

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

### 1    **Genetic meta-analysis identifies 9 novel loci and functional pathways for Alzheimer's disease 2    risk**

3  
4    Iris E Jansen<sup>1,2</sup>, Jeanne E Savage<sup>1</sup>, Kyoko Watanabe<sup>1</sup>, Julien Bryois<sup>3</sup>, Dylan M Williams<sup>3</sup>, Stacy  
5    Steinberg<sup>4</sup>, Julia Sealock<sup>5</sup>, Ida K Karlsson<sup>3</sup>, Sara Hägg<sup>3</sup>, Lavinia Athanasiou<sup>6,7</sup>, Nicola Voyle<sup>8</sup>,  
6    Petroula Proitsi<sup>8</sup>, Aree Witoelar<sup>6,9</sup>, Sven Stringer<sup>1</sup>, Dag Aarsland<sup>8,10</sup>, Ina S Almdahl<sup>11-13</sup>, Fred  
7    Andersen<sup>14</sup>, Sverre Bergh<sup>15,16</sup>, Francesco Bettella<sup>6,9</sup>, Sigurbjorn Bjornsson<sup>17</sup>, Anne Brækhus<sup>15,18</sup>,  
8    Geir Bråthen<sup>19,20</sup>, Christiaan de Leeuw<sup>1</sup>, Rahul S Desikan<sup>21</sup>, Srdjan Djurovic<sup>6,22</sup>, Logan  
9    Dumitrescu<sup>23</sup>, Tormod Fladby<sup>11,12</sup>, Timothy Homan<sup>23</sup>, Palmi V Jonsson<sup>17,24</sup>, Steven J Kiddle<sup>25</sup>, K  
10   Arvid Rongve<sup>27,28</sup>, Ingvild Saltvedt<sup>19,28</sup>, Sigrid B. Sandø<sup>19,20</sup>, Geir Selbæk<sup>15,29</sup>, Nathan Skenne<sup>30</sup>,  
11   Jon Snaedal<sup>17</sup>, Eystein Stordal<sup>31,32</sup>, Ingun D. Ulstein<sup>33</sup>, Yunpeng Wang<sup>6,9</sup>, Linda R White<sup>19,20</sup>, Jens  
12   Hjerling-Leffler<sup>30</sup>, Patrick F Sullivan<sup>3,34,35</sup>, Wiesje M van der Flier<sup>2</sup>, Richard Dobson<sup>8,36</sup>, Lea K.  
13   Davis<sup>37,38</sup>, Hreinn Stefansson<sup>4</sup>, Kari Stefansson<sup>4</sup>, Nancy L Pedersen<sup>3</sup>, Stephan Ripke<sup>39-41\*</sup>, Ole A  
14   Andreassen<sup>6,9\*</sup>, Danielle Posthuma<sup>1,42,\*#</sup>

- 15  
16   1. Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research,  
17   Amsterdam Neuroscience, VU University, Amsterdam, The Netherlands.  
18   2. Alzheimer Center and Department of Neurology, Amsterdam Neuroscience, VU University  
19   Medical Center, Amsterdam, The Netherlands.  
20   3. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm,  
21   Sweden.  
22   4. deCODE Genetics/Amgen, Reykjavik, Iceland.  
23   5. Interdisciplinary Graduate Program, Vanderbilt University, Nashville, USA.  
24   6. NORMENT, K.G. Jebsen Centre for Psychosis Research, Institute of Clinical Medicine,  
25   University of Oslo, Oslo, Norway.  
26   7. Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway.  
27   8. Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK.  
28   9. Institute of Clinical Medicine, University of Oslo, Oslo, Norway  
29   10. Center for Age-Related Diseases, Stavanger University Hospital, Stavanger, Norway.  
30   11. Department of Neurology, Akershus University Hospital, Lørenskog, Norway.  
31   12. AHUS Campus, University of Oslo, Oslo, Norway.  
32   13. Department of Psychiatry of Old Age, Oslo University Hospital, Oslo, Norway.  
33   14. Department of Community Medicine, University of Tromsø, Tromsø, Norway.  
34   15. Norwegian National Advisory Unit on Ageing and Health, Vestfold Hospital Trust, Tønsberg,  
35   Norway.  
36   16. Centre for Old Age Psychiatry Research, Innlandet Hospital Trust, Ottestad, Norway.  
37   17. Department of Geriatric Medicine, Landspítali University Hospital, Reykjavík, Iceland.  
38   18. Geriatric Department, University Hospital Oslo and University of Oslo, Oslo, Norway.  
39   19. Department of Neuroscience, Norwegian University of Science and Technology, Trondheim,  
40   Norway.  
41   20. Department of Neurology, St Olav's Hospital, Trondheim University Hospital, Trondheim,  
42   Norway.  
43   21. Neuroradiology Section, Department of Radiology and Biomedical Imaging, University of  
44   California, San Francisco, USA.

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

45 22. Department of Medical Genetics, Oslo University Hospital, Oslo, Norway.  
46 23. Vanderbilt Memory & Alzheimer's Center, Department of Neurology, Vanderbilt University  
47 Medical Center, Nashville, USA.  
48 24. Faculty of Medicine, University of Iceland, Reykjavik, Iceland.  
49 25. MRC Biostatistics Unit, Cambridge Institute of Public Health, University of Cambridge,  
50 Cambridge, UK.  
51 26. Department of Research and Innovation, Helse Fonna, Oslo, Norway.  
52 27. Department of Clinical Medicine, University of Bergen, Bergen, Norway.  
53 28. Department of Geriatrics, St. Olav's Hospital, Trondheim University Hospital, Trondheim,  
54 Norway.  
55 29. Institute of Health and Society, University of Oslo, Oslo, Norway.  
56 30. Laboratory of Molecular Neurobiology, Department of Medical Biochemistry and  
57 Biophysics, Karolinska Institutet, Stockholm, Sweden.  
58 31. Department of Psychiatry, Namsos Hospital, Namsos, Norway.  
59 32. Department of Mental Health, Norwegian University of Science and Technology, Trondheim,  
60 Norway.  
61 33. Memory Clinic, Geriatric Department, Oslo University Hospital, Oslo, Norway.  
62 34. Department of Genetics, University of North Carolina, Chapel Hill, USA.  
63 35. Department of Psychiatry, University of North Carolina, Chapel Hill, USA.  
64 36. Farr Institute of Health Informatics Research, University College London, London, UK.  
65 37. Department of Medicine, Division of Genetic Medicine, Vanderbilt University Medical  
66 Center, Nashville, US.  
67 38. Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, US.  
68 39. Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, USA.  
69 40. Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge,  
70 USA.  
71 41. Dept. of Psychiatry and Psychotherapy, Charité - Universitätsmedizin, Berlin, Germany.  
72 42. Department of Clinical Genetics, VU University Medical Center, Amsterdam, The  
73 Netherlands.  
74  
75 \* These authors contributed equally to this work  
76  
77 #Correspondence to: Danielle Posthuma: Department of Complex Trait Genetics, VU  
78 University, De Boelelaan 1085, 1081 HV, Amsterdam, The Netherlands. Phone: +31 20 598  
79 2823, Fax: +31 20 5986926, d.posthuma@vu.nl  
80 **Word count:** Introductory paragraph: 208; main text: 3,890; Online methods: 5,448  
81 **Display items:** 5 (Figures 4)  
82 Includes **Supplementary Figures 1-7, Supplementary Tables 1-20**.

83  
84

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

### 85   **Abstract**

86   Late onset Alzheimer's disease (AD) is the most common form of dementia with more than 35  
87   million people affected worldwide, and no curative treatment available. AD is highly heritable  
88   and recent genome-wide meta-analyses have identified over 20 genomic loci associated with  
89   AD, yet only explaining a small proportion of the genetic variance indicating that undiscovered  
90   loci exist. Here, we performed the largest genome-wide association study of clinically diagnosed  
91   AD and AD-by-proxy (71,880 AD cases, 383,378 controls). AD-by-proxy status is based on  
92   parental AD diagnosis, and showed strong genetic correlation with AD ( $r_g=0.81$ ). Genetic meta-  
93   analysis identified 29 risk loci, of which 9 are novel, and implicating 215 potential causative  
94   genes. Independent replication further supports these novel loci in AD. Associated genes are  
95   strongly expressed in immune-related tissues and cell types (spleen, liver and microglia).  
96   Furthermore, gene-set analyses indicate the genetic contribution of biological mechanisms  
97   involved in lipid-related processes and degradation of amyloid precursor proteins. We show  
98   strong genetic correlations with multiple health-related outcomes, and Mendelian  
99   randomisation results suggest a protective effect of cognitive ability on AD risk. These results  
100   are a step forward in identifying more of the genetic factors that contribute to AD risk and add  
101   novel insights into the neurobiology of AD to guide new drug development.

102

### 103   **Main text**

104   Alzheimer's disease (AD) is the most frequent neurodegenerative disease with roughly 35  
105   million affected to date.<sup>1</sup> Results from twin studies indicate that AD is highly heritable, with  
106   estimates ranging between 60 and 80%.<sup>2</sup> Genetically, AD can be roughly divided into 2

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

107 subgroups: 1) familial early-onset cases that are relatively often explained by rare variants with  
108 a strong effect,<sup>3</sup> and 2) late-onset cases that are influenced by multiple common variants with  
109 low effect sizes.<sup>4</sup> Segregation analyses have linked several genes to the first subgroup, including  
110 *APP*<sup>5</sup>, *PSEN1*<sup>6</sup> and *PSEN2*<sup>7</sup>. The identification of these genes has resulted in valuable insights  
111 into a molecular mechanism with an important role in AD pathogenesis, the amyloidogenic  
112 pathway,<sup>8</sup> providing a prominent example of how gene discovery can add to biological  
113 understanding of disease aetiology.

114 Besides the identification of a few rare genetic factors (e.g. *TREM2*<sup>9</sup> and *ABCA7*<sup>10</sup>),  
115 genome-wide association studies (GWAS) have mostly discovered common risk variants for the  
116 more complex late-onset type of AD. *APOE* is the strongest genetic risk locus for late-onset AD,  
117 where heterozygous and homozygous Apoe ε4 carriers are predisposed for a 3-fold and 15-fold  
118 increase in risk, respectively.<sup>11</sup> A total of 19 additional GWAS loci have been described using a  
119 discovery sample of 17,008 AD cases and 37,154 controls, followed by replication of the  
120 implicated loci with 8,572 AD patients and 11,312 controls.<sup>4</sup> The currently more than 20  
121 confirmed AD risk loci explain only a fraction of the heritability of AD and increasing the sample  
122 size is likely to boost the power for detection of more common risk variants, which will aid in  
123 understanding biological mechanisms involved in the risk for AD.

124 In the current study, we included 455,258 individuals of European ancestry, meta-  
125 analysed in 3 stages (**Figure 1**). These consisted of 24,087 clinically diagnosed late-onset AD  
126 cases, paired with 55,058 controls (phase 1). In phase 2, we analysed an AD-by-proxy  
127 phenotype, based on individuals in the UK Biobank (UKB) for whom parental AD status was  
128 available (N proxy cases=74,793; N proxy controls=328,320; **Online Methods**). The value of the

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

129 usage of by-proxy phenotypes for GWAS was recently demonstrated by Liu et al<sup>12</sup> for 12  
130 common diseases. In particular for AD, Liu et al<sup>12</sup> report substantial gains in statistical power by  
131 using a proxy phenotype, based on simulations and confirmed using empirical data from the 1<sup>st</sup>  
132 release of the UKBiobank. We here apply the proxy phenotype strategy for AD in the UKBv2  
133 sample. In this sample, parental diagnosis for AD was available for N=376,113 individuals, of  
134 whom 393 individuals had a known diagnosis of AD themselves (identified from medical  
135 register data). The high heritability of AD implies that case status for offspring can to some  
136 extent be inferred from parental case status and that offspring of AD parents are likely enriched  
137 for a higher genetic AD risk load. We thus defined individuals with one or two parents with AD  
138 as proxy cases (N=47,793), while putting more weight on the proxy cases with 2 parents.  
139 Similarly, the proxy controls include subjects with 2 parents without AD (N=328,320), where  
140 older cognitively normal parents were given more weight as proxy controls to account for the  
141 higher likelihood that younger parents may still develop AD. As the proxy phenotype is not a  
142 pure measure of an individual's AD status and may include individuals that never develop AD,  
143 genetic effect sizes will be somewhat underestimated. However, the proxy case-control sample  
144 is very large (N proxy cases=47,793; N proxy controls=328,320), and therefore increases power  
145 to detect genetic effects for AD substantially.<sup>12</sup> We first analysed the clinically defined case-  
146 control samples separately from the by-proxy case control sample to allow investigation of  
147 overlap in genetic signals for these two measurements of AD risk. Finally in phase 3, we meta-  
148 analysed all individuals of phase 1 and phase 2 together, and tested for replication in an  
149 independent sample.

150

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

151 *Genome-wide meta-analysis for AD status*

152 Phase 1 involved a genome-wide meta-analysis for AD case-control status using cohorts

153 collected as part of 3 independent main consortia (PGC-ALZ, IGAP and ADSP), totalling 79,145

154 individuals of European ancestry and 9,862,738 genetic variants passing quality control (**Figure**

155 **1, Supplementary Table 1**). The ADSP cohort obtained whole exome sequencing data from

156 4,343 cases and 3,163 controls, while the remaining datasets consisted of genotype single

157 nucleotide polymorphism (SNP) array data. AD patients were diagnosed according to generally

158 acknowledged diagnostic criteria, such as the NINCDS-ADRDA (See **Methods**). All cohorts for

159 which we had access to the raw genotypic data were subjected to a standardized quality

160 control pipeline, and GWA analyses were run per cohort and then included in a meta-analysis,

161 alongside one dataset (IGAP) for which only summary statistics were available (see **Methods**).

162 The full sample liability SNP-heritability ( $h^2_{SNP}$ ), estimated with the more conservative LD Score

163 regression (LDSC) method, was 0.055 (SE=0.0099), implying that 5.5% of AD heritability can be

164 explained by the tested SNPs. This is in line with previous estimates for IGAP (6.8%) also

165 estimated by LDSC regression method, which is based on summary statistics.<sup>13,14</sup> We do note

166 that previously reported estimates using a method based on raw genotypes (Genome-wide

167 Complex Trait Analysis, GCTA), estimated that up to 53% of total phenotypic variance in AD

168 could be explained by common SNPs, of which up to 6% could be explained by *APOE* alone, up

169 to 13% by the then known variants, and up to 25% by undiscovered loci.<sup>15,16</sup> The conservative

170 LDSC estimate of  $h^2_{SNP}$  is presumably a consequence of the underlying LDSC algorithm which is

171 based on common HapMap SNPs and excludes all variants with extreme associations.

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

172        The  $\lambda_{GC}=1.10$  indicated the presence of inflated genetic signal compared to the null  
173        hypothesis of no association. The linkage disequilibrium (LD) score intercept<sup>14</sup> was 1.044  
174        (SE=0.0084) indicating that most inflation could be explained by polygenic signal  
175        (**Supplementary Figure 1**). In the meta-analysis of AD case-control status, 1,067 variants  
176        indexed by 51 lead SNPs in approximate linkage equilibrium ( $r^2<0.1$ ) reached genome-wide  
177        significance (GWS;  $P<5\times10^{-8}$ ) (**Supplementary Figure 1**; **Supplementary Table 2**). These were  
178        located in 18 distinct genomic loci (**Table 1**). 15 of these loci confirmed previous findings  
179        (Lambert et al<sup>4</sup>) in a sample partially overlapping with that of the current study. The 3  
180        remaining loci (lead SNPs\* rs7657553, rs11257242 and rs2632516) have been linked more  
181        recently to AD in a genetic study<sup>17</sup> of AD-related cholesterol levels while conditioning on lipid  
182        levels and in a transethnic genome-wide association study of AD.<sup>18</sup>

183        We next (phase 2) performed a GWAS for AD-by-proxy using 376,113 individuals of  
184        European ancestry from the UKB version 2 release using parental AD status weighted by age  
185        and corrected for population frequency to construct an AD-by-proxy status (**Figure 1**; see  
186        **Methods**). The LD score intercept was 1.022 (SE=0.0099) indicating that most of the inflation in  
187        genetic signal ( $\lambda_{GC}=1.071$ ) could be explained by polygenic signal (**Supplementary Figure 1B**).  
188        For AD-by-proxy, 719 GWS variants were indexed by 61 lead SNPs in approximate linkage  
189        equilibrium ( $r^2<0.1$ ) reached genome-wide significance ( $P<5\times10^{-8}$ ), located in 13 loci  
190        (**Supplementary Figure 1A**). Of these, 8 loci overlapped with the significantly associated loci  
191        identified for clinical AD case control status (**Table 1**).

\* Although in other AD-related manuscripts this is common, we choose not to report the gene that is in closest proximity to the lead SNP as the ID for the locus, as this incorrectly implies that the gene is the causal gene for AD pathogenesis. We therefore believe it is preferred to use the rs-number of the most strongly associated SNP as an ID for the locus, and aim to highlight the most likely causal genes with more sophisticated functional interpretation analyses in later sections of this study.

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

192 We observed a strong genetic correlation of 0.81 (SE= 0.185, using LDSC) between AD  
193 status and AD-by-proxy, indicating substantial overlap between genetic effects beyond shared  
194 GWS SNPs. Sign concordance tests indicated that 50.4% of all LD-independent ( $r^2 < 0.1$ ) genome-  
195 wide SNPs (significant and non-significant) had consistent direction of effects between the two  
196 phenotypes (N=344,581 overlapping SNPs), slightly greater than the chance expectation of 50%  
197 (exact binomial test  $P=2.45 \times 10^{-7}$ ). Of the 51 lead SNPs identified by the case-control meta-  
198 analysis, all were available in UKB and 96.1% were sign-concordant ( $P=2.98 \times 10^{-12}$ ), while of the  
199 61 GWS lead SNPs identified in UKB, 48 were available in the case-control meta-analysis and  
200 99.7% of these were sign-concordant ( $P=5.98 \times 10^{-14}$ ). Such substantial overlap suggests that the  
201 AD-by-proxy phenotype captures a large part of the associated genetic effects on AD.

202 Given the high genetic overlap, in phase 3, we conducted a meta-analysis on the clinical  
203 AD case-control GWAS and the AD-by-proxy GWAS (**Figure 1**), comprising a total sample size of  
204 455,258 (71,880 (proxy) cases and 383,378 (proxy) controls). The LD score intercept was 1.0018  
205 (SE=0.0109) indicating again that most of the inflation in genetic signal ( $\lambda_{GC}=1.0833$ ) could be  
206 explained by polygenic signal (**Supplementary Figure 1b**). There were 2,357 GWS variants,  
207 which were represented by 94 lead SNPs, located in 29 loci (**Table 1, Figure 2**). These included  
208 15 of the 18 loci detected in our case-control analyses, all of the 13 detected in the AD-by-proxy  
209 analyses, as well as 9 loci that were sub-threshold in both individual analyses but reached  
210 significance in the meta-analysis. All 2,160 GWS SNPs that were available in both the case-  
211 control and AD-by-proxy sub-samples were sign concordant (exact binomial test  $P < 1 \times 10^{-300}$ ),  
212 including all of the 82 available independent lead SNPs ( $P=1.68 \times 10^{-23}$ ). Association was found  
213 with both AD and AD-by-proxy for 22 (out of 27 overlapping) loci for which SNP(s) in each locus

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

had a robust  $P$ -value ( $P < 0.05/94$  independent signals). Of the 29 associated loci, 16 were previously identified by the GWAS of Lambert et al.,<sup>4</sup> and 13 were not. Three of these (with lead SNPs rs184384746, rs187370608 and rs114360492) were only available in the UKB cohort (**Table 1**). Verifying our results against other<sup>9,19</sup> and more recent<sup>12,17,20</sup> genetic studies on AD, 4 loci (rs187370608, rs11257238, rs113260531 and rs28394864) were previously discovered, leaving 9 novel loci (rs4575098, rs184384746, rs6448453, rs114360492, rs442495, rs117618017, rs59735493, rs76726049 and rs76320948). Considering all loci of Lambert et al,<sup>4</sup> we were unable to replicate 4 loci (*MEF2C*, *NME8*, *CELF1* and *FERMT2\**) at a GWS level (observed  $P$ -values were  $1.6 \times 10^{-5}$  to 0.0011), which was mostly caused by a lower association signal in the UKB dataset (**Supplementary Table 3**). By contrast, Lambert et al<sup>4</sup> were unable to replicate the *DSG2* and *CD33* loci in the second stage of their study. In our study, *DSG2* is also not supported (meta-analysis  $P=0.030$ ; UKB analysis  $P=0.766$ ; **Table 1**), implying invalidation of this locus, while the *CD33* locus (rs3865444 in **Table 1**) is significantly associated with AD (meta-analysis  $P=6.34 \times 10^{-9}$ ; UKB analysis  $P=4.97 \times 10^{-5}$ ), implying a genuine genetic association to AD risk.

Next, we aimed to find further support for the novel findings of the phase 3 meta-analysis, by using an independent Icelandic cohort (deCODE<sup>21,22</sup>), including 6,593 AD cases and 174,289 controls (**Figure 1; see Methods; Supplementary Table 4**). We were unable to test two loci as the lead SNPs (and SNPs in high LD), either were not present in the 28,075 genomes of the Icelandic reference panel or were not imputed with sufficient quality. For 6 of the 7 novel loci tested for replication, we observed the same direction of effect in the deCODE cohort. Furthermore, 4 loci (rs6448453, rs442495, rs117618017, rs76320948) showed nominally

\* For straightforward comparison to this GWAS, we do here report the genes in closest proximity to the lead SNP. However, we would like to point out that GWAS findings implicate a genomic locus, and that the closest gene is not necessarily the causal gene.

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

236 significant association results ( $P<0.05$ ) for the same SNP or a SNP in high LD ( $r^2 > 0.9$ ) within the  
237 same locus (two-tailed binomial test  $P=1.9 \times 10^{-4}$ ). The locus on chromosome 1 (rs45759098) was  
238 very close to significance ( $P=0.053$ ). Apart from the novel loci, we also observed sign  
239 concordance for 95.6% of the lead SNPs in all loci from the meta-analysis ( $P=1.60 \times 10^{-20}$ ) that  
240 were available in deCODE (out of 94). As an additional method of testing for replication using  
241 genome-wide polygenic score prediction,<sup>23</sup> the current results explain 7.1% of the variance in  
242 clinical AD at a low best fitting  $P$ -threshold of  $1.69 \times 10^{-5}$  ( $P=1.80 \times 10^{-10}$ ) in an independent sample  
243 of 761 individuals (see **Methods**). When excluding the *APOE*-locus (chr19: 45020859-  
244 45844508), the results explain 3.9% of the variance with a best fitting  $P$ -threshold of  $3.5 \times 10^{-5}$   
245 ( $P=1.90 \times 10^{-6}$ ).

