.org/10.1016/j.biopsych.2016.04.017

Zhai, L., Zhang, H., Zhang, D., 2015. SLEEP DURATION AND DEPRESSION AMONG ADULTS: A META-ANALYSIS OF PROSPECTIVE STUDIES. Depress Anxiety 32, 664–670. https://doi.org/10.1002/da.22386

915 Table 1. SNP-based association test of MDD in the *MYRIP* gene region. The columns are the SNP ID,
916 chromosome, genome position of SNP, minor allele frequency, odds ratio, log of odds ratio, standard
917 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

918

919

920 Table 2. Comparison of SNPs within significant regions identified by both models and published GWAS
921 results for height and MDD. The columns are the name of trait, number of SNPs in regions identified
922 by SNP-RHM and Hap-RHM with p-value $< 5 \times 10^{-5}$ and SNPs in published GWAS (pubGWAS) for the
923 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

924

925

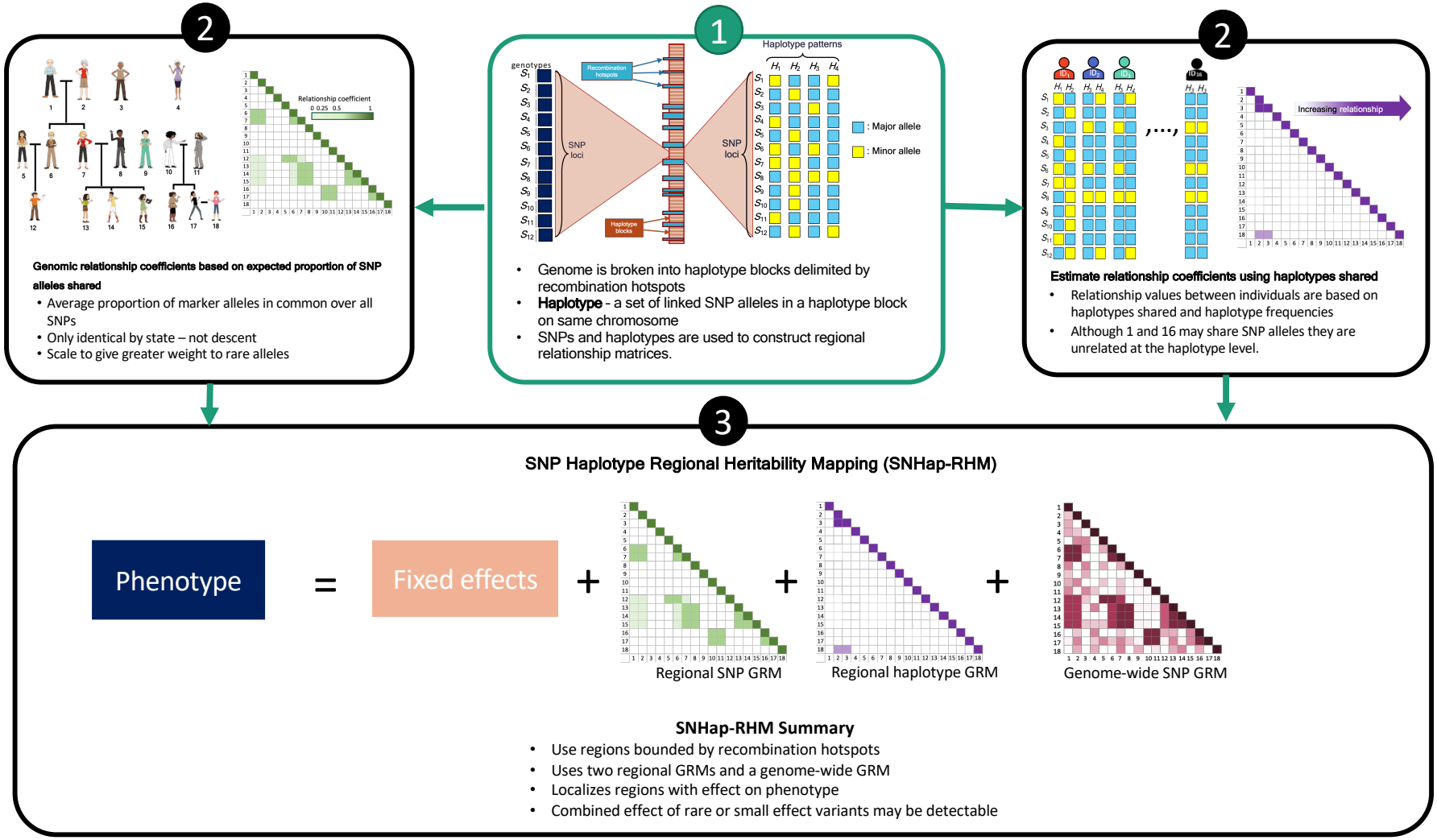

Figure 1. A Schema outlying SNHap-RHM

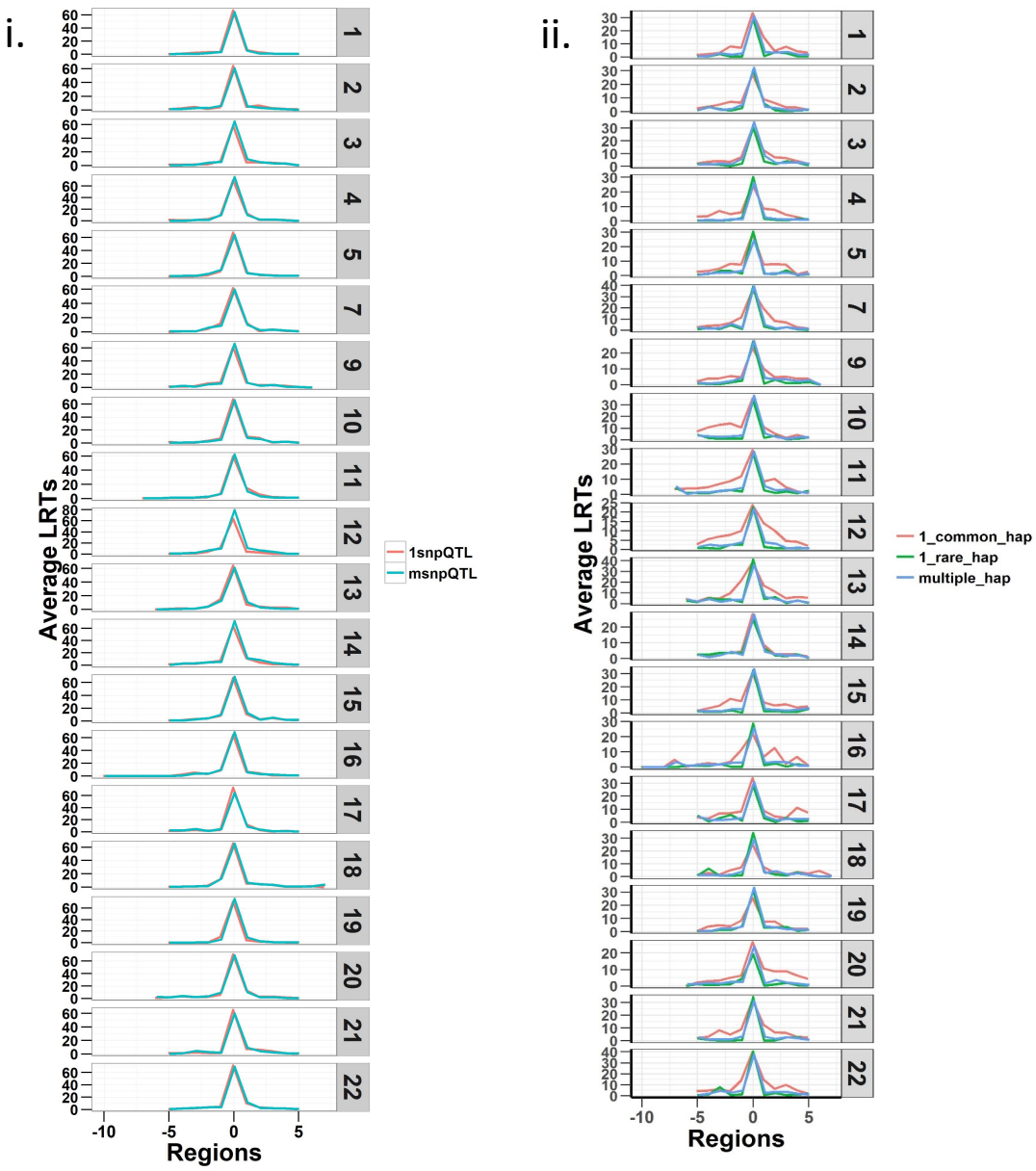

Figure 2. Plots of Likelihood ratio test (LRT) statistics at each QTL loci and 5 regions either side averaged for the 20 simulations of each of the five QTL phenotypes. Plot (i) is SNP QTL phenotypes analysed using the SNP-RHM and plot (ii) is the haplotype QTL phenotypes analysed using the Hap-RHM.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

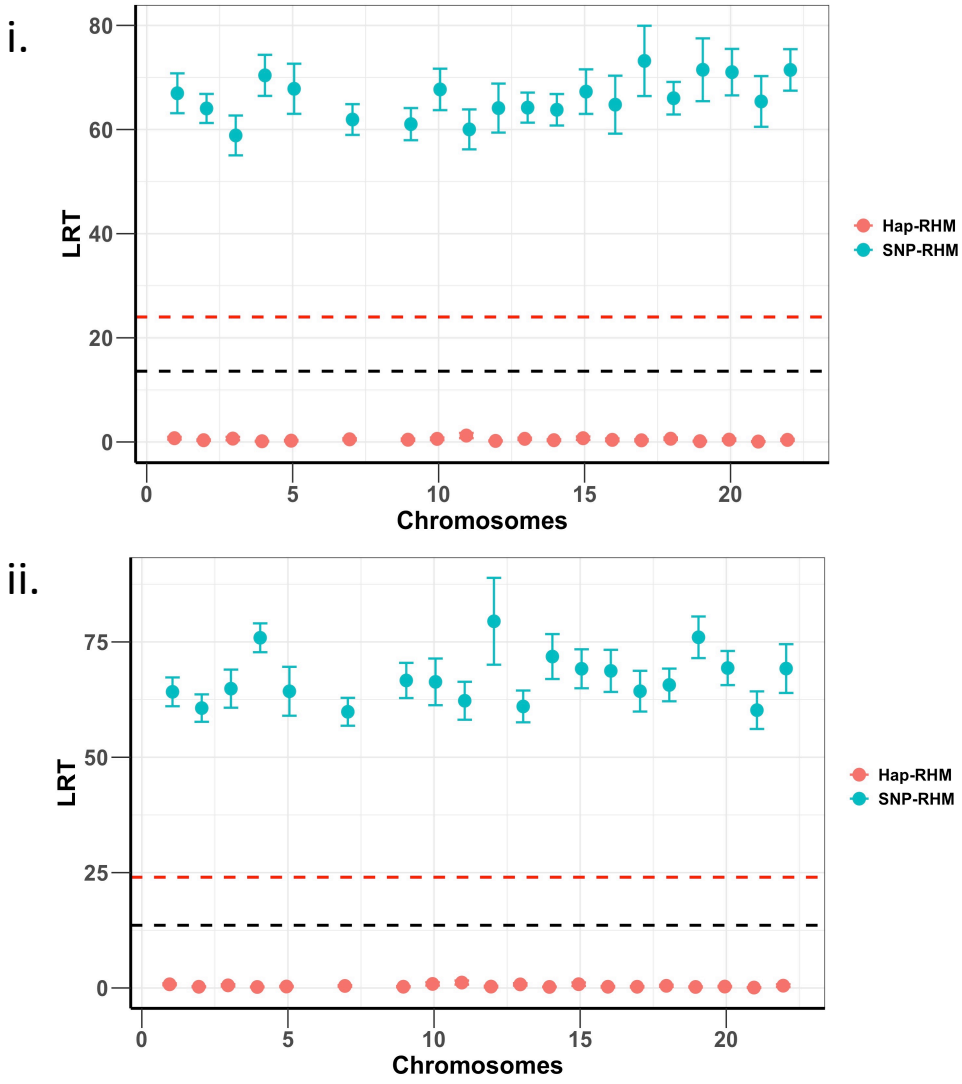

Figure 3. Plots of average LRT statistics over replicates of QTL loci across the chromosomes for the 20 simulations of each of the two SNP QTL phenotypes. The red dashed lines are genome-wide significance threshold (for 48,772 regions) and the black dashed lines are Bonferroni significance threshold (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.