246

247 *Functional interpretation of genetic variants contributing to AD and AD-by-proxy*

248 Next, we conducted a number of *in silico* follow-up analyses to interpret our findings in a  
249 biological context. Functional annotation of all GWS SNPs ( $n=2,178$ ) in the associated loci  
250 showed that SNPs were mostly located in intronic/intergenic areas, yet in regions that were  
251 enriched for chromatin states 4 and 5, implying effects on active transcription (**Figure 3A, 3B**  
252 and **3C; Supplementary Table 5**). 24 GWS SNPs were exonic non-synonymous (ExNS) (**Figure**  
253 **3A; Supplementary Table 6**) with likely deleterious implications on gene function. Converging  
254 evidence of strong association ( $Z > |7|$ ) and a high observed probability of a deleterious variant  
255 effect (CADD<sup>24</sup> score  $\geq 30$ ) was found for rs75932628 (*TREM2*), rs142412517 (*TOMM40*) and  
256 rs7412 (*APOE*). The first two missense mutations are rare (MAF=0.002 and 0.001, respectively)  
257 and the alternative alleles were associated with higher risk for AD. The latter *APOE* missense

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

258 mutation is the well-established protective allele Apoε2. The effect sizes for ExNS ranged from  
259 moderate to high. **Supplementary Tables 5 and 6** present a detailed annotation catalogue of  
260 variants in the associated genomic loci. Partitioned analysis,<sup>25</sup> excluding SNPs with extremely  
261 large effect sizes (i.e. *APOE* variants) showed enrichment for  $h^2_{SNP}$  for variants located in  
262 H3K27ac marks (Enrichment=3.18,  $P=9.63\times10^{-5}$ ), which are associated with activation of  
263 transcription, and in Super Enhancers (Enrichment=3.62,  $P=2.28\times10^{-4}$ ), which are genomic  
264 regions where multiple epigenetic marks of active transcription are clustered (**Figure 3D**;  
265 **Supplementary Table 7**). Heritability was also enriched in variants on chromosome 17  
266 (Enrichment=3.61,  $P=1.63\times10^{-4}$ ) and we observed a trend of enrichment for variants with high  
267 minor allele frequencies (Enrichment=3.31,  $P=2.85\times10^{-3}$ ), (**Supplementary Figure 3**;  
268 **Supplementary Tables 8 and 9**). Although a large proportion (23.9%) of the heritability can be  
269 explained by SNPs on chromosome 19, this enrichment is not significant, due to the large  
270 standard errors around this estimate (**Supplementary Table 8**). Overall these results suggest  
271 that, despite some nonsynonymous variants likely contributing to AD risk, most of the GWS  
272 SNPs are located in non-coding regions, and are enriched for regions that have an activating  
273 effect on transcription.

274

### 275 *Implicated genes*

276 To link the associated variants to genes, we applied three gene-mapping strategies  
277 implemented in FUMA<sup>26</sup> (**Online Methods**). We used all SNPs with a P-value  $< 5\times10^{-8}$  and  $r^2$  of  
278 0.6 with one of the independently associated SNPs, for gene-mapping. *Positional* gene-mapping  
279 aligned SNPs to 100 genes by their location within or immediately up/downstream (+/-10kb) of

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

280 known gene boundaries, *eQTL* (*expression quantitative trait loci*) gene-mapping matched cis-  
281 eQTL SNPs to 170 genes whose expression levels they influence in one or more tissues, and  
282 *chromatin interaction* mapping linked SNPs to 21 genes based on three-dimensional DNA-DNA  
283 interactions between each SNP's genomic region and nearby or distant genes, which we limited  
284 to include only interactions between annotated enhancer and promotor regions (**Figure 3B and**  
285 **3C; Supplementary Figure 4; Supplementary Tables 10 and 11**). This resulted in 192 uniquely  
286 mapped genes, 80 of which were implicated by at least two mapping strategies and 17 by all 3  
287 (**Figure 4E**). Eight genes (*HLA-DRB5*, *HLA-DRB1*, *HLA-DQA*, *HLA-DQB1*, *KAT8*, *PRSS36*, *ZNF232*  
288 and *CEACAM19*) are particularly notable as they are implicated via eQTL association in the  
289 hippocampus, a brain region highly affected early in AD pathogenesis (**Supplementary Table**  
290 **10**). Of special interest is the locus on chromosome 8 (rs4236673). In the GWAS by Lambert et  
291 al.<sup>4</sup>, this locus was defined as 2 distinct loci (*CLU* and *PTK2B*), while our meta-analysis specified  
292 this locus as a single locus based on LD-patterns. This is also supported by a chromatin  
293 interaction between the two regions (**Figure 3E**), which is observed in two immune-related  
294 tissues – the spleen and liver (**Supplementary Table 11**). Chromosome 16 contains a locus  
295 implicated by long-range eQTL association (**Figure 3F**) clearly illustrating more distant genes can  
296 be affected by a genetic factor (**Figure 3F**) and emphasising the relevance of considering  
297 putative causal genes or regulatory elements not solely on the physical location but also on  
298 epigenetic influences. **Supplementary Figure 4** displays chromatin interactions for all  
299 chromosomes containing significant GWAS loci.

300           Although these gene-mapping strategies imply multiple putative causal genes per GWAS  
301 locus, several of these genes in the novel loci (and significantly replicated by the deCODE

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

302 cohort) are of particular interest, as the genes have functional or previous genetic association  
303 to AD. For locus 1 in **Supplementary Table 10**, *ADAMTS4* encodes a protein of the ADAMTS  
304 family which has a function in neuroplasticity and has been extensively studied for their role in  
305 AD pathogenesis.<sup>27</sup> For locus 19, the obvious most likely causal gene is *ADAM10*, as this gene  
306 has been associated with AD by research focusing on rare coding variants in *ADAM10*.<sup>28</sup>  
307 However this is the first time that this gene is implicated as a common risk factor for AD. The  
308 lead SNP for locus 20 is a nonsynonymous variant in exon 1 of *APH1B*, which encodes for a  
309 protein subunit of the  $\gamma$ -secretase complex cleaving *APP*.<sup>29</sup> Although previously reported  
310 functional information on genes can be of great value, it is preferable to consider all implicated  
311 genes as putative causal factors to guide potential functional follow-up experiments.

312 We next performed genome-wide gene-based association analysis (GWGAS) using  
313 MAGMA.<sup>30</sup> This method annotates SNPs to known protein-coding genes to estimate aggregate  
314 associations based on all SNPs in a gene. It differs from the gene-mapping strategies in FUMA as  
315 it provides a statistical gene-based test, whereas FUMA maps individually significant SNPs to  
316 genes. With GWGAS, we identified 97 genes that were significantly associated to AD  
317 (**Supplementary Figure 5; Supplementary Table 12**), of which 74 were also mapped by FUMA  
318 (**Figure 4E**). In total, 16 genes were implicated by all four strategies (**Supplementary Table 13**),  
319 of which 7 genes (*HLA-DRA*, *HLA-DRB1*, *PTK2B*, *CLU*, *MS4A3*, *SCIMP* and *RABEP1*) are not  
320 located in the *APOE*-locus, and therefore of high interest for further investigation.

321

322 *Gene-sets implicated in AD and AD-by-proxy*

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

323 Using the gene-based P-values, we performed gene-set analysis for 6,994 biological-pathway-  
324 based gene-sets, 53 tissue expression-based gene-sets and 39 brain single-cell expression based  
325 gene-sets (24 derived from mouse data and 15 derived from human data). We found four Gene  
326 Ontology<sup>19</sup> gene-sets that were significantly associated with AD risk: *Protein lipid complex*  
327 ( $P=3.93\times10^{-10}$ ), *Regulation of amyloid precursor protein catabolic process* ( $P=8.16\times10^{-9}$ ), *High*  
328 *density lipoprotein particle* ( $P=7.81\times10^{-8}$ ), and *Protein lipid complex assembly* ( $P=7.96\times10^{-7}$ )  
329 (**Figure 4A; Supplementary Tables 14 and 15**). Conditional analysis on the *APOE* locus showed  
330 associations with AD for these four gene-sets independent of the effect of *APOE*, as they  
331 remained significantly associated ( $P<0.0125$ ), yet less strongly, suggesting that *APOE* is  
332 contributing a substantial part to the association signal, but does not completely drive the  
333 signal. There was overlap between genes included in the four gene-sets, and conditioning on  
334 each significant gene-set association showed that three gene-sets were associated with AD  
335 independently of each other (**Supplementary Tables 14 and 15**). All 25 genes of the *High*  
336 *density lipoprotein particle* pathway are also part of the *Protein lipid complex* (conditional  
337 analysis  $P=0.18$ ), and these pathways are therefore not interpretable as independent  
338 associations.

339 Linking gene-based P-values to tissue- and cell-type-specific gene-sets, no association  
340 survived the stringent Bonferroni correction, which corrected for all tested gene-sets (i.e. 6,994  
341 GO categories, 54 tissues and 39 cell types). However, we did observe associations when  
342 correcting only for the number of tests within all tissue types or cell-types. This was the case for  
343 gene expression across immune-related tissues (**Figure 4C; Supplementary Table 16**),  
344 particularly whole blood ( $P=5.61\times10^{-6}$ ), spleen ( $P=1.50\times10^{-5}$ ) and lung ( $P=4.67\times10^{-4}$ ). In brain

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

345 single-cell expression gene-set analyses, we found associations for microglia, both in the  
346 mouse-based expression dataset ( $P=1.96\times10^{-3}$ ) (**Figure 4B; Supplementary Table 17**) and the  
347 human-based expression dataset ( $P=2.56\times10^{-3}$ ) (**Supplementary Figure 6; Supplementary Table**  
348 **18**).

349

350 *Cross-trait genetic influences*

351 For a more comprehensive understanding of the genetic background of AD, we next tested  
352 whether AD is likely to share genetic factors with other phenotypes. This might reveal some  
353 functional insights about the genetic aetiology of AD. We conducted bivariate LD score<sup>14</sup>  
354 regression to test for genetic correlations between AD and 41 other traits for which large GWAS  
355 summary statistics were available. We observed significant negative genetic correlations with  
356 adult cognitive ability ( $r_g=-0.22$ ,  $P=7.28\times10^{-5}$ ), age of first birth ( $r_g=-0.33$ ,  $P=1.22\times10^{-4}$ ),  
357 educational attainment ( $r_g=-0.25$ ,  $P=5.01\times10^{-4}$ ), and confirmed a very strong positive  
358 correlation with previous GWAS of Alzheimer's disease ( $r_g=0.90$ ,  $P=3.29\times10^{-16}$ ) (**Figure 4D;**  
359 **Supplementary Table 19**).

360 We then used Generalised Summary-statistic-based Mendelian Randomisation<sup>31</sup> (GSMR;  
361 see **Methods**) to test for potential credible causal associations of genetically correlated  
362 outcomes which may directly influence the risk for AD. Due to the nature of AD being a late-  
363 onset disorder and summary statistics for most other traits being obtained from younger  
364 samples, we do not report tests for the opposite direction of potential causality (i.e. we did not  
365 test for a causal effect of a late-onset disease on an early onset disease). In this set of analyses,  
366 SNPs from the summary statistic of genetically correlated phenotypes were used as

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

367 instrumental variables to estimate the putative causal effect of these “exposure” phenotypes  
368 on AD risk by comparing the ratio of SNPs’ associations with each exposure to their associations  
369 with AD outcome (see **Methods**). Association statistics were standardized, such that the  
370 reported effects reflect the expected difference in odds ratio (OR) for AD as a function of every  
371 SD increase in the exposure phenotype. We observed a protective effect of cognitive ability  
372 (OR=0.89, 95% confidence interval[CI]: 0.85-0.92,  $P=5.07\times10^{-9}$ ), educational attainment  
373 (OR=0.88, 95%CI: 0.81-0.94,  $P=3.94\times10^{-4}$ ), and height (OR=0.96, 95%CI: 0.94-0.97,  $P=1.84\times10^{-8}$ )  
374 on risk for AD (**Supplementary Table 20**; **Supplementary Figure 7**). No substantial evidence of  
375 pleiotropy was observed between AD and these phenotypes, with <1% of overlapping SNPs  
376 being filtered as outliers (**Supplementary Figure 7**).  
377

### 378 *Discussion*

379 By using a non-conventional approach of including a by-proxy phenotype for AD to increase  
380 sample size, we have identified 9 novel loci and gained novel biological knowledge on AD  
381 aetiology. Both the high genetic correlation between the standard case-control status and the  
382 UKB by proxy phenotype ( $r_g=0.81$ ) and the high rate of novel loci replication in the independent  
383 deCODE cohort, suggest that this strategy is robust. Through extensive in silico functional  
384 follow-up analysis, and in line with previous research,<sup>20,32</sup> we emphasise the crucial causal role  
385 of the immune system - rather than immune response as a consequence of disease pathology -  
386 by establishing variant enrichments for immune-related body tissues (whole blood, spleen,  
387 liver) and for the main immune cells of the brain (microglia). Furthermore, we observe  
388 informative eQTL associations and chromatin interactions within immune-related tissues for

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

389 the identified genomic risk loci. Together with the AD-associated genetic effects on lipid  
390 metabolism in our study, these biological implications strengthen the hypothesis that AD  
391 pathogenesis involves an interplay between inflammation and lipids, as lipid changes might  
392 harm immune responses of microglia and astrocytes, and vascular health of the brain.<sup>33</sup>

393 In accordance with previous clinical research, our study suggests an important role for  
394 protective effects of several human traits on AD. As an example, cognitive reserve has been  
395 proposed as a protective mechanism in which the brain aims to control brain damage with prior  
396 existing cognitive processing strategies.<sup>34</sup> Our findings imply that some component of the  
397 genetic factors for AD might affect cognitive reserve, rather than being involved in AD-  
398 pathology-related damaging processes, influencing AD pathogenesis in an indirect way through  
399 cognitive reserve. Similarly, in a largescale community-based study it was observed that AD  
400 incidence rates declined over decades, which was specific for individuals with at minimum a  
401 high school diploma.<sup>35</sup> Combined with our Mendelian randomization results for educational  
402 attainment, this suggests that the protective effect of educational attainment on AD is  
403 influenced by genetics.

404 The results of this study could furthermore serve as a valuable resource (e.g.  
405 Supplementary Tables 10 and 13) for selection of promising genes for functional follow-up  
406 experiments and identify targets for drug development. We anticipate that functional  
407 interpretation strategies and follow-up experiments will result in a comprehensive  
408 understanding of late-onset AD aetiology, which will serve as a solid foundation for future AD  
409 drug development and stratification approaches.

410

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

411 URLs:

412 <http://ukbiobank.ac.uk>

413 <https://www.ncbi.nlm.nih.gov/gap>

414 <http://fuma.ctglab.nl>

415 <http://ctg.cnrc.nl/software/magma>

416 [http://genome.sph.umich.edu/wiki/METAL\\_Program](http://genome.sph.umich.edu/wiki/METAL_Program)

417 <https://github.com/bulik/ldsc>

418 <http://ldsc.broadinstitute.org/>

419 <https://data.broadinstitute.org/alkesgroup/LDSCORE/>

420 <http://www.genecards.org>

421 <http://www.med.unc.edu/pgc/results-and-downloads>

422 <http://software.broadinstitute.org/gsea/msigdb/collections.jsp>

423 <https://www.ebi.ac.uk/gwas/>

424 <https://github.com/ivankosmos/RegionAnnotator>

425 <http://cnsgenomics.com/software/gsmr/>

426

427 **Acknowledgments:** This work was funded by The Netherlands Organization for Scientific Research (NWO VICI 453-14-005) and the Sophia Foundation for Scientific Research (grant nr: S14-27). The analyses were carried out on the Genetic Cluster Computer, which is financed by the Netherlands Scientific Organization (NWO: 480-05-003), by the VU University, Amsterdam, The Netherlands, and by the Dutch Brain Foundation, and is hosted by the Dutch National Computing and Networking Services SurfSARA. The work was also funded by The Research

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

433 Council of Norway (#251134, #248778, #223273, #213837, #225989), KG Jebsen Stiftelsen, The  
434 Norwegian Health Association, European Community's JPND Program, ApGeM RCN #237250,  
435 and the European Community's grant # PIAPP-GA-2011-286213 PsychDPC. This research has  
436 been conducted using the UK Biobank resource under application number 16406 and the public  
437 ADSP dataset, obtained through the Database of Genotypes and Phenotypes (dbGaP) under  
438 accession number phs000572 (see sections below).

439 Genotyping for the Swedish Twin Studies of Aging was supported by NIH/NIA grant R01  
440 AG037985. Genotyping in TwinGene was supported by NIH/NIDDK U01 DK066134. WvdF is  
441 recipient of Joint Programming for Neurodegenerative Diseases (JPND) grants PERADES (ANR-  
442 13-JPRF-0001) and EADB (733051061). AddNeuroMed consortium was led by Simon Lovestone,  
443 Bruno Vellas, Patrizia Mecocci, Magda Tsolaki, Iwona Kłoszewska, Hilkka Soininen. This work  
444 was supported by InnoMed (Innovative Medicines in Europe), an integrated project funded by  
445 the European Union of the Sixth Framework program priority (FP6-2004- LIFESCIHEALTH-5). JB  
446 was supported by a grant from the Swiss National Science Foundation. JHL was supported by  
447 the Swedish Research Council (Vetenskapsrådet, award 2014-3863), the Wellcome Trust  
448 (108726/Z/15/Z), and the Swedish Brain Foundation (Hjärnfonden). NS was supported by the  
449 Wellcome Trust (108726/Z/15/Z). RD was supported by National Institute for Health Research  
450 University College London Hospital's Biomedical Research Centre, Arthritis Research UK, the  
451 British Heart Foundation, Cancer Research UK, the Chief Scientist Office, the Economic and  
452 Social Research Council, the Engineering and Physical Sciences Research Council, the National  
453 Institute for Social Care and Health Research, and the Wellcome Trust (grant number  
454 MR/K006584/1), Innovative Medicines Initiative Joint Undertaking under EMIF grant agreement

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

455 number 115372, resources of which are composed of financial contribution from the European  
456 Union's Seventh Framework Program (FP7/2007-2013) and EFPIA companies' in kind  
457 contribution. SJK was supported by an MRC Career Development Award in Biostatistics  
458 (MR/L011859/1).

459 We thank the International Genomics of Alzheimer's Project (IGAP) for providing  
460 summary results data for these analyses. The investigators within IGAP contributed to the  
461 design and implementation of IGAP and/or provided data but did not participate in analysis or  
462 writing of this report. IGAP was made possible by the generous participation of the control  
463 subjects, the patients, and their families. The i-Select chips was funded by the French National  
464 Foundation on Alzheimer's disease and related disorderacknows. EADI was supported by the  
465 LABEX (laboratory of excellence program investment for the future) DISTALZ grant, Inserm,  
466 Institut Pasteur de Lille, Université de Lille 2 and the Lille University Hospital. GERAD was  
467 supported by the Medical Research Council (Grant n° 503480), Alzheimer's Research UK (Grant  
468 n° 503176), the Wellcome Trust (Grant n° 082604/2/07/Z) and German Federal Ministry of  
469 Education and Research (BMBF): Competence Network Dementia (CND) grant n° 01GI0102,  
470 01GI0711, 01GI0420. CHARGE was partly supported by the NIH/NIA grant R01 AG033193 and  
471 the NIA AG081220 and AGES contract N01-AG-12100, the NHLBI grant R01 HL105756, the  
472 Icelandic Heart Association, and the Erasmus Medical Center and Erasmus University. ADGC  
473 was supported by the NIH/NIA grants: U01 AG032984, U24 AG021886, U01 AG016976, and the  
474 Alzheimer's Association grant ADGC-10-196728. This paper represents independent research  
475 funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at  
476 South London and Maudsley NHS Foundation Trust and King's College London. The views

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

477 expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the  
478 Department of Health.

479 The Alzheimer's Disease Sequencing Project (ADSP) is comprised of two Alzheimer's  
480 Disease (AD) genetics consortia and three National Human Genome Research Institute (NHGRI)  
481 funded Large Scale Sequencing and Analysis Centers (LSAC). The two AD genetics consortia are  
482 the Alzheimer's Disease Genetics Consortium (ADGC) funded by NIA (U01 AG032984), and the  
483 Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) funded by NIA (R01  
484 AG033193), the National Heart, Lung, and Blood Institute (NHLBI), other National Institute of  
485 Health (NIH) institutes and other foreign governmental and non-governmental organizations.

486 The Discovery Phase analysis of sequence data is supported through UF1AG047133 (to Drs.  
487 Schellenberg, Farrer, Pericak-Vance, Mayeux, and Haines); U01AG049505 to Dr. Seshadri;  
488 U01AG049506 to Dr. Boerwinkle; U01AG049507 to Dr. Wijsman; and U01AG049508 to Dr.  
489 Goate and the Discovery Extension Phase analysis is supported through U01AG052411 to Dr.  
490 Goate, U01AG052410 to Dr. Pericak-Vance and U01 AG052409 to Drs. Seshadri and Fornage.

491 Data generation and harmonization in the Follow-up Phases is supported by U54AG052427 (to  
492 Drs. Schellenberg and Wang). The ADGC cohorts include: Adult Changes in Thought (ACT), the  
493 Alzheimer's Disease Centers (ADC), the Chicago Health and Aging Project (CHAP), the Memory  
494 and Aging Project (MAP), Mayo Clinic (MAYO), Mayo Parkinson's Disease controls, University of  
495 Miami, the Multi-Institutional Research in Alzheimer's Genetic Epidemiology Study (MIRAGE),  
496 the National Cell Repository for Alzheimer's Disease (NCRAD), the National Institute on Aging  
497 Late Onset Alzheimer's Disease Family Study (NIA-LOAD), the Religious Orders Study (ROS), the  
498 Texas Alzheimer's Research and Care Consortium (TARC), Vanderbilt University/Case Western

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

499 Reserve University (VAN/CWRU), the Washington Heights-Inwood Columbia Aging Project  
500 (WHICAP) and the Washington University Sequencing Project (WUSP), the Columbia University  
501 Hispanic- Estudio Familiar de Influencia Genetica de Alzheimer (EFIGA), the University of  
502 Toronto (UT), and Genetic Differences (GD). The CHARGE cohorts are supported in part by  
503 National Heart, Lung, and Blood Institute (NHLBI) infrastructure grant HL105756 (Psaty),  
504 RC2HL102419 (Boerwinkle) and the neurology working group is supported by the National  
505 Institute on Aging (NIA) R01 grant AG033193. The CHARGE cohorts participating in the ADSP  
506 include the following: Austrian Stroke Prevention Study (ASPS), ASPS-Family study, and the  
507 Prospective Dementia Registry-Austria (ASPS/PRODEM-Aus), the Atherosclerosis Risk in  
508 Communities (ARIC) Study, the Cardiovascular Health Study (CHS), the Erasmus Rucphen Family  
509 Study (ERF), the Framingham Heart Study (FHS), and the Rotterdam Study (RS). ASPS is funded  
510 by the Austrian Science Fond (FWF) grant number P20545-P05 and P13180 and the Medical  
511 University of Graz. The ASPS-Fam is funded by the Austrian Science Fund (FWF) project  
512 I904),the EU Joint Programme - Neurodegenerative Disease Research (JPND) in frame of the  
513 BRIDGET project (Austria, Ministry of Science) and the Medical University of Graz and the  
514 Steiermärkische Krankenanstalten Gesellschaft. PRODEM-Austria is supported by the Austrian  
515 Research Promotion agency (FFG) (Project No. 827462) and by the Austrian National Bank  
516 (Anniversary Fund, project 15435. ARIC research is carried out as a collaborative study  
517 supported by NHLBI contracts (HHSN268201100005C, HHSN268201100006C,  
518 HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C,  
519 HHSN268201100011C, and HHSN268201100012C). Neurocognitive data in ARIC is collected by  
520 U01 2U01HL096812, 2U01HL096814, 2U01HL096899, 2U01HL096902, 2U01HL096917 from the

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

521 NIH (NHLBI, NINDS, NIA and NIDCD), and with previous brain MRI examinations funded by R01-  
522 HL70825 from the NHLBI. CHS research was supported by contracts HHSN268201200036C,  
523 HHSN268200800007C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082,  
524 N01HC85083, N01HC85086, and grants U01HL080295 and U01HL130114 from the NHLBI with  
525 additional contribution from the National Institute of Neurological Disorders and Stroke  
526 (NINDS). Additional support was provided by R01AG023629, R01AG15928, and R01AG20098  
527 from the NIA. FHS research is supported by NHLBI contracts N01-HC-25195 and  
528 HHSN268201500001I. This study was also supported by additional grants from the NIA (R01s  
529 AG054076, AG049607 and AG033040 and NINDS (R01 NS017950). The ERF study as a part of  
530 EUROSPAN (European Special Populations Research Network) was supported by European  
531 Commission FP6 STRP grant number 018947 (LSHG-CT-2006-01947) and also received funding  
532 from the European Community's Seventh Framework Programme (FP7/2007-2013)/grant  
533 agreement HEALTH-F4-2007-201413 by the European Commission under the programme  
534 "Quality of Life and Management of the Living Resources" of 5th Framework Programme (no.  
535 QLG2-CT-2002-01254). High-throughput analysis of the ERF data was supported by a joint grant  
536 from the Netherlands Organization for Scientific Research and the Russian Foundation for Basic  
537 Research (NWO-RFBR 047.017.043). The Rotterdam Study is funded by Erasmus Medical Center  
538 and Erasmus University, Rotterdam, the Netherlands Organization for Health Research and  
539 Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of  
540 Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European  
541 Commission (DG XII), and the municipality of Rotterdam. Genetic data sets are also supported  
542 by the Netherlands Organization of Scientific Research NWO Investments (175.010.2005.011,

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

543 911-03-012), the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, the  
544 Research Institute for Diseases in the Elderly (014-93-015; RIDE2), and the Netherlands  
545 Genomics Initiative (NGI)/Netherlands Organization for Scientific Research (NWO) Netherlands  
546 Consortium for Healthy Aging (NCHA), project 050-060-810. All studies are grateful to their  
547 participants, faculty and staff. The content of these manuscripts is solely the responsibility of  
548 the authors and does not necessarily represent the official views of the National Institutes of  
549 Health or the U.S. Department of Health and Human Services. The three LSACs are: the Human  
550 Genome Sequencing Center at the Baylor College of Medicine (U54 HG003273), the Broad  
551 Institute Genome Center (U54HG003067), and the Washington University Genome Institute  
552 (U54HG003079). Biological samples and associated phenotypic data used in primary data  
553 analyses were stored at Study Investigators institutions, and at the National Cell Repository for  
554 Alzheimer's Disease (NCRAD, U24AG021886) at Indiana University funded by NIA. Associated  
555 Phenotypic Data used in primary and secondary data analyses were provided by Study  
556 Investigators, the NIA funded Alzheimer's Disease Centers (ADCs), and the National Alzheimer's  
557 Coordinating Center (NACC, U01AG016976) and the National Institute on Aging Genetics of  
558 Alzheimer's Disease Data Storage Site (NIAGADS, U24AG041689) at the University of  
559 Pennsylvania, funded by NIA, and at the Database for Genotypes and Phenotypes (dbGaP)  
560 funded by NIH. This research was supported in part by the Intramural Research Program of the  
561 National Institutes of health, National Library of Medicine. Contributors to the Genetic Analysis  
562 Data included Study Investigators on projects that were individually funded by NIA, and other  
563 NIH institutes, and by private U.S. organizations, or foreign governmental or nongovernmental  
564 organizations.

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

565 We thank the numerous participants, researchers, and staff from many studies who  
566 collected and contributed to the data. Summary statistics will be made available for download  
567 upon publication from <http://ctglab.vu.nl>.

568

569 **Author Contributions:** I.E.J. and J.E.S. performed the analyses. D.P. and O.E.A. conceived the  
570 idea of the study. D.P. and S.R. supervised analyses. Sv.St. performed QC on the UK Biobank  
571 data and wrote the analysis pipeline. K.W. constructed and applied the FUMA pipeline for  
572 performing follow-up analyses. J.B. conducted the single cell enrichment analyses. J.H.L and  
573 N.S. contributed data. D.P. and I.E.J. wrote the first draft of the paper. All other authors  
574 contributed data and critically reviewed the paper.

575

576 **Author Information:** PF Sullivan reports the following potentially competing financial interests:  
577 Lundbeck (advisory committee), Pfizer (Scientific Advisory Board member), and Roche (grant  
578 recipient, speaker reimbursement). JHL: Cartana (Scientific Advisor) and Roche (grant recipient).  
579 Ole A Andreassen: (Lundbeck) speaker's honorarium. Stacy Steinberg, Hreinn Stefansson and  
580 Kari Stefansson are employees of deCODE Genetics/Amgen. All other authors declare no  
581 financial interests or potential conflicts of interest.

582 Correspondence and requests for materials should be addressed to [d.posthuma@vu.nl](mailto:d.posthuma@vu.nl).

583  
584  
585  
586  
587  
588  
589  
590

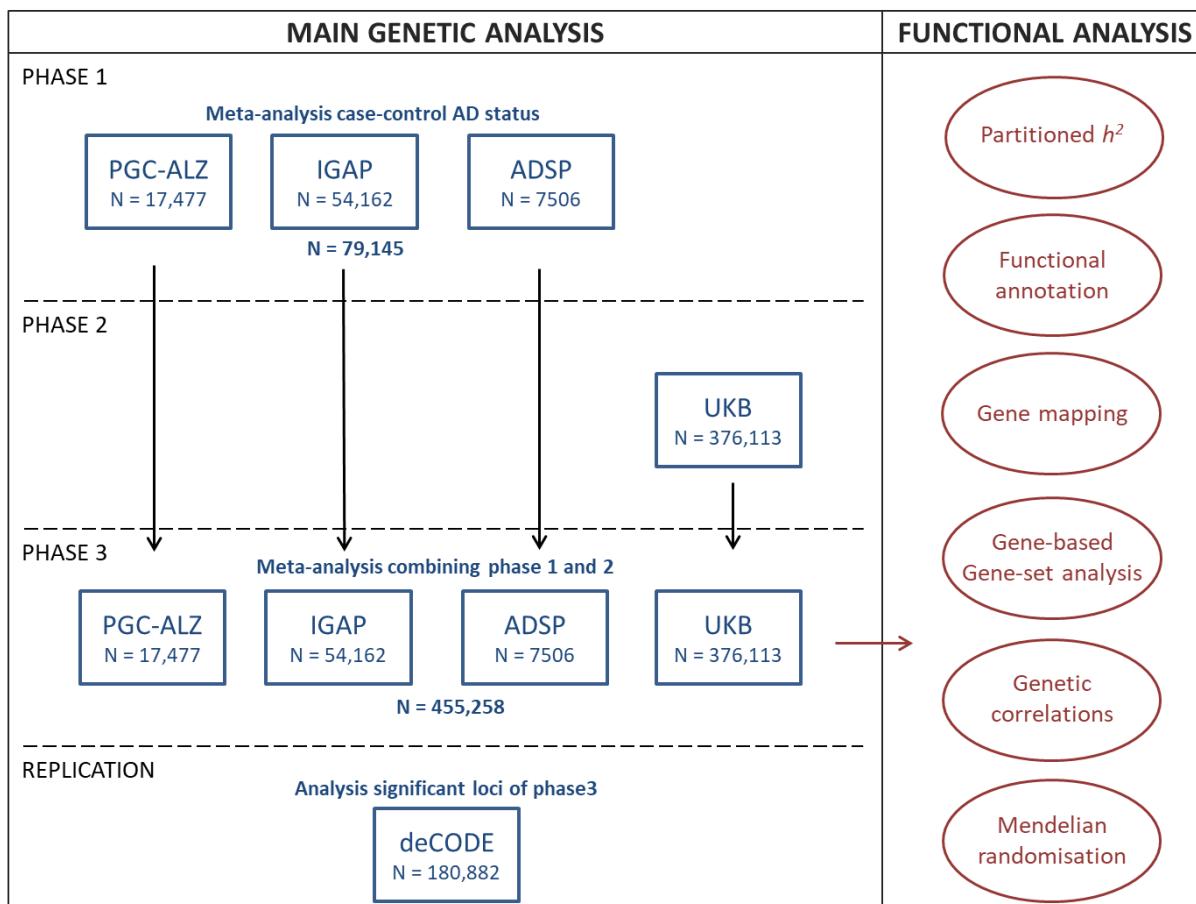
## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

**Table 1. Summary statistics for the meta-analysis of case-control status, by proxy phenotype and both. Independent lead SNPs are defined by  $r^2 < .1$ ; distinct genomic loci are  $>250\text{kb}$  apart. Alleles = the effect and non-effect allele. OR = odds ratio, only displayed for dichotomous phenotype. beta = effect size of the first allele displayed in the corresponding alleles column. Meta-analysis effect direction (column V) is in the following order: ADSP, IGAP, UKB, PGC-ADZ, note that the first cohort is often missing as this concerns exome sequencing data. Corrected P value for significance = 5E-08 (marked as boldfaced and underlined values). Note that the lead SNP can differ between the distinct analyses, while it tags the same locus.**

Locus	Chr	SNP	Case-control status			By proxy AD			Overall			<i>p</i>	direction						
			bp	alleles	OR	beta	<i>p</i>	SNP	bp	alleles	beta			SNP	bp	alleles	MAF	HRC	beta
1	1	rs4575098	161155392	A/G	1.024	0.024	0.000157	rs4575098	161155392	A/G	0.009	6.88E-08	rs4575098	161155392	A/G	0.228	0.016	<b>2.05E-10</b>	?+++
2	1	rs6556401	207692049	A/G	1.062	0.060	<b>1.39E-17</b>	rs679515	207750568	T/C	0.010	<b>8.85E-10</b>	rs2093760	207786828	A/G	0.192	0.024	<b>1.10E-18</b>	+++
3	2	rs463105	127891427	C/A	1.062	0.060	<b>3.58E-29</b>	rs463105	127891427	C/A	0.018	<b>5.46E-26</b>	rs4663105	127891427	C/A	0.411	0.030	<b>3.38E-44</b>	?+++
4	2	rs10933431	233981912	G/C	0.971	-0.029	1.67E-06	rs10933431	233981912	G/C	-0.008	2.51E-06	rs10933431	233981912	G/C	0.244	-0.015	<b>8.92E-10</b>	?--?
5	3	NA						rs184384746	57226150	T/C	0.009	<b>1.24E-08</b>	rs184384746	57226150	T/C	0.001	0.195	<b>1.24E-08</b>	?--?
6	4	rs6448453	11026028	A/G	1.014	0.014	0.023641	rs6448451	11024682	G/C	0.009	<b>1.19E-08</b>	rs6448453	11026028	A/G	0.262	0.014	<b>1.93E-09</b>	?++
7	4	rs7657553	11723235	A/G	1.033	0.033	<b>2.16E-08</b>	rs7657553	11723235	A/G	0.000	7.90E-01	rs7657553	11723235	A/G	0.282	0.005	0.051	?++
8	6	rs2698533	32550322	A/C	1.037	0.036	<b>2.66E-08</b>	rs6931277	32583357	T/A	-0.009	1.78E-07	rs6931277	32583357	T/A	0.156	-0.019	<b>8.41E-11</b>	?--?
9	6	NA						rs187370608	40942196	A/G	0.014	<b>1.45E-16</b>	rs187370608	40942196	A/G	0.002	0.226	<b>1.45E-16</b>	?++?
10	6	rs381563	47432637	C/T	1.033	0.032	<b>5.35E-09</b>	rs9381563	47432637	C/T	0.007	8.10E-06	rs9381563	47432637	C/T	0.356	0.014	<b>2.52E-10</b>	?+++
11	7	rs1859788	99971834	A/G	0.968	-0.033	<b>6.05E-09</b>	rs7384878	99932049	C/T	-0.011	<b>2.38E-10</b>	rs1859788	99971834	A/G	0.324	-0.018	<b>2.22E-15</b>	---
12	7	rs11763230	143108841	T/C	0.957	-0.044	<b>2.58E-11</b>	rs7810606	143108158	T/C	-0.008	1.01E-06	rs7810606	143108158	T/C	0.485	-0.014	<b>3.59E-11</b>	?--?
13	7	NA						rs114360492	145950029	T/C	0.010	<b>2.10E-09</b>	rs114360492	145950029	T/C	0.002	0.174	<b>2.10E-09</b>	?++?
14	8	rs4236673	27464929	A/G	0.952	-0.049	<b>6.36E-20</b>	rs1532278	27466315	T/C	-0.010	<b>7.45E-09</b>	rs4236673	27464929	A/G	0.378	-0.020	<b>2.61E-19</b>	---
15	10	rs11257242	11721119	C/G	0.970	-0.031	<b>2.38E-08</b>	rs11257238	11717397	C/T	0.007	5.84E-05	rs11257238	11717397	C/T	0.361	0.013	<b>1.26E-08</b>	?+++
16	11	rs7935829	59942815	G/A	0.962	-0.039	<b>8.21E-13</b>	rs1582763	60021948	A/G	-0.010	<b>4.72E-09</b>	rs2081545	59958380	A/C	0.382	-0.017	<b>1.55E-15</b>	---
17	11	rs10792832	85867875	A/G	0.954	-0.047	<b>1.12E-17</b>	rs3841443	85850243	C/T	-0.011	<b>5.31E-11</b>	rs8676161	85776544	G/A	0.317	-0.020	<b>2.19E-18</b>	2---
18	11	rs11218343	121435587	C/T	0.919	-0.084	<b>5.57E-11</b>	rs11218343	121435587	C/T	-0.008	2.81E-06	rs11218343	121435587	C/T	0.044	-0.035	<b>1.09E-11</b>	?---
19	14	rs12590654	92938855	A/G	0.969	-0.031	<b>1.98E-08</b>	rs12590654	92938855	A/G	-0.008	3.70E-06	rs12590654	92938855	A/G	0.336	-0.014	<b>1.65E-10</b>	?--?
20	15	rs442495	59022615	C/T	0.980	-0.020	0.000309	rs442495	59022615	C/T	-0.009	2.65E-07	rs442495	59022615	C/T	0.354	-0.013	<b>1.31E-09</b>	?--?
21	15	rs117618017	63569902	T/C	1.018	0.018	0.022178	rs117618017	63569902	T/C	0.008	2.64E-07	rs117618017	63569902	T/C	0.125	0.018	<b>3.35E-08</b>	+++
22	16	rs59735493	31133100	A/G	0.981	-0.019	0.000825	rs59735493	31133100	A/G	-0.008	3.72E-06	rs59735493	31133100	A/G	0.298	-0.013	<b>3.98E-08</b>	?---
23	17	rs13260531	51318980	A/G	1.038	0.037	3.21E-06	rs9916042	4984447	A/G	0.009	<b>4.73E-08</b>	rs13260531	51318980	A/G	0.126	0.019	<b>9.16E-10</b>	?+++
24	17	rs28394864	47450775	A/G	1.021	0.021	7.29E-05	rs28394864	47450775	A/G	0.007	6.80E-06	rs28394864	47450775	A/G	0.453	0.012	<b>1.87E-08</b>	?+++
25	17	rs2632516	56409089	C/G	0.968	-0.032	<b>1.42E-09</b>	rs2632516	56409089	C/G	-0.005	5.29E-03	rs2632516	56409089	C/G	0.465	-0.010	9.66E-07	?--?
26	18	rs8093731	29088958	T/C	0.886	-0.121	<b>4.63E-08</b>	rs8093731	29088958	T/C	0.000	7.66E-01	rs8093731	29088958	T/C	0.019	-0.017	0.030	?--?
27	18	rs76726049	56189459	C/T	1.053	0.052	0.039261	rs76726049	56189459	C/T	0.009	1.83E-07	rs76726049	56189459	C/T	0.011	0.055	<b>3.30E-08</b>	?+++
28	19	rs147929	1063443	A/G	1.041	0.040	<b>8.64E-09</b>	rs3752241	1063524	G/C	-0.009	<b>2.87E-08</b>	rs112178892	1093323	G/C	0.147	0.019	<b>7.93E-11</b>	?++
29	19	rs1289512	45351516	G/C	1.521	0.419	<b>2.70E-194</b>	rs75627662	45413576	T/C	0.061	<b>9.51E-296</b>	rs1289512	45351516	G/C	0.036	0.200	<b>5.79E-276</b>	?+++
30	20	rs76320948	46241841	T/C	1.079	0.076	<b>1.54E-05</b>	rs76320948	46241841	T/C	0.007	1.80E-05	rs76320948	46241841	T/C	0.031	0.034	<b>4.64E-08</b>	?++?
31	19	rs3865444	51727962	A/C	0.969	-0.032	<b>4.25E-08</b>	rs3865444	51727962	A/C	-0.007	4.97E-05	rs3865444	51727962	A/C	0.299	-0.013	<b>6.34E-09</b>	?--?
32	20	rs6014724	54998544	G/A	0.953	-0.048	8.72E-08	rs6014724	54998544	T/C	-0.007	6.32E-06	rs6014724	54998544	G/A	0.095	-0.022	<b>6.56E-10</b>	?--?

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

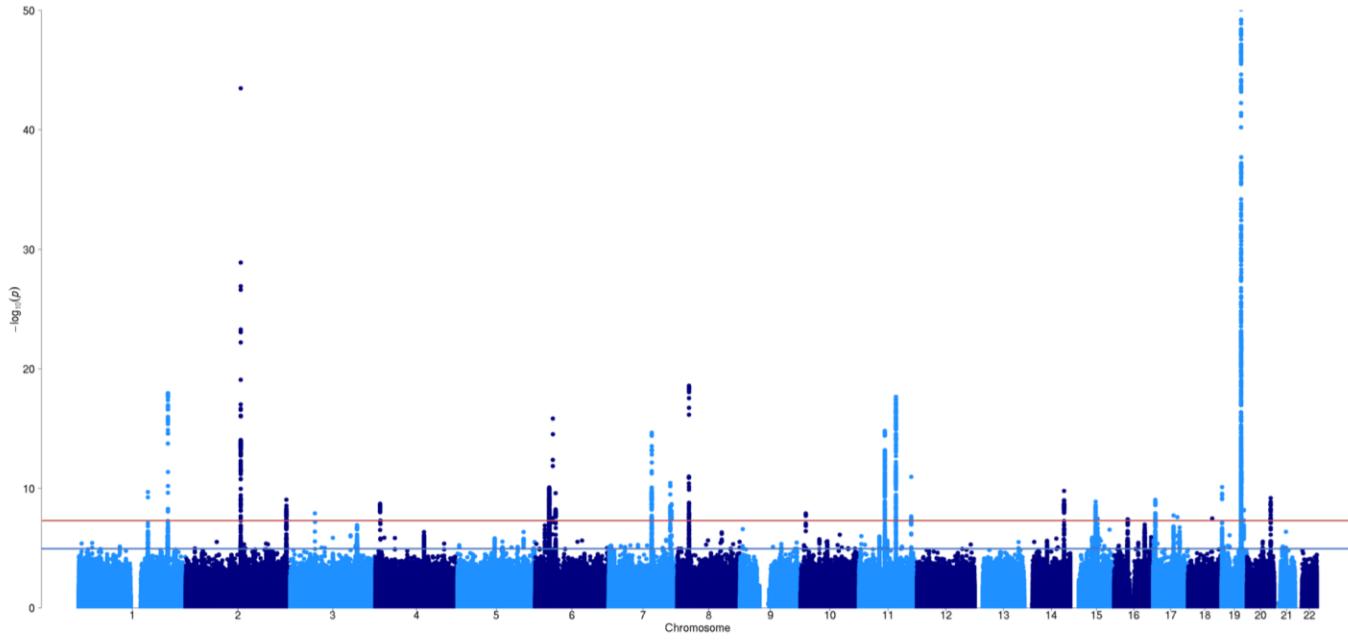
639 **Figure 1. Overview of analyses steps.** The main genetic analysis encompasses the procedures to detect  
640 GWAS risk loci for AD. The functional analysis part includes the *in silico* functional follow-up procedures  
641 with the aim to put the genetic findings in biological context. N = total of individuals within specified  
642 dataset.  
643  
644  
645



646  
647  
648  
649  
650  
651  
652  
653  
654

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

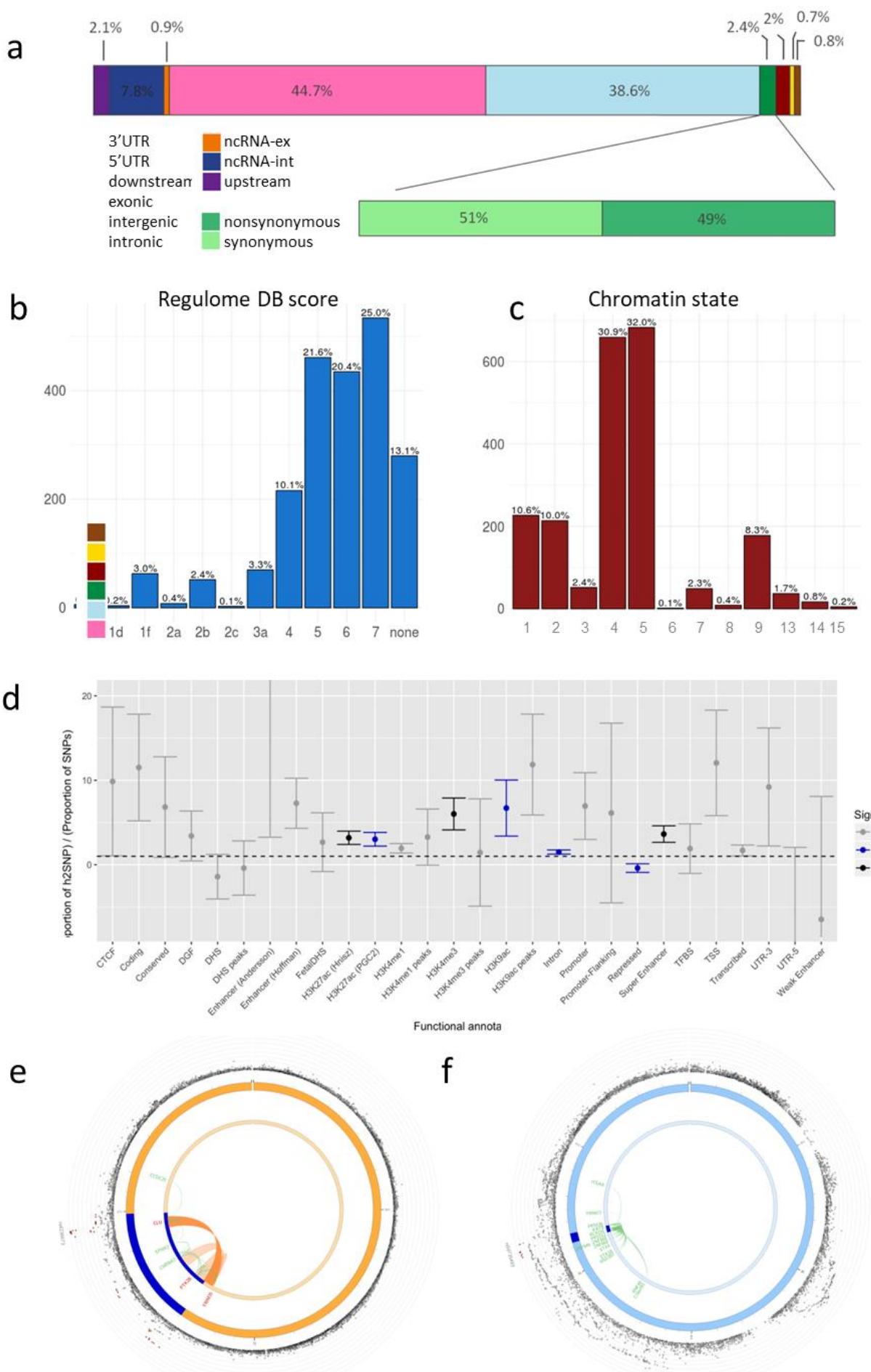
655 **Figure 2. GWAS results for AD risk (N=455,258).** Manhattan plot displays all associations per variant  
656 ordered according to their genomic position on the x-axis and showing the strength of the association  
657 with the  $-\log_{10}$  transformed P-values on the y-axis. The y-axis is limited at 50 to enable visualization of  
658 non-APOE loci. The original  $-\log_{10}$  for the APOE locus is 276.  
659



660  
661  
662  
663  
664  
665 **Figure 3. Functional annotation of association results.** a) Heritability enrichment of 28 functional  
666 variant annotations calculated with stratified LD score regression. UTR=untranslated region;  
667 CTCF=CCCTC-binding factor; DHS=DNasel hypersensitive site; TFBS=transcription factor binding site;  
668 DGF=DNAasel digital genomic footprint; b) Functional effects of genome-wide significant variants in  
669 genomic risk loci of the meta-analysis – the second bar shows the distribution for exonic variants only; c)  
670 Distribution of RegulomeDB score for variants in genomic risk loci, with a low score indicating a higher  
671 probability of having a regulatory function; d) Distribution of minimum chromatin state across 127 tissue  
672 and cell types for genome-wide significant variants in genomic risk loci, with lower states indicating  
673 higher accessibility and states 1-7 referring to open chromatin states. e) Zoomed-in circos plot of  
674 chromosome 8; f) Zoomed-in circos plot of chromosome 16. Circos plots show implicated genes by  
675 significant loci, where blue areas indicate genomic risk loci, green indicates eQTL associations and  
676 orange indicates chromatin interactions. Genes mapped by both eQTL and chromatin interactions are  
677 red. The outer layer shows a Manhattan plot containing the negative log10-transformed P-value of each  
678 SNP in the GWAS meta-analysis of AD. Full circos plots of all autosomal chromosomes are provided in  
679 Supplementary Figure 4.