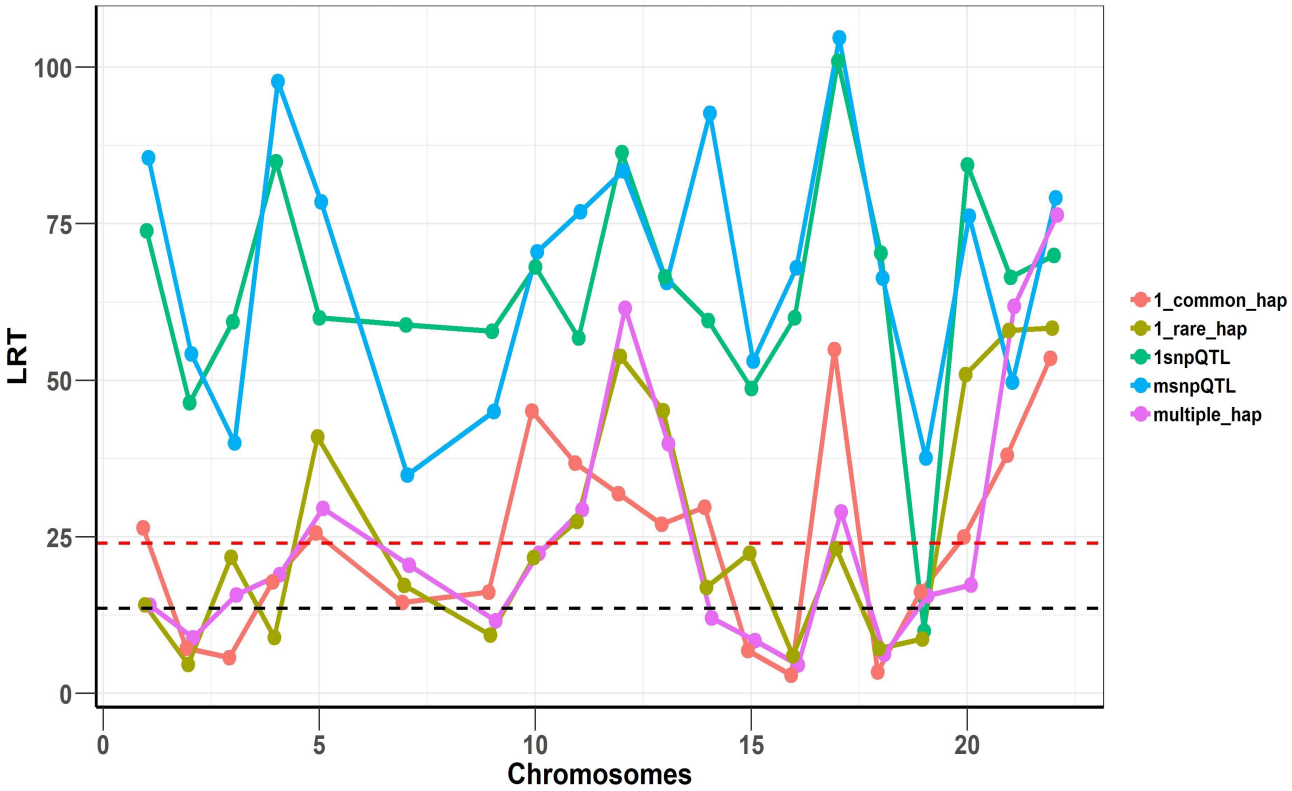

Figure 4. Joint analysis of the SNP and haplotype phenotypes using SNHap-RHM. The plot is an analysis of one replicate of each of the simulated phenotypes. The LRT statistics are plotted over QTL loci across the chromosomes. The red dashed lines are genome-wide significance threshold (for 48,772 regions)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