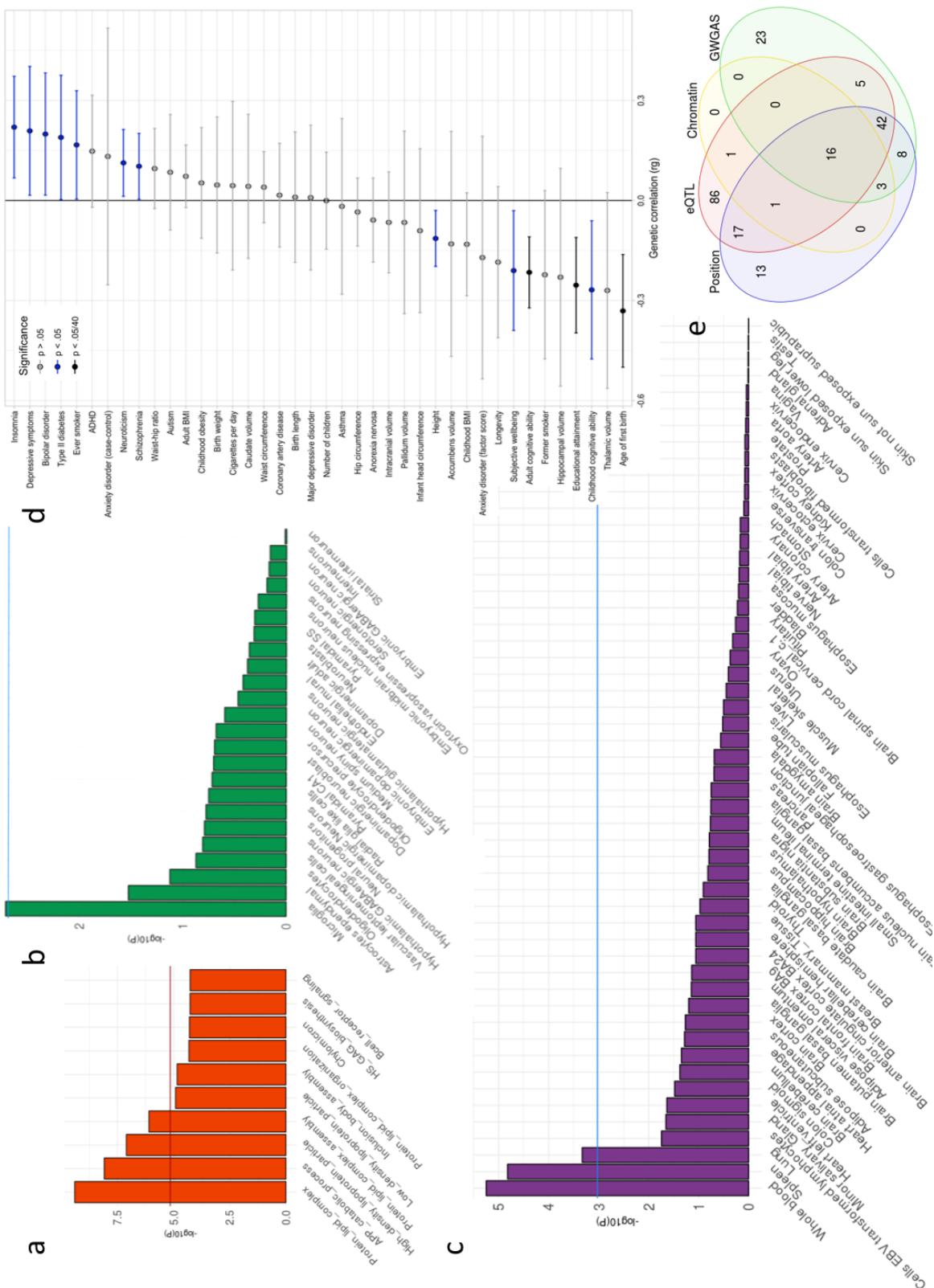
## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

680



## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

**Figure 4. Functional implications based on gene-set analysis, genetic correlations and functional annotations.** The gene-set results are displayed per category of biological mechanisms (A), brain cell types (B) and tissue types (C). The red horizontal lines indicates the significance threshold corrected for all gene-set tests of all categories, while the blue horizontal lines display the significance threshold corrected only for the number of tests within the three categories (i.e. gene-ontology, tissue expression, single cell expression). (D) Genetic correlations between AD and other heritable traits. (E) Venn diagram showing the number of genes mapped by four distinct strategies.



## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

706 **Online methods**

707

708 1.1 Study Cohorts

709 *1.1.1 PGC-ALZ cohorts*

710 Three non-public datasets (the Norwegian DemGene network, The Swedish Twin Studies of  
711 Aging and TwinGene) were meta-analyzed as part of the Alzheimer workgroup initiative of the  
712 Psychiatric Genomic Consortium (PGC-ALZ).

713 We collected genotype data from the Norwegian DemGene Network consisting of 2,224  
714 cases and 1,855 healthy controls. The DemGene Study is a Norwegian network of clinical sites  
715 collecting cases from Memory Clinics based on standardised examination of cognitive,  
716 functional and behavioural measures and data on progression of most patients. We diagnosed  
717 2,224 cases of AD from 7 studies: the Norwegian Register of persons with Cognitive Symptoms  
718 (NorCog), the Progression of Alzheimer's Disease and Resource use (PADR), the Dementia Study  
719 of Western Norway (DemVest), the AHUS study, the Dementia Study in Rural Northern Norway  
720 (NordNorge), the HUNT Dementia Stud, the Nursing Home study, and the TrønderBrain study.  
721 These cases were diagnosed according to the recommendations from the National Institute on  
722 Aging-Alzheimer's Association (NIA-AA) (AHUS), the NINCDS-ADRDA criteria (DemVest and  
723 TrønderBrain) or the ICD-10 research criteria (NorCog, PADR, NordNorge and HUNT). The  
724 controls from Norway were obtained through the AHUS, NordNorge, HUNT and TrønderBrain  
725 studies. The controls were screened with standardized interview and cognitive tests. Genotypes  
726 of the 4079 individuals from the DemGene Study were obtained with Human Omni Express-24  
727 v.1.1 (Illumina Inc., San Diego, CA, USA) at deCODE Genetics (Reykjavik, Iceland). To increase

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

728 the statistical power of our association analysis, the controls were combined with additional  
729 5786 population controls from Norwegian blood donor samples (Oslo University Hospital,  
730 Ullevål Hospital, Oslo) and controls from Thematically Organized Psychosis (TOP) Research  
731 Study (between 25-65 years). Control subjects of the TOP Research Study were of Caucasian  
732 origin without history of moderate/severe head injury, neurological disorder, mental  
733 retardation and were excluded if they or any of their close relatives had a lifetime history of a  
734 severe psychiatric disorder, a history of medical problems thought to interfere with brain  
735 function or significant illicit drug use.

736 The Swedish Twin Studies of Aging (STSA) (n cases = 398, n controls = 1079) includes  
737 three sub-studies of aging within the Swedish Twin Registry<sup>36</sup>: The Swedish Adoption/Twin  
738 Study of Aging (SATSA)<sup>37</sup>, Aging in Women and MEN (GENDER)<sup>38</sup>, and The Study of Dementia in  
739 Swedish Twins (HARMONY)<sup>39</sup>. Informed consent was obtained from all participants and the  
740 studies were approved by the Regional Ethics Board in Stockholm and the Institutional Review  
741 Board at the University of Southern California. DNA was extracted from blood samples and  
742 genotyped using Illumina Infinium PsychArray. Alzheimer's disease patients were diagnosed as  
743 part of the studies according to the NINCDS/ADRDA criteria<sup>40</sup>. In addition, information on  
744 disease after last study participation was retrieved from three population-based health care  
745 registers: The National Patient Register, the Causes of Death Register, and the Prescribed Drug  
746 Register.

747 TwinGene<sup>36</sup> is a population-based study of older twins drawn from the Swedish Twin  
748 Registry. Written informed consent was obtained from all participants and the study was  
749 approved by the Regional Ethics Board in Stockholm. DNA was extracted from blood samples

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

750 and genotyped using Illumina Human OmniExpress for 1791 individuals. Information about  
751 Alzheimer's disease (n cases = 343, n controls = 9070) was extracted from the National Patient  
752 Register, the Causes of Death Register, and the Prescribed Drug Register, all of which are  
753 population-based health care registers with nationwide coverage.

754

755 *1.1.2 IGAP*

756 Publically available ([http://web.pasteur-lille.fr/en/recherche/u744/igap/igap\\_download.php](http://web.pasteur-lille.fr/en/recherche/u744/igap/igap_download.php))  
757 genome-wide association analysis results of the International Genomics of Alzheimer's Project  
758 (IGAP)<sup>4</sup> were included as one of the four cohorts that were meta-analysed in our effort. IGAP is  
759 a large two-stage study based upon genome-wide association studies (GWAS) on individuals of  
760 European ancestry. We focused on the results of stage 1, for which IGAP used genotyped and  
761 imputed data of 7,055,881 single nucleotide polymorphisms (SNPs) to meta-analyse four  
762 previously-published GWAS datasets consisting of 17,008 Alzheimer's disease cases and 37,154  
763 controls (The European Alzheimer's disease Initiative – EADI, the Alzheimer Disease Genetics  
764 Consortium – ADGC, the Cohorts for Heart and Aging Research in Genomic Epidemiology  
765 consortium – CHARGE, the Genetic and Environmental Risk in AD consortium – GERAD). As the  
766 purpose of stage 2 (11,632 SNPs were genotyped and tested for association in an independent  
767 set of 8,572 Alzheimer's disease cases and 11,312 controls) was replication of the significantly  
768 associated loci of stage 1, we limited the inclusion of the summary statistics for our own  
769 analyses to stage 1. Written informed consent was obtained from study participants or, for  
770 those with substantial cognitive impairment, from a caregiver, legal guardian or other proxy,

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

771 and the study protocols for all populations were reviewed and approved by the appropriate  
772 institutional review boards.

773

774 *1.1.3 ADSP*

775 The Alzheimer's Disease Sequencing Project (ADSP) collaboration has the aim to identify novel  
776 genetic factors that contribute to AD risk by studying genetic sequencing data. ADSP has made  
777 their sequencing data available through the Genotypes and Phenotypes database (dbGaP) under  
778 the study accession: phs000572.v7.p ([https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs000572.v1.p1](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000572.v1.p1)). We have obtained access to 10,907 individuals  
780 (5,771 cases, 5,136 controls) with whole-exome sequencing data to include as the second  
781 cohort within our meta-analysis. A substantial proportion of the ADSP individuals were  
782 previously also included in IGAP. We applied two strategies to prevent inflated meta-analysis  
783 results due to sample overlap: (1) exclusion of ADSP individuals that were duplicates based on  
784 genotype data comparison of individual level genetic data between IGAP and ADSP, (2) perform  
785 meta-analysis while correcting for cross-study LD score regression intercept (see section 1.4.).  
786 To accomplish the first approach we obtained access for all IGAP datasets for which individual  
787 level genotype data was available through dbGaP (phs000160.v1.p1 - [https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs000160.v1.p1](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000160.v1.p1);  
788 phs000219.v1.p1 - [https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs000219.v1.p1](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000219.v1.p1);  
789 phs000372.v1.p1 - [https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs000372.v1.p1](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000372.v1.p1);  
790 https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\_id=phs000372.v1.p1;  
791 phs000168.v2.p2 - [https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs000168.v2.p2](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000168.v2.p2) - [https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs000168.v2.p2](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000168.v2.p2)

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

793 phs000168.v2.p2; phs000234.v1.p1 - [https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs000234.v1.p1](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000234.v1.p1)) or NIAGADS (NG00026 -  
794 https://www.niagads.org/datasets/ng00026; NG00028 -  
795 https://www.niagads.org/datasets/ng00028; NG00029 - <https://www.niagads.org/datasets/ng00029>; NG00031 - <https://www.niagads.org/datasets/ng00030> ; NG00031 -  
796 https://www.niagads.org/datasets/ng00031; NG00034 -  
797 https://www.niagads.org/datasets/ng00034). By calculating identity-by-descent using PLINK<sup>41</sup>,  
800 we identified duplicates, which were excluded from the ADSP WES dataset for subsequent  
801 analyses.

802

### 803 1.1.1 UK Biobank study

804 The current study used data from the UK Biobank<sup>42</sup> (UKB; [www.ukbiobank.ac.uk](http://www.ukbiobank.ac.uk)), a large  
805 population-based cohort that includes over 500,000 participants and aims to improve insight  
806 into a wide variety of health-related determinants and outcomes across the UK. Between 2006  
807 and 2010, approximately 9.2 million invitations to participate in the study were sent to  
808 individuals aged 40-69 years who were registered with the National Health Service (NHS) and  
809 were living within 25 miles from one of the 22 study research centers. In total, 503,325  
810 participants were recruited in the study, from which we used a subsample of individuals of  
811 European ancestry with available phenotypic and genotypic data ( $M$  age = 56.5, 54.0% female),  
812 described in more detail below. Besides phenotypic information obtained from the NHS  
813 registries and associated medical records, participants completed an in-person visit at one of  
814 the study research centers where extensive self-report data were collected by questionnaire in

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

815 addition to anthropometric assessments, DNA collection from blood samples, and magnetic  
816 resonance imaging of body and brain. All participants provided written informed consent; the  
817 UKB received ethical approval from the National Research Ethics Service Committee North  
818 West-Haydock (reference 11/NW/0382), and all study procedures were in accordance with the  
819 World Medical Association for medical research. Access to the UK Biobank data was obtained  
820 under application number 16406.

821

### 822 1.2 UKB by proxy phenotype

823 A proxy phenotype for Alzheimer's disease case-control status in UKB was assessed as part of  
824 the self-report questionnaire administered during the in-person assessment. Participants were  
825 asked to report whether their biological mother or father ever suffered from Alzheimer's  
826 disease/dementia, and to report each parent's current age (or age at death, if applicable). Of  
827 376,113 individuals in our analytic subsample who completed these questions, a diagnosis was  
828 reported for 32,327 mothers (8.6%) and 17,014 fathers (4.5%), resulting in 47,793 participants  
829 (12.7%) with one or both parents affected. We created a proxy phenotype from these questions  
830 to index genetic risk for Alzheimer's based on parents' diagnoses. The phenotype was  
831 constructed as a linear count of the number of affected biological parents (0, 1, or 2). The  
832 contribution for each unaffected parent to this count was weighted by the parent's age/age at  
833 death to account for the fact that they may not yet have passed through the period of risk for  
834 this late-onset disease. Specifically, each affected parent contributed one full unit of "risk" to  
835 the count, while each unaffected parent contributed a proportion of one unit of "risk" inversely  
836 related to their age. This was calculated as the ratio of parent's age to age 100 (approximately

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

837 the 95th percentile for life expectancy in developed countries, such that weight=(100-age)/100.  
838 The weight for unaffected parents was capped at 0.32, corresponding to a risk equivalent to  
839 that of the maximum population prevalence of AD.<sup>43</sup> The phenotype thus ranged approximately  
840 from 0 to 2, with values near zero when both parents were unaffected (lower for older parents  
841 and possible values below zero if both parents were over age 100) and values of two when both  
842 parents were affected. Participants who were uncertain or chose not to answer questions  
843 about either parent's disease status or age were excluded from the analyses, resulting in a final  
844 N=364,859.

845 Additional information on Alzheimer's disease risk was obtained from national medical  
846 records linked to participant data. This information pertained to the participants themselves  
847 (not their parents), and was extracted from hospital records obtained between 1996 and the  
848 present or from national death registries in the case of participants who passed away after  
849 initial enrolment in the study, as described in more detail in the UKB resources  
850 (<http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=146641>;  
851 <http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=115559>). Briefly, primary and secondary  
852 diagnoses from inpatient hospital stays and primary and secondary causes of death from death  
853 records were recorded using ICD-10 codes. Participants with a diagnosis of "Alzheimer's  
854 disease" (diseases of the nervous system chapter; code G30) or "Dementia in Alzheimer's  
855 disease" (mental and behavioral disorders chapter; code F00) from any record of a hospital stay  
856 or as a cause of death were treated as Alzheimer's cases as given the maximum possible "risk"  
857 score of 2, regardless of the affection status of their parents. The reported rate of Alzheimer's  
858 in parents of cases (27.4%) was more than double that of non-cases (12.7%;  $\chi^2(1)=71.7$ ,

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

859  $P=2.45E-17$ ). There were 393 individuals in the analytic subsample classified as affected by  
860 these records; due to the small number of cases and the limited representativeness of these  
861 types of health records, we used this information to supplement the proxy parent phenotype  
862 rather than as a primary outcome. This information reduces the possibility of misclassification  
863 in the proxy phenotype method, and also allows us to evaluate the performance of the proxy  
864 phenotype method.

865

### 866 1.3 Genome-wide association analysis

867 Except for IGAP (obtained summary statistics), we performed genome-wide association  
868 analyses for the ADSP, PGC-ALZ and UKB cohorts. For the UKB dataset, quality control and  
869 imputation procedures were slightly different, and therefore described separately in the  
870 sections below.

871

#### 872 *1.3.1a Quality control and imputation procedures for ADSP and PGC-ALZ datasets*

873 Prior to individual quality control steps, all datasets were filtered on a max missingness of 5%.  
874 Individuals were excluded when identified as a low quality sample (individual call rate < 0.98),  
875 heterozygosity outlier ( $F +/- .20$ ), gender mismatch (females:  $F > 0.2$ , males:  $F < 0.2$ ) when  
876 comparing phenotypic and genotypic data, population outlier (defined by principal component  
877 boundaries of 1000 Genomes European samples) or being related to another sample ( $PI\_HAT >$   
878 0.2). Inclusion criteria for variants encompassed a call rate > 0.98, a case-control missingness  
879 difference < 0.02, a Hardy-Weinberg equilibrium  $p$ -value <  $10 \times 10^{-6}$  for controls ( $< 10 \times 10^{-10}$  for  
880 cases) and a valid association  $p$ -value (excluding the variants with low allele frequencies).

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

881 Pre-imputation, the ADSP and PGC-ALZ datasets were checked for palindromic variants  
882 with allele frequency close to 0.5, incorrect reference allele definitions, false strand designation  
883 and extreme deviations from expected allele frequencies. Subsequently the ADSP and PGC-ALZ  
884 datasets were imputed with the 1000 Genomes Phase 3<sup>44</sup> reference panel. The reported SNPs  
885 all have a considerable imputation quality (INFO score>0.591) and variants with a low allele  
886 frequency (MAF<0.01) were excluded, resulting in a total of 7508 individuals (4343 cases and  
887 3165 controls) and 260,934 variants for the ADSP cohort and 17477 individuals (2,736 cases and  
888 14,471 controls) and 9,629,492 variants for the PGC-ALZ cohort.

889

### 890 *1.3.1b Quality control and imputation for UKB dataset*

891 We used second-release genotype data that were made available by UKB in July 2017.  
892 Genotype data collection and processing are described by the UKB in a previous overview  
893 paper<sup>45</sup>. DNA was extracted from blood samples and genotyping was completed for 488,366  
894 individuals on one of two Affymetrix genotyping arrays with custom content, the UK BiLEVE  
895 Axiom array (N=49,949) or UK Biobank Axiom array (N=438,417), covering 812,428 genetic  
896 markers common to both arrays. Of these, 488,377 individuals and 805,426 markers passed the  
897 genotype quality control checks conducted by UKB (see  
898 <http://www.biorxiv.org/content/early/2017/07/20/166298> for details). Samples were excluded  
899 for low DNA concentration, call rate < 95%, excess heterozygosity, sex chromosome  
900 abnormality, or sample duplication. Variants were excluded if they exhibited poor clustering of  
901 allele calls, batch, plate, array, or sex effects, departures from HWE, or discordance between  
902 technical replicate samples.

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

903        After quality control, the samples were imputed to approximately 92 million SNPs using  
904        both the reference panel of the Haplotype Reference Consortium (HRC)<sup>46</sup> as well as a combined  
905        reference panel of the 1000 Genomes Project<sup>44</sup> and UK10K. As recommended by UKB, we  
906        removed variants that were not imputed on the HRC reference panel due to technical errors in  
907        the imputation process of the combined panel. We converted imputed variants to hard calls  
908        (certainty > 0.9), filtered by imputation quality (INFO score >0.9), and excluded multi-allelic  
909        SNPs, indels, SNPs without unique rsID, and SNPs with minor allele frequency (MAF) <0.0001,  
910        resulting in 10,847,151 SNPs available for analysis.

911        For the present study, we selected unrelated individuals of European ancestry. To  
912        empirically determine ancestry, we projected genetic principal components from known  
913        ancestral populations in the 1000 Genomes Project onto the UKB genotypes and assigned  
914        individuals to the continental ancestral superpopulation with the closest Mahalanobis  
915        distance.<sup>47</sup> Within-ancestry principal components were created using FlashPCA2<sup>48</sup> to correct for  
916        any residual population stratification within the European ancestry subset. Unrelated  
917        individuals (less than 3rd degree relatives, as indicated by genomic relatedness coefficients  
918        calculated by UKB) were selected by sequentially removing participants with the greatest  
919        number of relatives until no related pairs remained. After applying these filtering criteria and  
920        removing any participants with missing phenotypic or covariate data and participants who  
921        withdrew consent, 364,859 individuals remained for analysis in the UKB sample.

922

923        *1.3.2 Single-marker association analysis*

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

924 Genome-wide association analysis (GWAS) for the ADSP, PGC-ALZ and UKB datasets was  
925 performed in PLINK<sup>41</sup>, using logistic regression for dichotomous phenotypes (cases versus  
926 controls for ADSP and PGC-ALZ cohorts), and linear regression for phenotypes analysed as  
927 continuous outcomes (by proxy parental AD phenotype for UKB cohort). For the ADSP and PGC-  
928 ALZ cohorts, association tests were adjusted for gender, batch (if applicable), and the first 4  
929 principal components. Twenty principal components were calculated, and depending on the  
930 dataset being tested, additional principal components (on top of the standard inclusion of 4  
931 PCAs) were added if significantly associated to the phenotype. Furthermore, for the PGC-ALZ  
932 cohorts age was included as a covariate. For 4,537 controls of the DemGene cohort, no detailed  
933 age information was available, besides the age range the subjects were in (20-45 years). We  
934 therefore set the age of these individuals conservatively to 20 years. For the ADSP dataset, age  
935 was not included as a covariate due to the enrichment for older controls (mean age cases =  
936 73.1 years (SE=7.8); mean age controls = 86.1 years (SE=4.5)) in their collection procedures.  
937 Correcting for age in ADSP would remove a substantial part of genuine association signals (e.g.  
938 well-established *APOE* locus rs11556505 is strongly associated to AD ( $P=1.08\times10^{-99}$ ), which is  
939 lost when correcting for age ( $P=0.0054$ ). For the UKB dataset, 12 components were included as  
940 covariates, as well as age, sex, genotyping array, and assessment centre. We used the genome-  
941 wide threshold for significance of  $P<5\times10^{-8}$ .

942

943 1.3.3 *Multivariate genome-wide meta-analysis*

944 Two meta-analyses were performed, including: 1) cohorts with case-control phenotypes (IGAP,  
945 ADSP and PGC-ALZ datasets), 2) all cohorts, also including the by proxy phenotype of UKB.

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

946 The per SNP test statistics is defined by

947

948 
$$Z_k = \frac{\sum_i w_i Z_i}{\sqrt{\sum_i w_i^2 + \sum_i \sum_j w_i w_j |CTI_{ij}| (i \neq j)}}$$

949

950 where  $w_i$  and  $Z_i$  are the squared root of the sample size and the test statistics of SNP  $k$  in cohort  
951  $i$ , respectively. CTI is the cross trait LD score intercept estimated by LDSC using genome-wide  
952 summary statistics as

953 
$$CTI_{ij} = \frac{N_{sij} \rho_{ij}}{\sqrt{N_i N_j}}$$

954

955 where  $N_{sij}$  and  $\rho_{ij}$  are the number of overlapping samples and the phenotypic correlation  
956 between cohort  $i$  and  $j$ , respectively.<sup>14</sup> The test statistics per SNP per GWAS were converted  
957 from the P-value by using the sign of either beta or odds ratio. When direction is aligned the  
958 conversion is two-sided. To avoid infinite values, we replaced P-value 1 with 0.999999 and P-  
959 value < 1e-323 to 1e-323 (the minimum >0 value in Python).

960 The effective sample size ( $N_{eff}$ ) is computed for each SNP  $k$  from the matrix  $M$ ,  
961 containing the sample size  $N_i$  of each cohort  $i$  on the diagonal and the estimated number of  
962 shared data points  $N_{sij} \rho_{ij} = CTI_{ij} \sqrt{N_i N_j}$  for each pair of cohorts  $i$  and  $j$  as the off-diagonal  
963 values.  $N_{eff}$  is computed recursively as follows. Starting with the first cohort in  $M$ ,  $N_{eff}$  is first  
964 increased by  $M_{1,1}$ , corresponding to the sample size of that cohort. The proportion of samples  
965 shared between cohort 1 and each other cohort  $j$  is then computed as  $p_{1,j} = M_{1,j}/M_{jj}$ , and  $M$  is  
966 then adjusted to remove this overlap, multiplying all values in each column  $j$  by  $1-p_{1,j}$ . This  
967 amounts to reducing the sample size of each other cohort  $j$  by the number of samples it shares

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

968 with cohort 1, and reducing the shared samples between cohort  $j$  and subsequent cohorts by  
969 the same proportion. After this, the first row and column of  $M$  are discarded, and the same  
970 process is applied to the new  $M$  matrix. This is repeated until  $M$  is empty. The script for the  
971 multivariate GWAS is available from <https://github.com/Kyoko-wtnb/mvGWAMA>.

972

973 1.5 Replication of meta-analysis result in an Icelandic sample

974 The study group included 6,593 Alzheimer's disease cases (4,923 of whom were chip-typed) and  
975 174,289 controls (88,581 of whom were chip-typed). In 16% of patients, the diagnosis of  
976 Alzheimer's disease was established at the Memory Clinic of the University Hospital according  
977 to the criteria for definite, probable, or possible Alzheimer's disease of the National Institute of  
978 Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related  
979 Disorders Association (NINCDS-ADRDA). In 77% of patients, the diagnosis has been registered  
980 according to the criteria for code 331.0 in ICD-9, or for F00 and G30 in ICD-10 in health records.  
981 Seven percent of the patients were identified in the Directorate of Health medication database  
982 as having been prescribed Donepezil (Aricept). The controls were drawn from various research  
983 projects at deCODE Genetics.

984 The study was approved by the National Bioethics Committee and the Icelandic Data Protection  
985 Authority. Written informed consent was obtained from all participants or their guardians  
986 before blood samples were drawn. All sample identifiers were encrypted in accordance with  
987 the regulations of the Icelandic Data Protection Authority.

988 Chip-typing and long-range phasing of 155,250 individuals was carried out as described  
989 previously.<sup>21</sup> Imputation of the variants found in 28,075 whole-genome sequenced individuals

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

990 into the chip-typed individuals and 285,664 close relatives was performed as detailed earlier.<sup>21</sup>  
991 Association analysis was carried out using logistic regression with Alzheimer's disease status as  
992 the response and genotype counts and a set of nuisance variables including sex, county of birth,  
993 and current age as predictors.<sup>22</sup> Correction for inflation of test statistics due to relatedness and  
994 population stratification was performed using the intercept estimate from LD score regression<sup>14</sup>  
995 (1.29).

996

### 997 1.6 Genomic risk loci definition

998 We used FUMA<sup>26</sup>, an online platform for functional mapping and annotation of genetic variants,  
999 to define genomic risk loci and obtain functional information of relevant SNPs in these loci. We  
1000 first identified independent significant SNPs that have a genome-wide significant P-value  
1001 ( $<5\times10^{-8}$ ) and are independent from each other at  $r^2<0.6$ . These SNPs were further represented  
1002 by lead SNPs, which are a subset of the independent significant SNPs that are in approximate  
1003 linkage equilibrium with each other at  $r^2>0.6$ . We then defined associated genomic risk loci by  
1004 merging any physically overlapping lead SNPs (LD blocks  $<250\text{kb}$  apart). Borders of the genomic  
1005 risk loci were defined by identifying all SNPs in LD ( $r^2>0.6$ ) with one of the independent  
1006 significant SNPs in the locus, and the region containing all these candidate SNPs was considered  
1007 to be a single independent genomic risk locus. LD information was calculated using the UK  
1008 Biobank genotype data as a reference.

1009

### 1010 1.7 Cohort Heritability and Genetic Correlation

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

1011 LD score regression<sup>14</sup> was used to estimate genomic inflation and heritability of the AD in each  
1012 of the 7 cohorts (PGC-ALZ, ADSP, IGAP, UKB, DemGene, STSA, TwinGene) using their post-  
1013 quality control summary statistics, and to estimate the cross-cohort genetic correlations.<sup>49</sup> Pre-  
1014 calculated LD scores from the 1000 Genomes European reference population were obtained  
1015 from <https://data.broadinstitute.org/alkesgroup/LDSCORE/>. Genetic correlations were  
1016 calculated on HapMap3 SNPs only. LD score regression was also used on the case-control and  
1017 by-proxy phenotype result to estimate heritability and genetic correlations for the two  
1018 phenotype definitions.

1019

### 1020 1.8 Polygenic risk scoring

1021 We calculated polygenic scores (PGS) based on the SNP effect sizes estimated in the meta-  
1022 analyses. PGS were calculated using an independent genotype dataset of 761 individuals (379  
1023 cases and 382 controls) from the ADDNeuroMed study.<sup>50</sup> The same QC and imputation  
1024 approach was applied as for the other datasets with genotype-level data (see Method section  
1025 1.3.1a). PRSice PGS were calculated on hard-called imputed genotypes using *P*-value thresholds  
1026 from 0.0 to 0.5 in steps ranging from  $5 \times 10^{-8}$  to 0.001. The explained variance ( $\Delta R^2$ ) was derived  
1027 from a linear model in which the AD phenotype was regressed on each PGS while controlling for  
1028 the same covariates as in each cohort-specific GWAS, compared to a linear model with GWAS  
1029 covariates only.

1030

### 1031 1.9 Stratified Heritability

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

1032 To test whether specific categories of SNP annotations were enriched for heritability, we  
1033 partitioned the SNP heritability for binary annotations using stratified LD score regression  
1034 (<https://github.com/bulik/ldsc>)<sup>14</sup>. Heritability enrichment was calculated as the proportion of  
1035 heritability explained by a SNP category divided by the proportion of SNPs that are in that  
1036 category. Partitioned heritability was computed by 28 functional annotation categories, by  
1037 minor allele frequency (MAF) in six percentile bins and by 22 chromosomes. Annotations for  
1038 binary categories of functional genomic characteristics (e.g. coding or regulatory regions) were  
1039 obtained from the LD score website (<https://github.com/bulik/ldsc>). The Bonferroni-corrected  
1040 significance threshold for 56 annotations was set at:  $P<0.05/56=8.93\times10^{-4}$ .

1041

### 1042 1.10 Functional Annotation of SNPs

1043 Functional annotation of SNPs implicated in the meta-analysis was performed using FUMA<sup>26</sup>  
1044 (<http://fuma.ctglab.nl/>). We selected all *candidate SNPs* in the associated genomic loci having  
1045 an  $r^2 \geq 0.6$  with one of the independent significant SNPs (see above), a *P*-value ( $P<1\times10^{-8}$ ) and a  
1046 MAF>0.0001 for annotations. Functional consequences for these SNPs were obtained by  
1047 matching SNPs' chromosome, base-pair position, and reference and alternative alleles to  
1048 databases containing known functional annotations, including ANNOVAR<sup>51</sup> categories,  
1049 Combined Annotation Dependent Depletion (CADD) scores<sup>24</sup>, RegulomeDB<sup>52</sup> (RDB) scores, and  
1050 chromatin states<sup>53,54</sup>. ANNOVAR annotates the functional consequence of SNPs on genes (e.g.  
1051 intron, exon, intergenic). CADD scores predict how deleterious the effect of a SNP with higher  
1052 scores referring to higher deleteriousness. A CADD score above 12.37 is the threshold to be  
1053 potentially pathogenic<sup>55</sup>. The RegulomeDB score is a categorical score based on information

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

1054 from expression quantitative trait loci (eQTLs) and chromatin marks, ranging from 1a to 7 with  
1055 lower scores indicating an increased likelihood of having a regulatory function. Scores are as  
1056 follows: 1a=eQTL + Transcription Factor (TF) binding + matched TF motif + matched DNase  
1057 Footprint + DNase peak; 1b=eQTL + TF binding + any motif + DNase Footprint + DNase peak;  
1058 1c=eQTL + TF binding + matched TF motif + DNase peak; 1d=eQTL + TF binding + any motif +  
1059 DNase peak; 1e=eQTL + TF binding + matched TF motif; 1f=eQTL + TF binding / DNase peak;  
1060 2a=TF binding + matched TF motif + matched DNase Footprint + DNase peak; 2b=TF binding +  
1061 any motif + DNase Footprint + DNase peak; 2c=TF binding + matched TF motif + DNase peak;  
1062 3a=TF binding + any motif + DNase peak; 3b=TF binding + matched TF motif; 4=TF binding +  
1063 DNase peak; 5=TF binding or DNase peak; 6=other;7=None. The chromatin state represents the  
1064 accessibility of genomic regions (every 200bp) with 15 categorical states predicted by a hidden  
1065 Markov model based on 5 chromatin marks for 127 epigenomes in the Roadmap Epigenomics  
1066 Project<sup>39</sup>. A lower state indicates higher accessibility, with states 1-7 referring to open  
1067 chromatin states. We annotated the minimum chromatin state across tissues to SNPs. The 15-  
1068 core chromatin states as suggested by Roadmap are as follows: 1=Active Transcription Start Site  
1069 (TSS); 2=Flanking Active TSS; 3=Transcription at gene 5' and 3'; 4=Strong transcription; 5= Weak  
1070 Transcription; 6=Genic enhancers; 7=Enhancers; 8=Zinc finger genes & repeats;  
1071 9=Heterochromatic; 10=Bivalent/Poised TSS; 11=Flanking Bivalent/Poised TSS/Enh; 12=Bivalent  
1072 Enhancer; 13=Repressed PolyComb; 14=Weak Repressed PolyComb; 15=Quiescent/Low.  
1073

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

1074 1.11 Gene-mapping

1075 Genome-wide significant loci obtained by GWAS were mapped to genes in FUMA<sup>26</sup> using three  
1076 strategies:

1077 1. Positional mapping maps SNPs to genes based on physical distance (within a 10kb  
1078 window) from known protein coding genes in the human reference assembly  
1079 (GRCh37/hg19).

1080 2. eQTL mapping maps SNPs to genes with which they show a significant eQTL association  
1081 (i.e. allelic variation at the SNP is associated with the expression level of that gene).

1082 eQTL mapping uses information from 45 tissue types in 3 data repositories (GTEx<sup>56</sup>,  
1083 Blood eQTL browser<sup>57</sup>, BIOS QTL browser<sup>58</sup>), and is based on cis-eQTLs which can map  
1084 SNPs to genes up to 1Mb apart. We used a false discovery rate (FDR) of 0.05 to define  
1085 significant eQTL associations.

1086 3. Chromatin interaction mapping was performed to map SNPs to genes when there is a  
1087 three-dimensional DNA-DNA interaction between the SNP region and another gene  
1088 region. Chromatin interaction mapping can involve long-range interactions as it does not  
1089 have a distance boundary. FUMA currently contains Hi-C data of 14 tissue types from  
1090 the study of Schmitt et al<sup>59</sup>. Since chromatin interactions are often defined in a certain  
1091 resolution, such as 40kb, an interacting region can span multiple genes. If a SNPs is  
1092 located in a region that interacts with a region containing multiple genes, it will be  
1093 mapped to each of those genes. To further prioritize candidate genes, we selected only  
1094 genes mapped by chromatin interaction in which one region involved in the interaction  
1095 overlaps with a predicted enhancer region in any of the 111 tissue/cell types from the

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

1096 Roadmap Epigenomics Project<sup>54</sup> and the other region is located in a gene promoter  
1097 region (250bp up and 500bp downstream of the transcription start site and also  
1098 predicted by Roadmap to be a promoter region). This method reduces the number of  
1099 genes mapped but increases the likelihood that those identified will have a plausible  
1100 biological function. We used a FDR of  $1\times10^{-5}$  to define significant interactions, based on  
1101 previous recommendations<sup>44</sup> modified to account for the differences in cell lines used  
1102 here.

1103

### 1104 1.12 Gene-based analysis

1105 To account for the distinct types of genetic data in this study, genotype array (PGC-ALZ, IGAP,  
1106 UKB) and whole-exome sequencing data (ADSP), we first performed two gene-based genome-  
1107 wide association analysis (Gwgas) using MAGMA<sup>30</sup>, followed by a meta-analysis. SNP-based P-  
1108 values from the meta-analysis of the 3 genotype-array-based datasets were used as input for  
1109 the first Gwgas, while the unimputed individual-level sequence data of ADSP was used as  
1110 input for the second Gwgas. 18,233 protein-coding genes (each containing at least one SNP in  
1111 the GWAS) from the NCBI 37.3 gene definitions were used as basis for Gwgas in MAGMA.  
1112 Bonferroni correction was applied to correct for multiple testing ( $P<2.74\times10^{-6}$ ).

1113

### 1114 1.13 Gene-set analysis

1115 Results from the Gwgas analyses were used to test for association in 7,087 predefined gene-  
1116 sets of four types:

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

1117 1. 6,994 curated gene-sets representing known biological and metabolic pathways derived  
1118 from Gene Ontology (5917 gene-sets), Biocarta (217 gene-sets), KEGG (186 gene-sets),  
1119 Reactome (674 gene-sets) catalogued by and obtained from the MsigDB version 6.1<sup>60</sup>  
1120 (<http://software.broadinstitute.org/gsea/msigdb/collections.jsp>)  
1121 2. Gene expression values from 54 (53 + 1 calculated 1<sup>st</sup> PC of three tissue subtypes)  
1122 tissues obtained from GTEx<sup>56</sup>, log2 transformed with pseudocount 1 after winsorization  
1123 at 50 and averaged per tissue.  
1124 3. Cell-type specific expression in 173 types of brain cells (24 broad categories of cell types,  
1125 'level 1' and 129 specific categories of cell types 'level 2'), which were calculated  
1126 following the method described in<sup>32</sup>. Briefly, brain cell-type expression data was drawn  
1127 from single-cell RNA sequencing data from mouse brains. For each gene, the value for  
1128 each cell-type was calculated by dividing the mean Unique Molecular Identifier (UMI)  
1129 counts for the given cell type by the summed mean UMI counts across all cell types.  
1130 Single-cell gene-sets were derived by grouping genes into 40 equal bins based on  
1131 specificity of expression.  
1132 4. Nucleus specific gene expression of 15 distinct human brain cell-types from the study  
1133 described in<sup>61</sup>. The value for each cell-type was calculated with the same method as  
1134 explained in point 3 above.  
1135 These gene-sets were tested using MAGMA. We computed competitive *P*-values, which  
1136 represent the test of association for a specific gene-set compared with genes not in the gene-  
1137 set to correct for baseline level of genetic association in the data. The Bonferroni-corrected  
1138 significance threshold was 0.05/7,087 gene-sets=7.06×10<sup>-6</sup>. The suggestive significance

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

1139 threshold was defined by the number of tests within the category. Conditional analyses were  
1140 performed as a follow-up using MAGMA to test whether each significant association observed  
1141 was independent of all others and of *APOE* (a gene-set including all genes within genomic  
1142 region chr19:45,020,859-45,844,508). Furthermore, the association between each of the  
1143 significant gene-sets was tested conditional on each of the other significantly associated gene-  
1144 sets. Gene-sets that retained their association after correcting for other sets were considered to  
1145 represent independent signals. We note that this is not a test of association per se, but rather a  
1146 strategy to identify, among gene-sets with known significant associations and overlap in genes,  
1147 which set (s) are responsible for driving the observed association.

1148

### 1.14 Cross-Trait Genetic Correlation

1149 Genetic correlations ( $r_g$ ) between AD and 41 phenotypes were computed using LD score  
1150 regression<sup>14</sup>, as described above, based on GWAS summary statistics obtained from publicly  
1151 available databases (<http://www.med.unc.edu/pgc/results-and-downloads>; <http://ldsc.broadinstitute.org/>); **Supplementary Table 19**). The Bonferroni-corrected significance  
1152 threshold was  $0.05/41$  traits= $1.22 \times 10^{-3}$ .

1153

### 1.15 Mendelian Randomisation

1154 To infer credible causal associations between AD and traits that are genetically correlated with  
1155 AD, we performed Generalised Summary-data based Mendelian Randomisation<sup>31</sup> (GSMR;  
1156 <http://cnsgenomics.com/software/gsmr/>). This method utilizes summary-level data to test for  
1157 putative causal associations between a risk factor (exposure) and an outcome by using

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

1161 independent genome-wide significant SNPs as instrumental variables as an index of the  
1162 exposure. HEIDI-outlier detection was used to filter genetic instruments that showed clear  
1163 pleiotropic effects on the exposure phenotype and the outcome phenotype. We used a  
1164 threshold p-value of 0.01 for the outlier detection analysis in HEIDI, which removes 1% of SNPs  
1165 by chance if there is no pleiotropic effect. To test for a potential causal effect of various  
1166 outcomes on risk for AD, we selected phenotypes in non-overlapping samples that showed  
1167 (suggestive) significant ( $P<0.05$ ) genetic correlations ( $r_g$ ) with AD. With this method it is typical  
1168 to test for bi-directional causation by repeating the analyses while switching the role of the  
1169 exposure and the outcome; however, because AD is a late-onset disease, it makes little sense to  
1170 estimate its causal effect on outcomes that develop earlier in life, particularly when the  
1171 summary statistics for these outcomes were derived mostly from younger samples than those  
1172 of AD cases. Therefore, we conducted these analyses only in one direction. For genetically  
1173 correlated phenotypes, we selected independent ( $r^2=<0.1$ ), GWS lead SNPs as instrumental  
1174 variables in the analyses. The method estimates a putative causal effect of the exposure on the  
1175 outcome ( $b_{xy}$ ) as a function of the relationship between the SNPs' effects on the exposure ( $b_{zx}$ )  
1176 and the SNPs' effects on the outcome ( $b_{zy}$ ), given the assumption that the effect of non-  
1177 pleiotropic SNPs on an exposure (x) should be related to their effect on the outcome (y) in an  
1178 independent sample only via mediation through the phenotypic causal pathway ( $b_{xy}$ ). The  
1179 estimated causal effect coefficients ( $b_{xy}$ ) are approximately equal to the natural log odds ratio  
1180 ( $OR$ )<sup>31</sup> for a case-control trait. An OR of 2 can be interpreted as a doubled risk compared to the  
1181 population prevalence of a binary trait for every SD increase in the exposure trait. For  
1182 quantitative traits the  $b_{zx}$  and  $b_{zy}$  can be interpreted as a one standard deviation increase

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

1183 explained in the outcome trait for every SD increase in the exposure trait. This method can help  
1184 differentiate the causal direction of association between two traits, but cannot make any  
1185 statement about the intermediate mechanisms involved in any potential causal process.