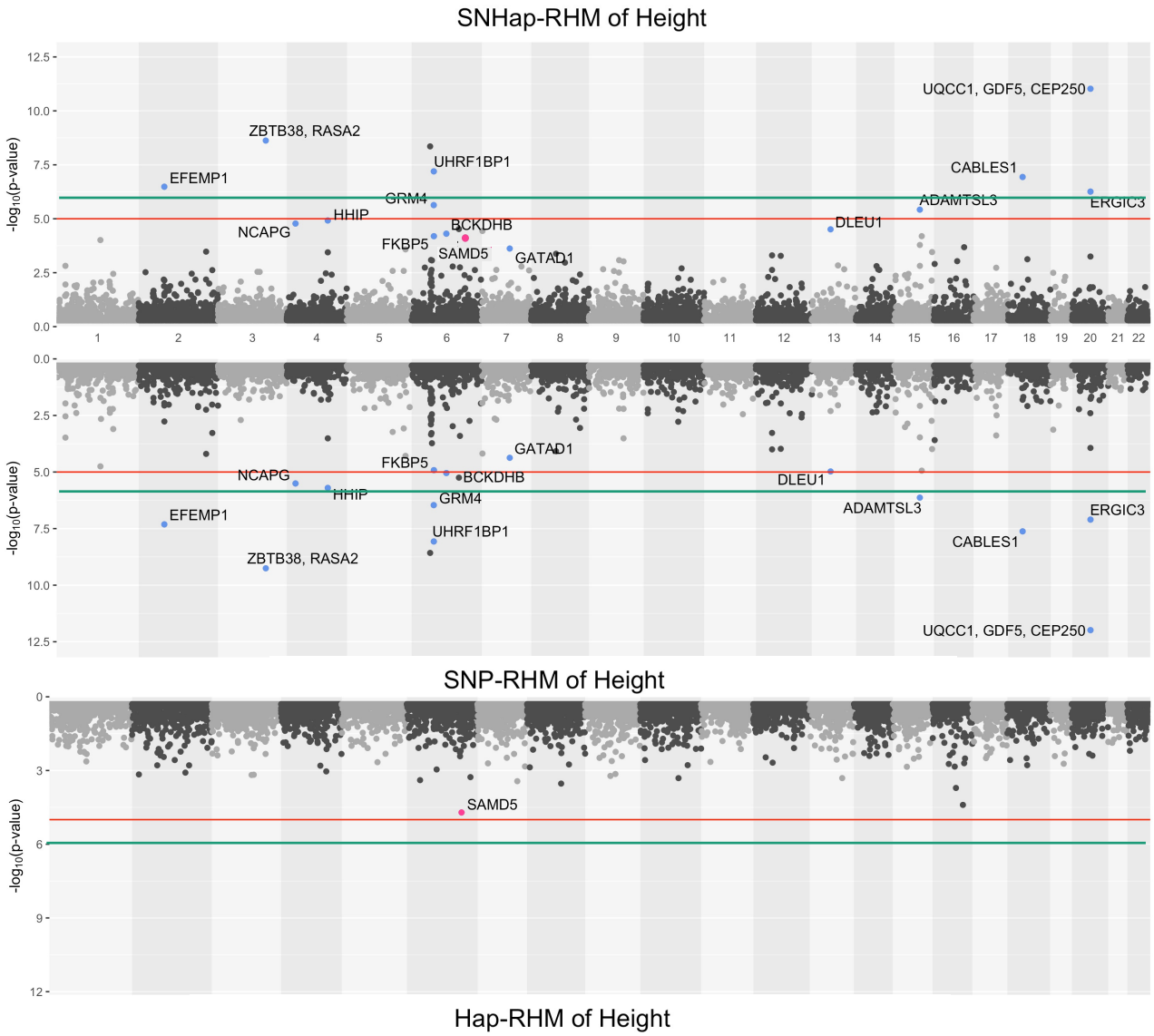

Figure 5. The genome-wide evidence of haplotype block association for height. Analysis done with SNHap-RHM, SNP-RHM and Hap-RHM. The points are plots of -log10 of the p-values of regions tested with the LRT for the regional GREML analyses. The green lines are the Bonferroni-corrected genome-wide significance 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.