1186

1187 *Data availability*

1188 Summary statistics will be made available for download upon publication (<https://ctg.cnrc.nl>).

1189

## 1190 References

1191

- 1192 1. Prince M, Bryce R, Albanese E, Wimo A, Ribeiro W, Ferri CP. The global prevalence of dementia: a systematic review and metaanalysis. *Alzheimer's & dementia : the journal of the Alzheimer's Association* 2013; **9**(1): 63-75.e2.
- 1193 2. Gatz M, Reynolds CA, Fratiglioni L, et al. Role of genes and environments for explaining Alzheimer disease. *Archives of general psychiatry* 2006; **63**(2): 168-74.
- 1194 3. Cacace R, Sleegers K, Van Broeckhoven C. Molecular genetics of early-onset Alzheimer's disease revisited. *Alzheimer's & dementia : the journal of the Alzheimer's Association* 2016; **12**(6): 733-48.
- 1195 4. Lambert JC, Ibrahim-Verbaas CA, Harold D, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature genetics* 2013; **45**(12): 1452-8.
- 1196 5. Goate A, Chartier-Harlin MC, Mullan M, et al. Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. *Nature* 1991; **349**(6311): 704-6.
- 1197 6. Sherrington R, Rogaei EI, Liang Y, et al. Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. *Nature* 1995; **375**(6534): 754-60.
- 1198 7. Sherrington R, Froelich S, Sorbi S, et al. Alzheimer's disease associated with mutations in presenilin 2 is rare and variably penetrant. *Human molecular genetics* 1996; **5**(7): 985-8.
- 1199 8. Karran E, Mercken M, De Strooper B. The amyloid cascade hypothesis for Alzheimer's disease: an appraisal for the development of therapeutics. *Nature reviews Drug discovery* 2011; **10**(9): 698-712.
- 1200 9. Jonsson T, Stefansson H, Steinberg S, et al. Variant of TREM2 associated with the risk of Alzheimer's disease. *The New England journal of medicine* 2013; **368**(2): 107-16.
- 1201 10. Steinberg S, Stefansson H, Jonsson T, et al. Loss-of-function variants in ABCA7 confer risk of Alzheimer's disease. *Nature genetics* 2015; **47**(5): 445-7.
- 1202 11. Liu CC, Liu CC, Kanekiyo T, Xu H, Bu G. Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. *Nature reviews Neurology* 2013; **9**(2): 106-18.
- 1203 12. Liu JZ, Erlich Y, Pickrell JK. Case-control association mapping by proxy using family history of disease. *Nature genetics* 2017; **49**(3): 325-31.

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

- 1221 13. Zheng J, Erzurumluoglu AM, Elsworth BL, et al. LD Hub: a centralized database and web  
1222 interface to perform LD score regression that maximizes the potential of summary level GWAS  
1223 data for SNP heritability and genetic correlation analysis. *Bioinformatics (Oxford, England)*  
1224 2017; **33**(2): 272-9.
- 1225 14. Bulik-Sullivan BK, Loh PR, Finucane HK, et al. LD Score regression distinguishes  
1226 confounding from polygenicity in genome-wide association studies. *Nature genetics* 2015;  
1227 **47**(3): 291-5.
- 1228 15. Ridge PG, Mukherjee S, Crane PK, Kauwe JS. Alzheimer's disease: analyzing the missing  
1229 heritability. *PloS one* 2013; **8**(11): e79771.
- 1230 16. Ridge PG, Hoyt KB, Boehme K, et al. Assessment of the genetic variance of late-onset  
1231 Alzheimer's disease. *Neurobiology of aging* 2016; **41**: 200.e13.-e20.
- 1232 17. Desikan RS, Schork AJ, Wang Y, et al. Polygenic Overlap Between C-Reactive Protein,  
1233 Plasma Lipids, and Alzheimer Disease. *Circulation* 2015; **131**(23): 2061-9.
- 1234 18. Jun GR, Chung J, Mez J, et al. Transtheortic genome-wide scan identifies novel Alzheimer's  
1235 disease loci. *Alzheimer's & dementia : the journal of the Alzheimer's Association* 2017; **13**(7):  
1236 727-38.
- 1237 19. Guerreiro R, Wojtas A, Bras J, et al. TREM2 variants in Alzheimer's disease. *The New  
1238 England journal of medicine* 2013; **368**(2): 117-27.
- 1239 20. Sims R, van der Lee SJ, Naj AC, et al. Rare coding variants in PLCG2, ABI3, and TREM2  
1240 implicate microglial-mediated innate immunity in Alzheimer's disease. *Nature genetics* 2017;  
1241 **49**(9): 1373-84.
- 1242 21. Gudbjartsson DF, Helgason H, Gudjonsson SA, et al. Large-scale whole-genome  
1243 sequencing of the Icelandic population. *Nature genetics* 2015; **47**(5): 435-44.
- 1244 22. Steinthorsdottir V, Thorleifsson G, Sulem P, et al. Identification of low-frequency and  
1245 rare sequence variants associated with elevated or reduced risk of type 2 diabetes. *Nature  
1246 genetics* 2014; **46**(3): 294-8.
- 1247 23. Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software. *Bioinformatics  
1248 (Oxford, England)* 2015; **31**(9): 1466-8.
- 1249 24. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework  
1250 for estimating the relative pathogenicity of human genetic variants. *Nature genetics* 2014;  
1251 **46**(3): 310-5.
- 1252 25. Finucane HK, Bulik-Sullivan B, Gusev A, et al. Partitioning heritability by functional  
1253 annotation using genome-wide association summary statistics. *Nature genetics* 2015; **47**(11):  
1254 1228-35.
- 1255 26. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and  
1256 annotation of genetic associations with FUMA. *Nature communications* 2017; **8**(1): 1826.
- 1257 27. Gurses MS, Ural MN, Gulec MA, Akyol O, Akyol S. Pathophysiological Function of  
1258 ADAMTS Enzymes on Molecular Mechanism of Alzheimer's Disease. *Aging and disease* 2016;  
1259 **7**(4): 479-90.
- 1260 28. Suh J, Choi SH, Romano DM, et al. ADAM10 missense mutations potentiate beta-  
1261 amyloid accumulation by impairing prodomain chaperone function. *Neuron* 2013; **80**(2): 385-  
1262 401.
- 1263 29. Dries DR, Yu G. Assembly, maturation, and trafficking of the gamma-secretase complex  
1264 in Alzheimer's disease. *Current Alzheimer research* 2008; **5**(2): 132-46.

## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

- 1265 30. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis  
1266 of GWAS data. *PLoS computational biology* 2015; **11**(4): e1004219.
- 1267 31. Zhu Z, Zheng Z, Zhang F, et al. Causal associations between risk factors and common  
1268 diseases inferred from GWAS summary data. *Nature communications* 2018; **9**(1): 224.
- 1269 32. Skene NG, Grant SG. Identification of Vulnerable Cell Types in Major Brain Disorders  
1270 Using Single Cell Transcriptomes and Expression Weighted Cell Type Enrichment. *Frontiers in*  
1271 *neuroscience* 2016; **10**: 16.
- 1272 33. Loewendorf A, Fonteh A, Mg H, Me C. Inflammation in Alzheimer's Disease: Cross-talk  
1273 between Lipids and Innate Immune Cells of the Brain; 2015.
- 1274 34. Stern Y. Cognitive reserve in ageing and Alzheimer's disease. *The Lancet Neurology*  
1275 2012; **11**(11): 1006-12.
- 1276 35. Satizabal C, Beiser AS, Seshadri S. Incidence of Dementia over Three Decades in the  
1277 Framingham Heart Study. *The New England journal of medicine* 2016; **375**(1): 93-4.
- 1278 36. Magnusson PK, Almqvist C, Rahman I, et al. The Swedish Twin Registry: establishment of  
1279 a biobank and other recent developments. *Twin research and human genetics : the official*  
1280 *journal of the International Society for Twin Studies* 2013; **16**(1): 317-29.
- 1281 37. Finkel D, Pedersen NL. Processing Speed and Longitudinal Trajectories of Change for  
1282 Cognitive Abilities: The Swedish Adoption/Twin Study of Aging. *Aging, Neuropsychology, and*  
1283 *Cognition* 2004; **11**(2-3): 325-45.
- 1284 38. Gold CH, Malmberg B, McClearn GE, Pedersen NL, Berg S. Gender and health: a study of  
1285 older unlike-sex twins. *The journals of gerontology Series B, Psychological sciences and social*  
1286 *sciences* 2002; **57**(3): S168-76.
- 1287 39. Gatz M, Fratiglioni L, Johansson B, et al. Complete ascertainment of dementia in the  
1288 Swedish Twin Registry: the HARMONY study. *Neurobiology of aging* 2005; **26**(4): 439-47.
- 1289 40. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis  
1290 of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of  
1291 Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* 1984;  
1292 **34**(7): 939-44.
- 1293 41. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK:  
1294 rising to the challenge of larger and richer datasets. *GigaScience* 2015; **4**: 7.
- 1295 42. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying  
1296 the causes of a wide range of complex diseases of middle and old age. *PLoS medicine* 2015;  
1297 **12**(3): e1001779.
- 1298 43. Hebert LE, Weuve J, Scherr PA, Evans DA. Alzheimer disease in the United States (2010-  
1299 2050) estimated using the 2010 census. *Neurology* 2013; **80**(19): 1778-83.
- 1300 44. The Genomes Project C. A global reference for human genetic variation. *Nature* 2015;  
1301 **526**: 68.
- 1302 45. Davies G, Marioni RE, Liewald DC, et al. Genome-wide association study of cognitive  
1303 functions and educational attainment in UK Biobank (N=112 151). *Molecular psychiatry* 2016;  
1304 **21**(6): 758-67.
- 1305 46. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for  
1306 genotype imputation. *Nature genetics* 2016; **48**(10): 1279-83.

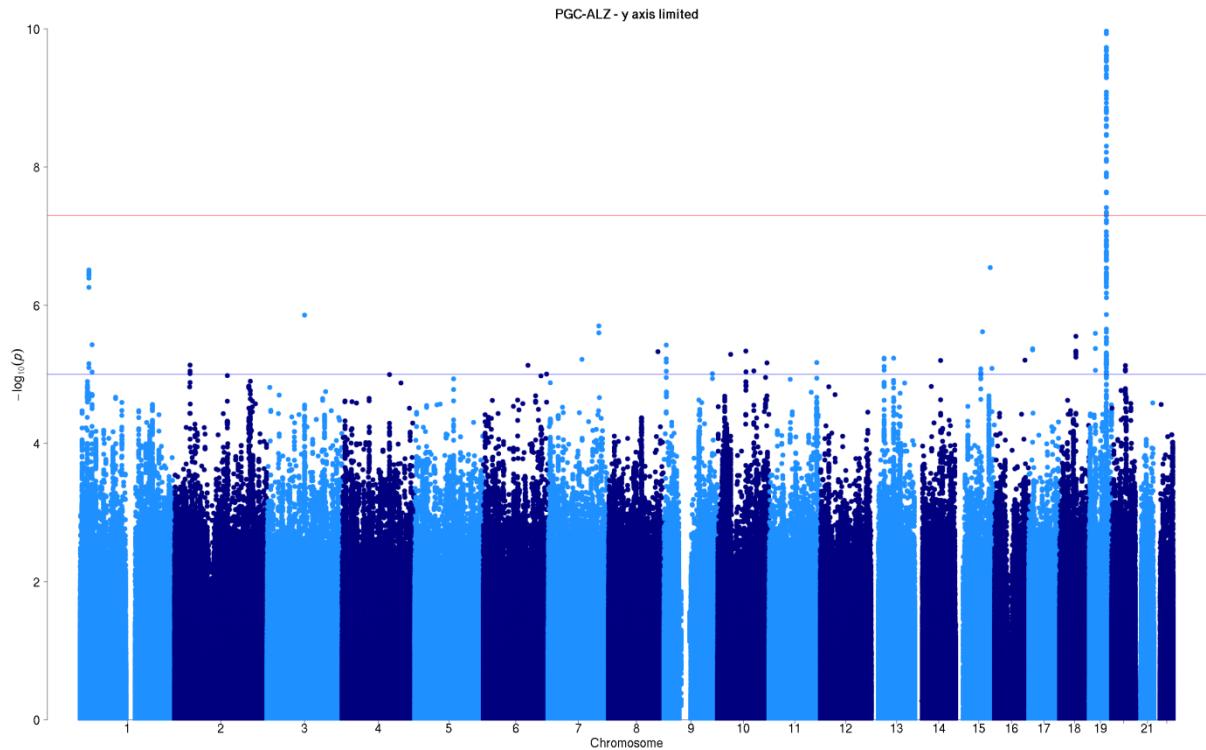
## GWAS META-ANALYSIS OF ALZHEIMER'S DISEASE RISK

- 1307 47. Peterson RE, Edwards AC, Bacanu SA, Dick DM, Kendler KS, Webb BT. The utility of  
1308 empirically assigning ancestry groups in cross-population genetic studies of addiction. *The  
1309 American journal on addictions* 2017; **26**(5): 494-501.
- 1310 48. Abraham G, Qiu Y, Inouye M. FlashPCA2: principal component analysis of Biobank-scale  
1311 genotype datasets. *Bioinformatics (Oxford, England)* 2017; **33**(17): 2776-8.
- 1312 49. Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across  
1313 human diseases and traits. *Nature genetics* 2015; **47**(11): 1236-41.
- 1314 50. Lovestone S, Francis P, Kloszewska I, et al. AddNeuroMed--the European collaboration  
1315 for the discovery of novel biomarkers for Alzheimer's disease. *Annals of the New York Academy  
1316 of Sciences* 2009; **1180**: 36-46.
- 1317 51. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from  
1318 high-throughput sequencing data. *Nucleic acids research* 2010; **38**(16): e164.
- 1319 52. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal  
1320 genomes using RegulomeDB. *Genome research* 2012; **22**(9): 1790-7.
- 1321 53. Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and  
1322 characterization. *Nature methods* 2012; **9**(3): 215-6.
- 1323 54. Kundaje A, Meuleman W, Ernst J, et al. Integrative analysis of 111 reference human  
1324 epigenomes. *Nature* 2015; **518**(7539): 317-30.
- 1325 55. Amendola LM, Dorschner MO, Robertson PD, et al. Actionable exomic incidental findings  
1326 in 6503 participants: challenges of variant classification. *Genome research* 2015; **25**(3): 305-15.
- 1327 56. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue  
1328 gene regulation in humans. *Science (New York, NY)* 2015; **348**(6235): 648-60.
- 1329 57. Westra HJ, Peters MJ, Esko T, et al. Systematic identification of trans eQTLs as putative  
1330 drivers of known disease associations. *Nature genetics* 2013; **45**(10): 1238-43.
- 1331 58. Zhernakova DV, Deelen P, Vermaat M, et al. Identification of context-dependent  
1332 expression quantitative trait loci in whole blood. *Nature genetics* 2017; **49**(1): 139-45.
- 1333 59. Schmitt AD, Hu M, Jung I, et al. A Compendium of Chromatin Contact Maps Reveals  
1334 Spatially Active Regions in the Human Genome. *Cell reports* 2016; **17**(8): 2042-59.
- 1335 60. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a  
1336 knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of  
1337 the National Academy of Sciences of the United States of America* 2005; **102**(43): 15545-50.
- 1338 61. Habib N, Avraham-David I, Basu A, et al. Massively parallel single-nucleus RNA-seq with  
1339 DroNc-seq. *Nature methods* 2017; **14**(10): 955-8.
- 1340

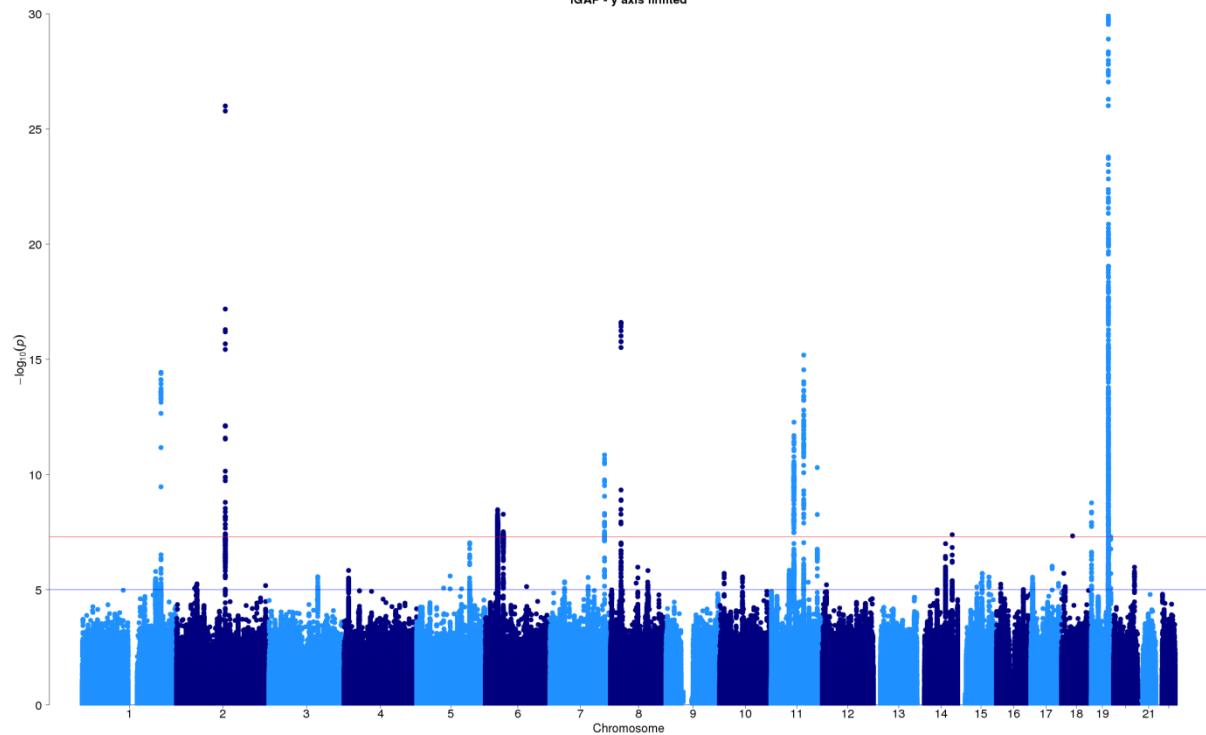
It is made available under a CC-BY-NC-ND 4.0 International license.

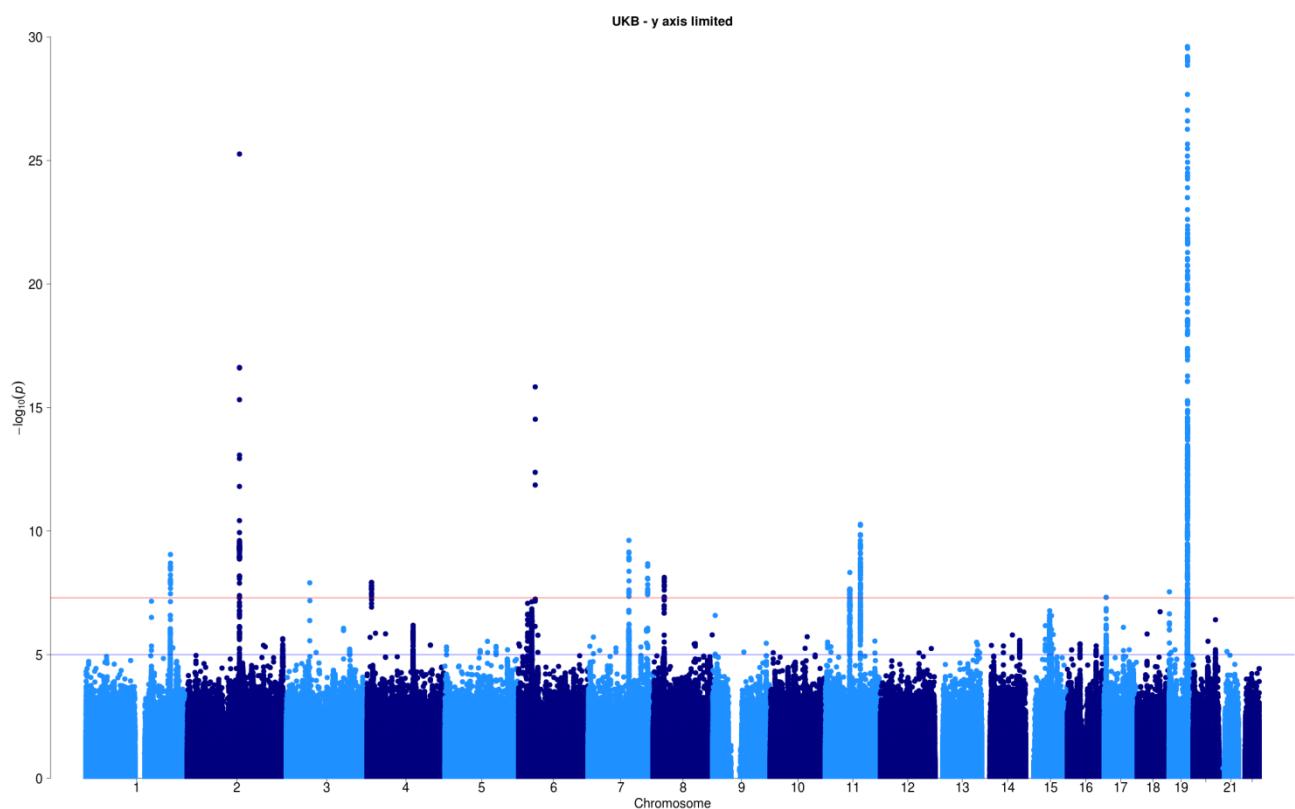
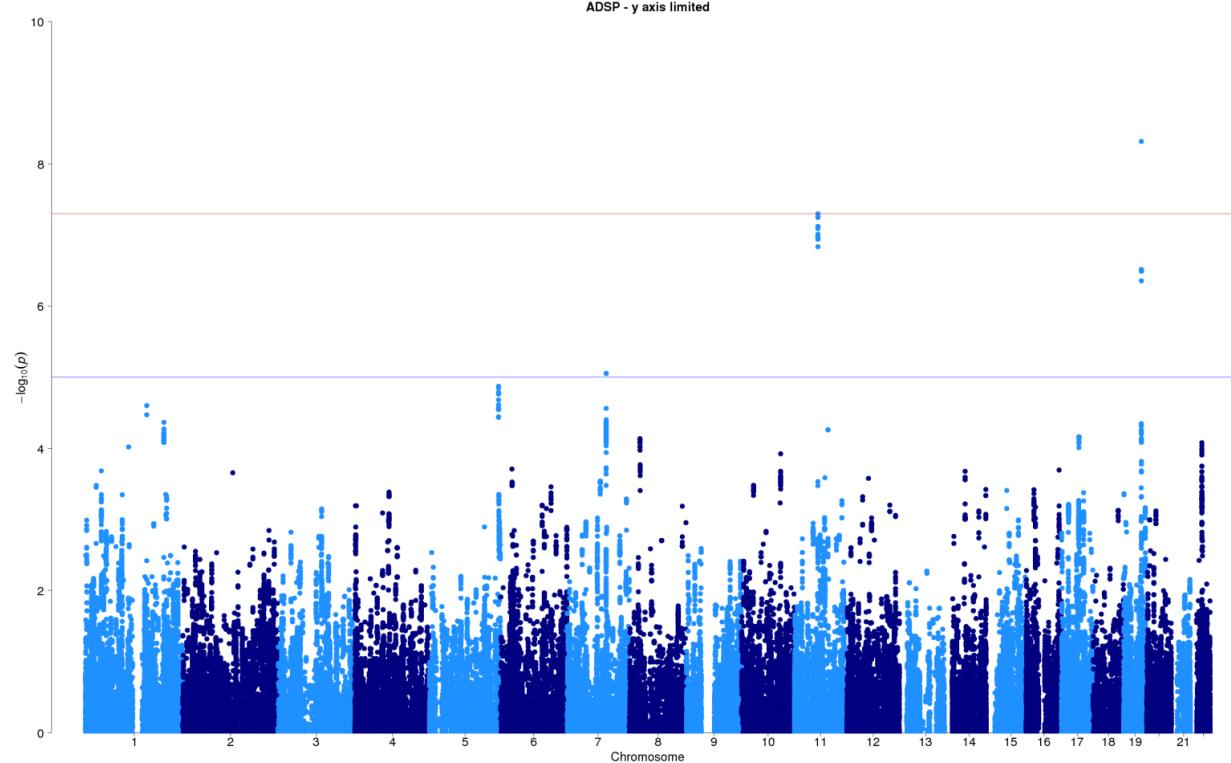
**Supplementary Figure 1. Manhattan and QQ plots of single variant association results per main cohort.** For each cohort, Manhattan and QQ plots are shown. A) The Manhattan plot displays all associations per variant ordered according to their genomic position on the x-axis and showing the strength of the association with the  $-\log_{10}$  transformed  $P$ -values on the y-axis. The y-axis is limited to enable visualization of non-*APOE* loci. B) The QQ plot displays the expected  $-\log_{10}$  transformed  $p$ -values on the x-axis and the observed  $-\log_{10}$  transformed  $p$ -values on the y-axis.

**A**

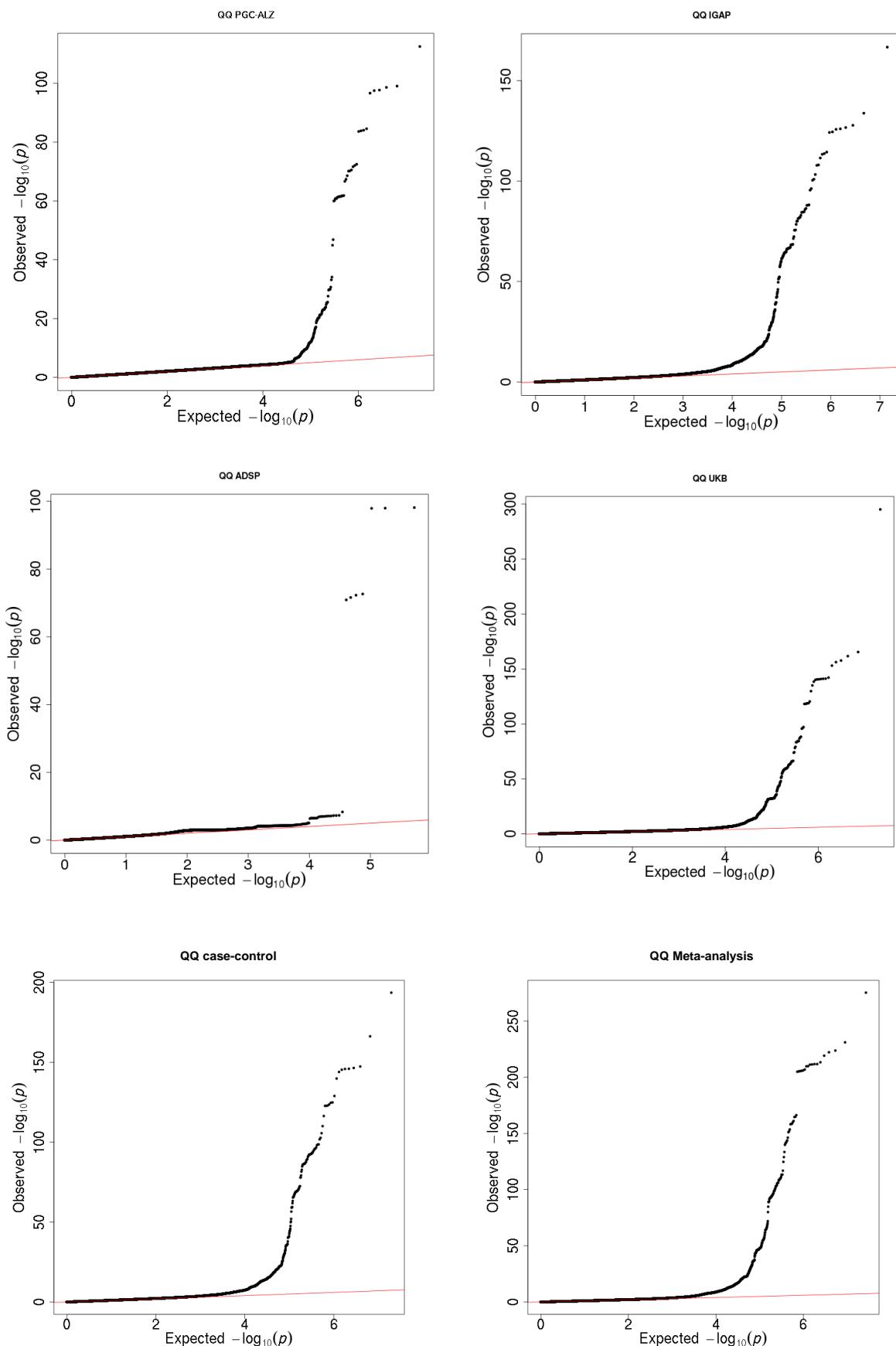


IGAP - y axis limited

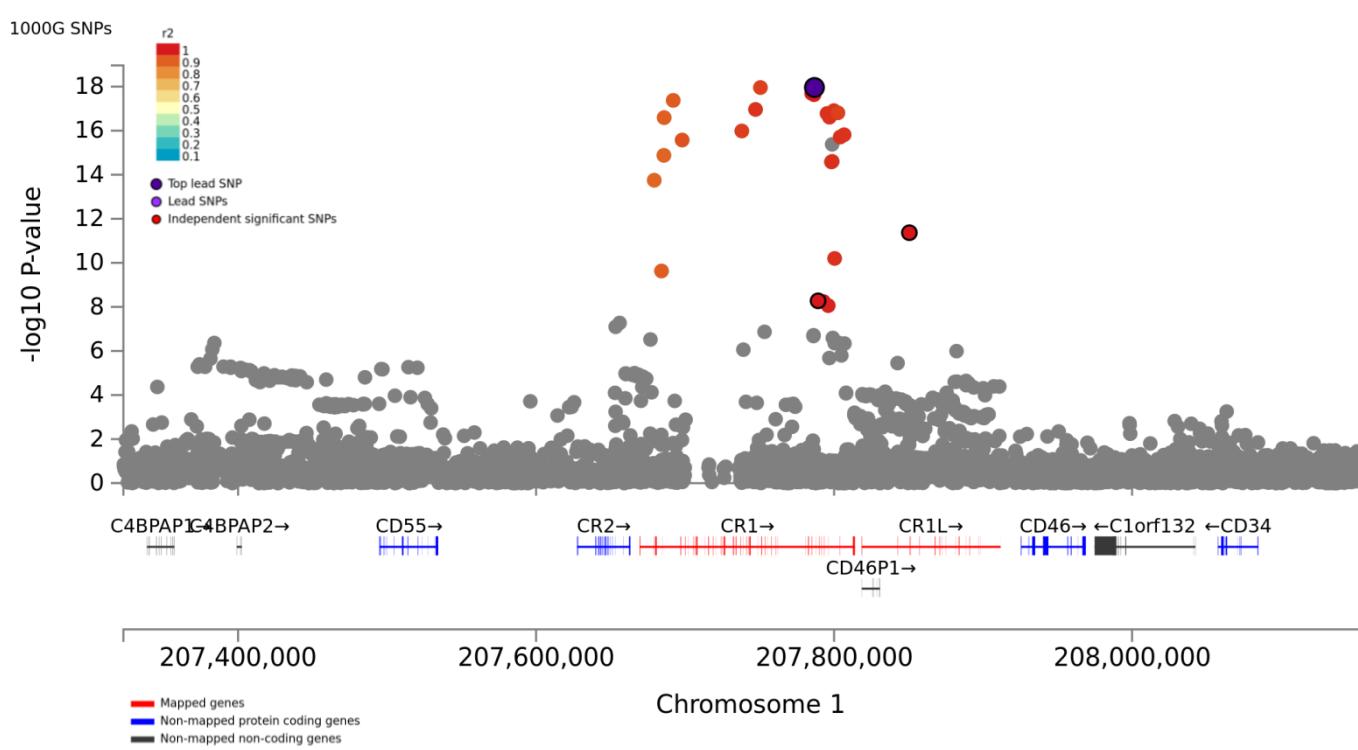
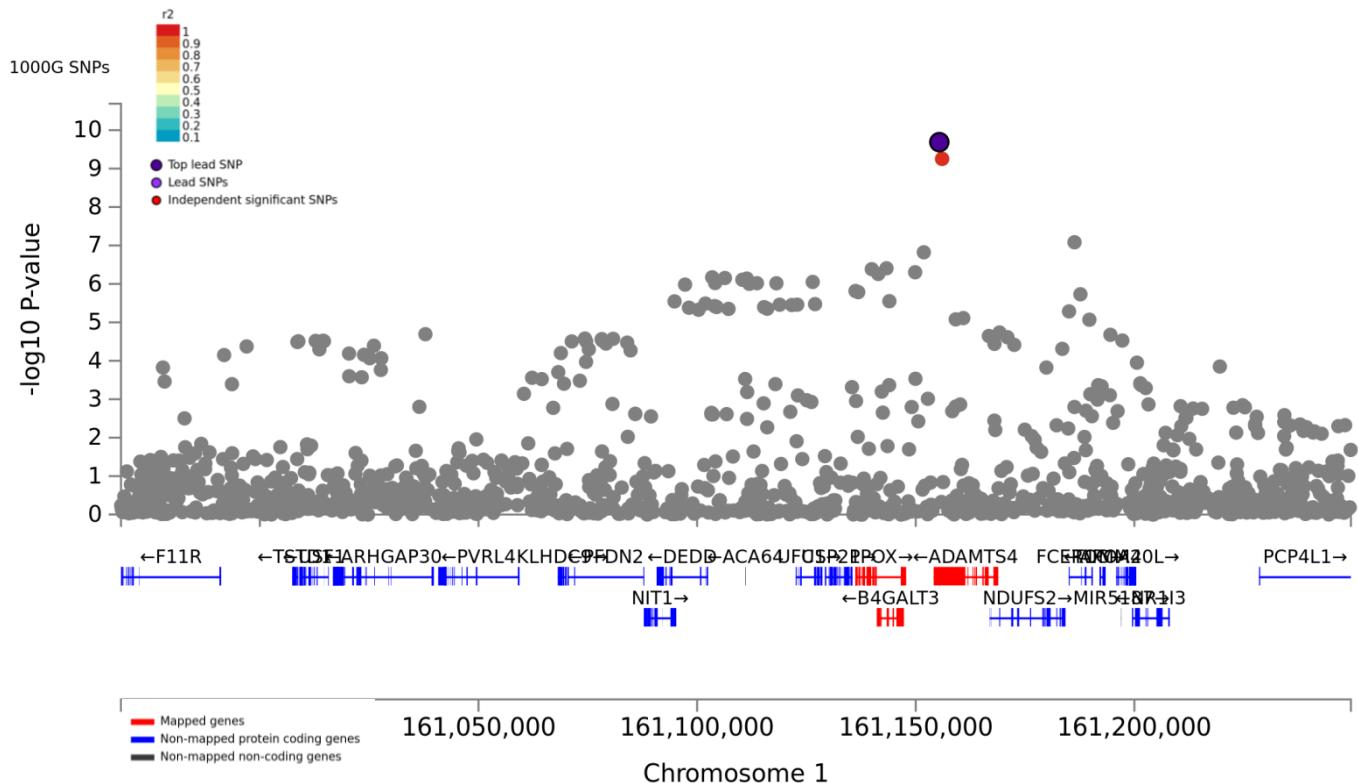


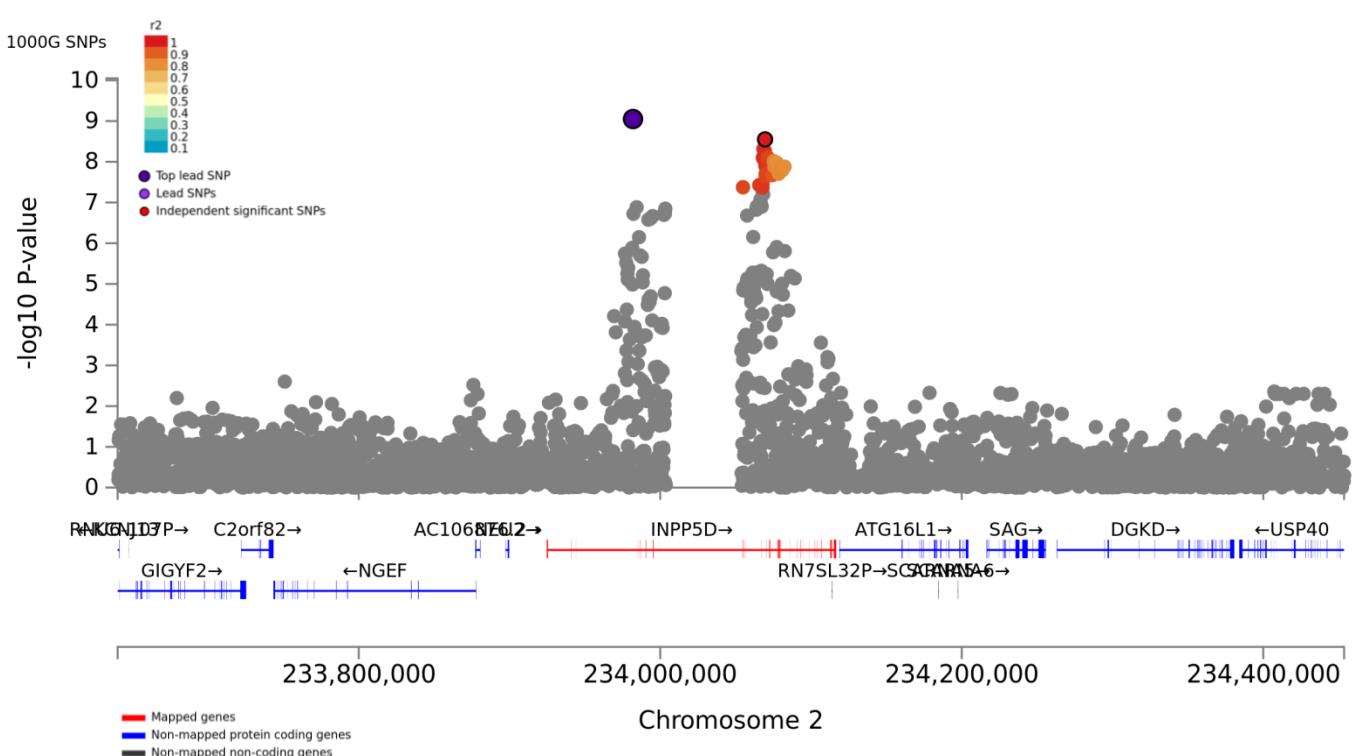
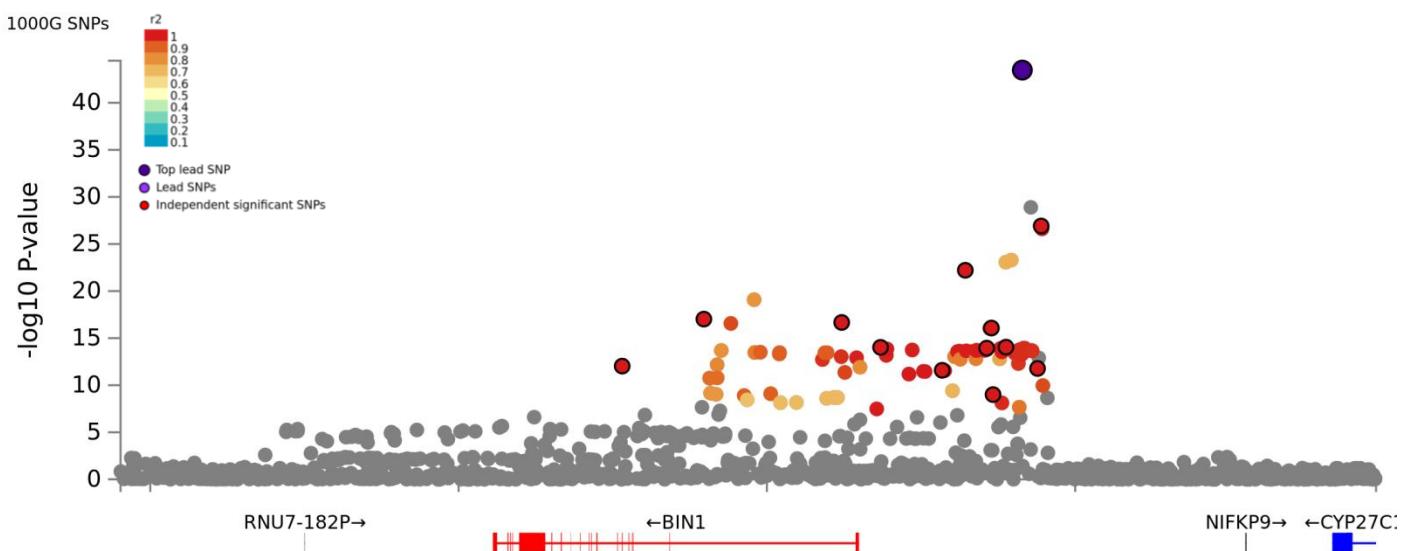


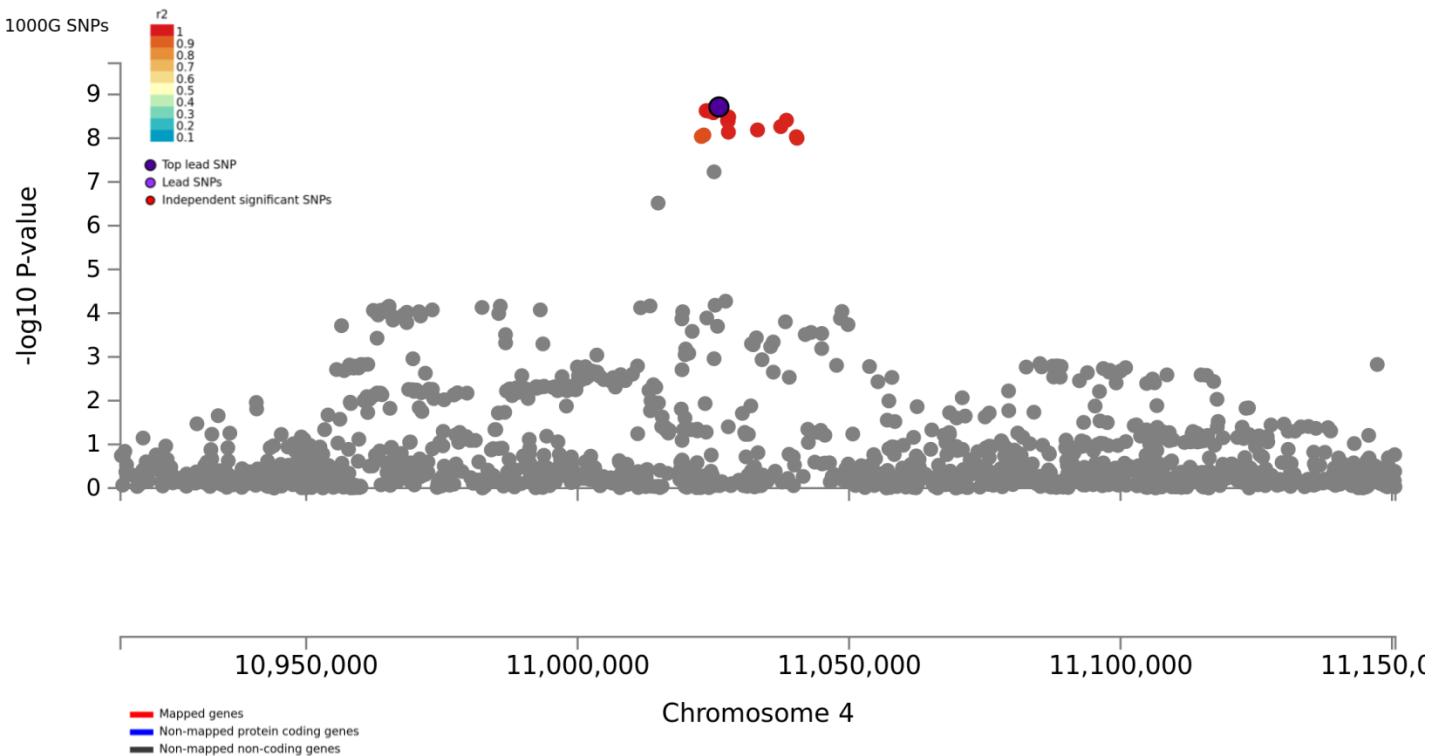
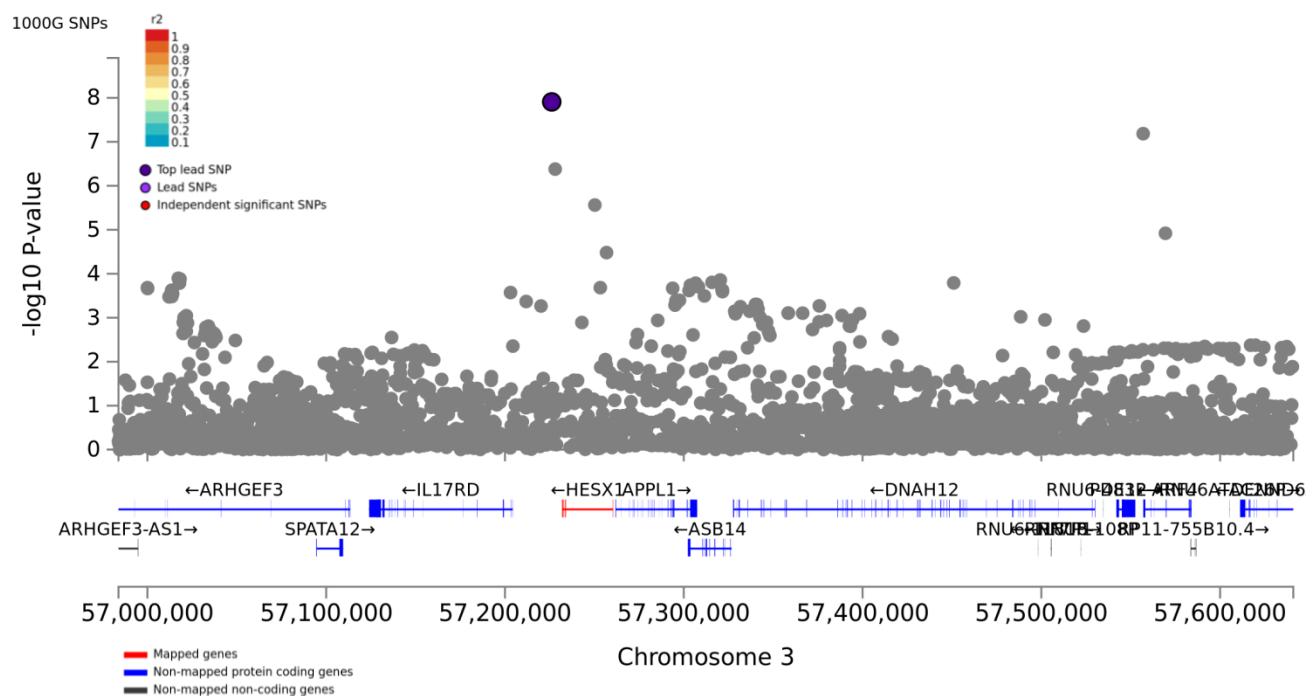
B