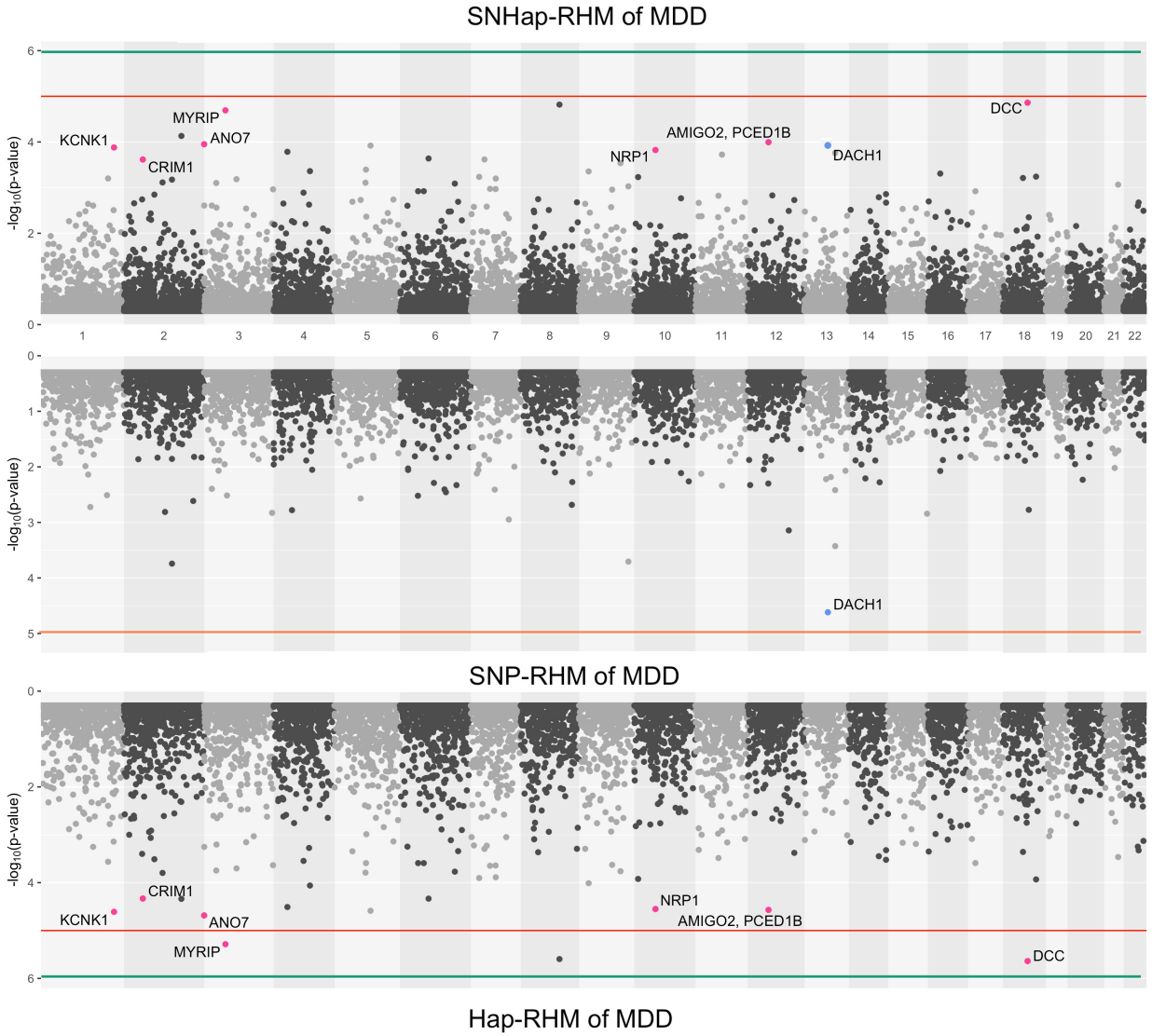

Figure 6. The genome-wide evidence of haplotype block association for Major Depressive Disorder. Analysis done with SNHap-RHM, SNP-RHM and Hap-RHM. The points are plots of -log10 of the p-values of regions tested with the LRT for the regional GREML analyses. The green lines are the Bonferroni-corrected 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.