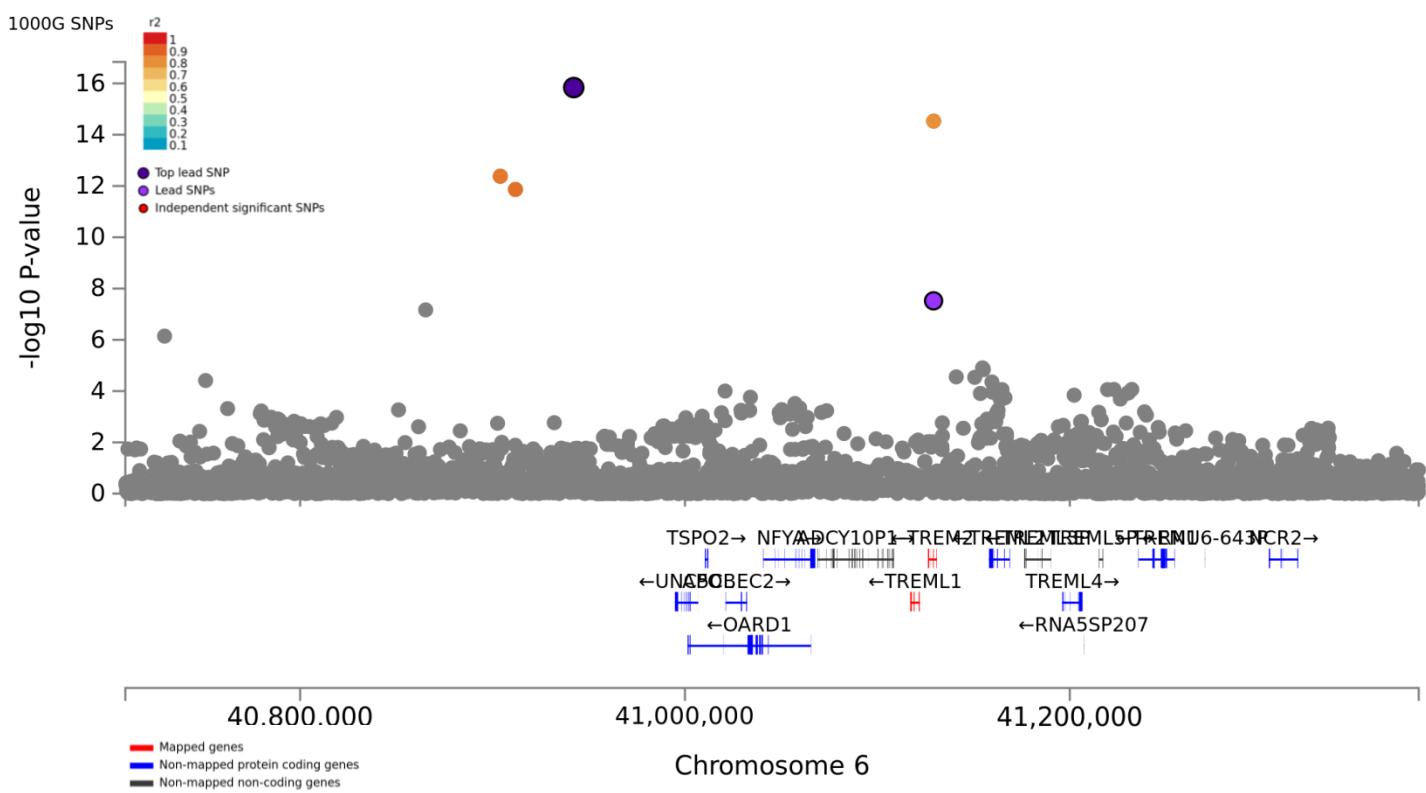
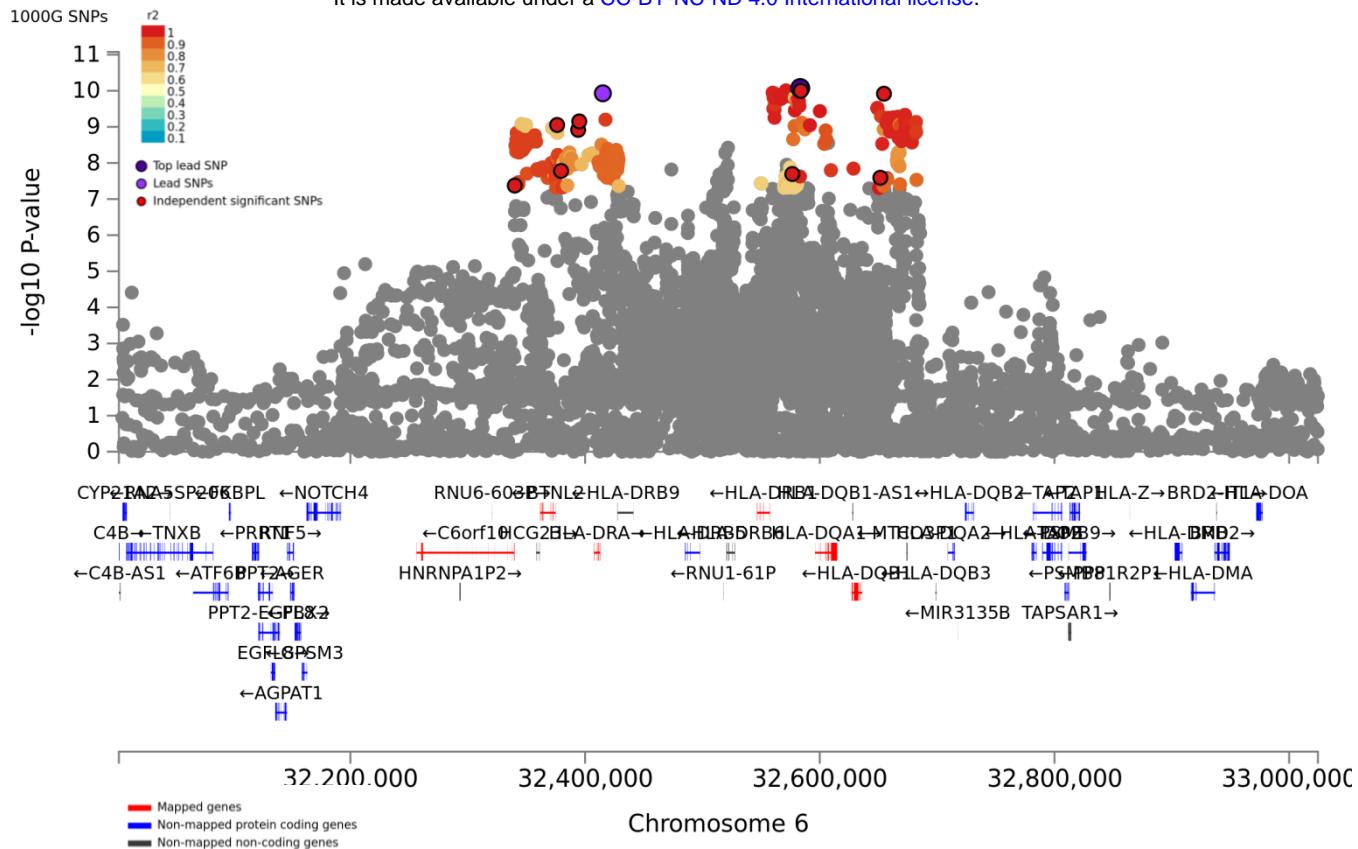


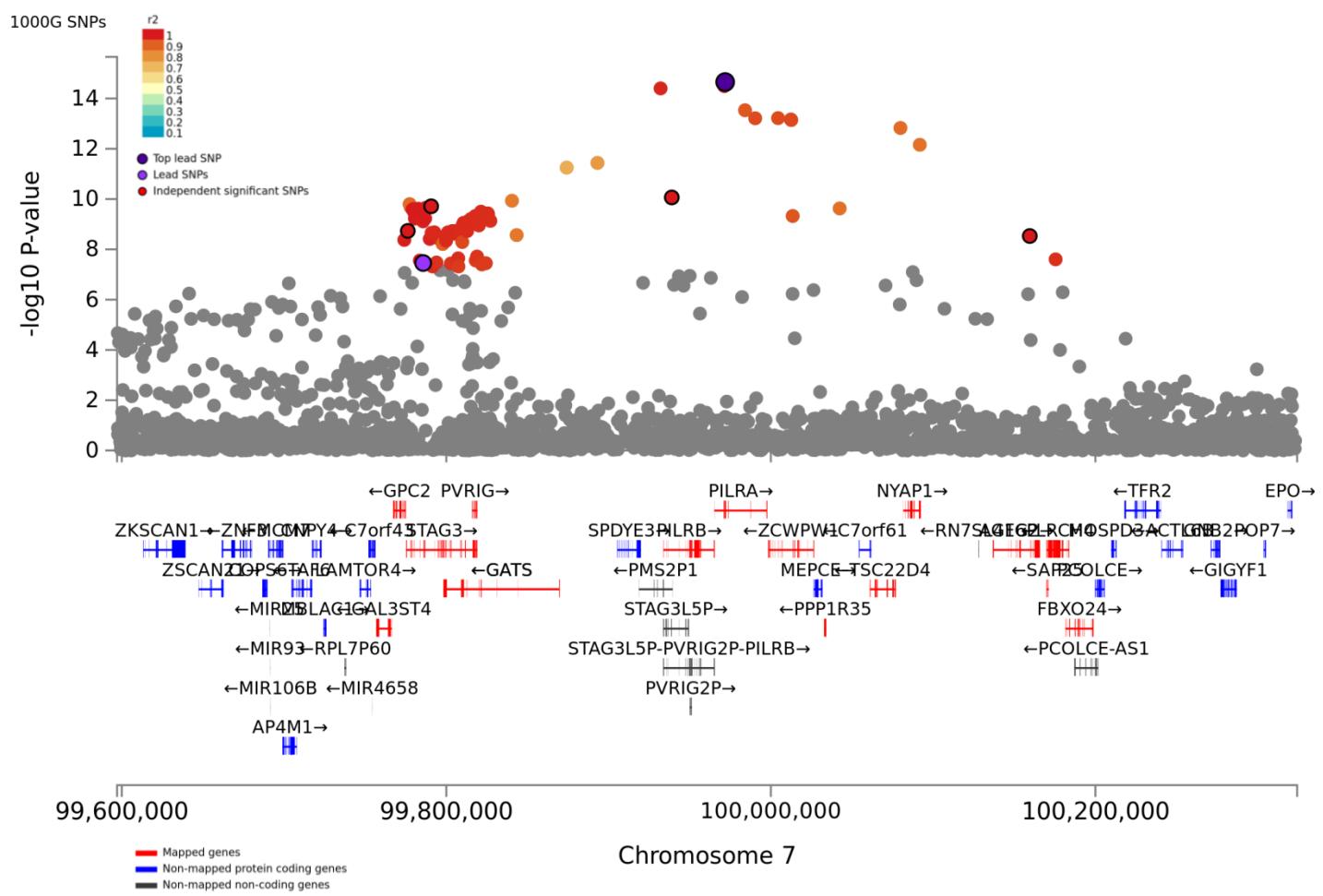
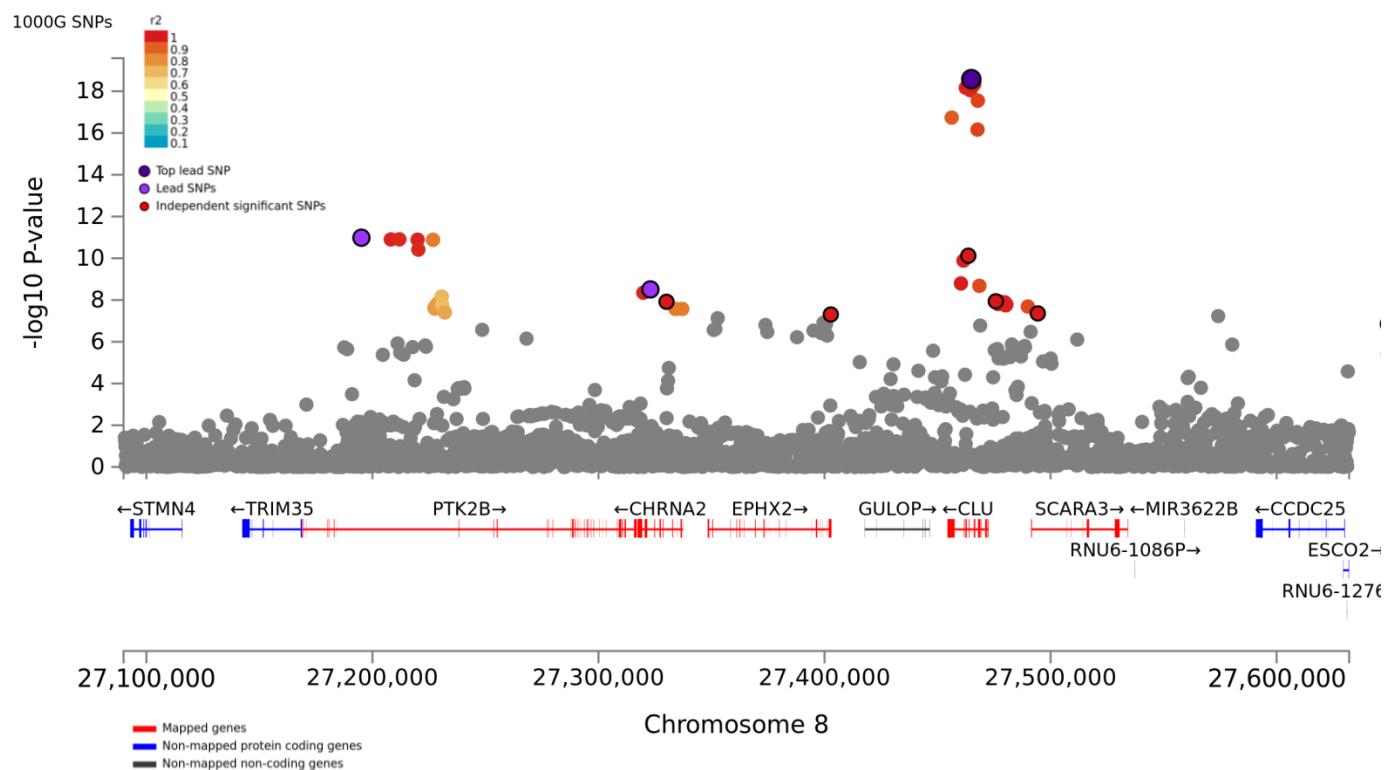
**Supplementary Figure 2. Regional plot for the 29 significant loci of the meta-analysis.** Every point represents a SNP, which are colour-coded based on the highest  $r^2$  to one of the most significant SNPs, if greater or equal to  $r^2$  of 0.6. Other SNPs are coloured in grey.

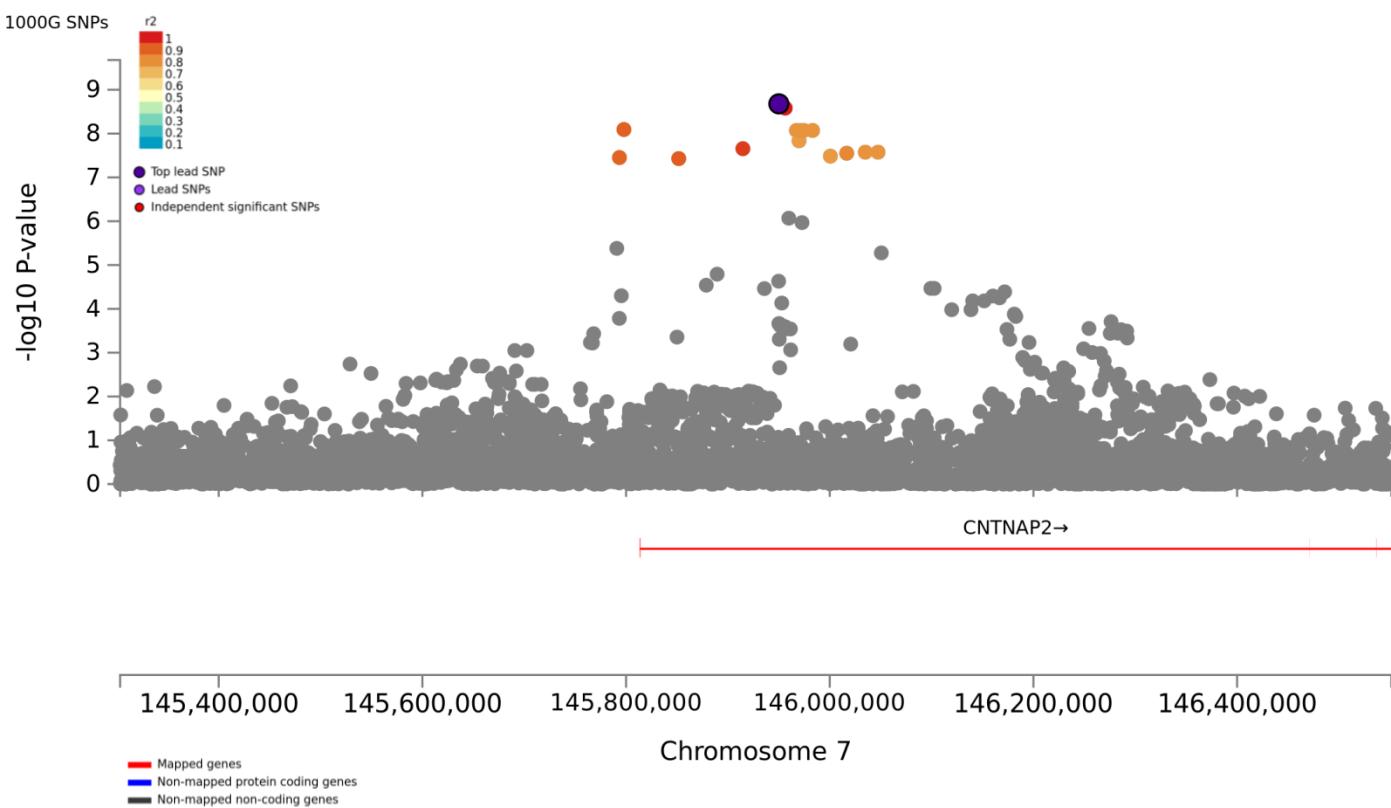
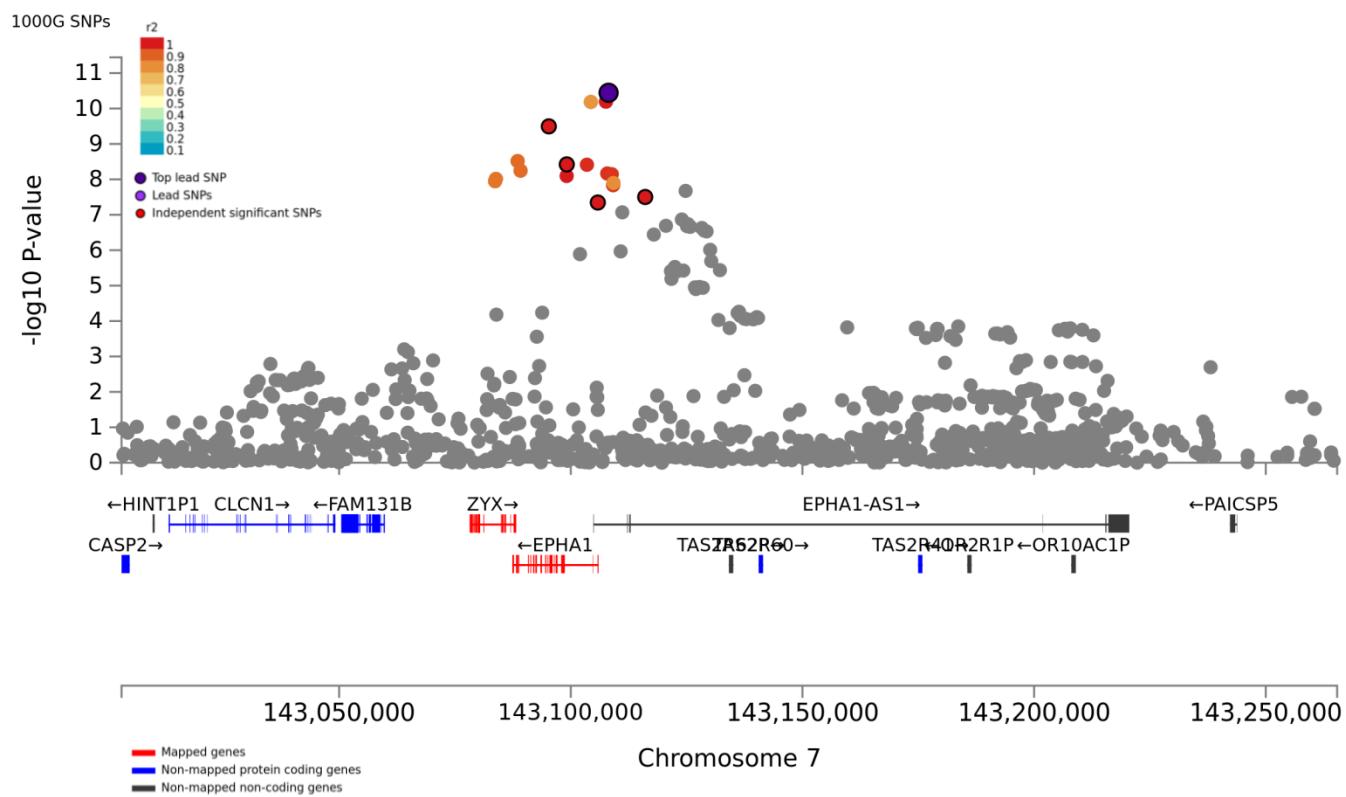


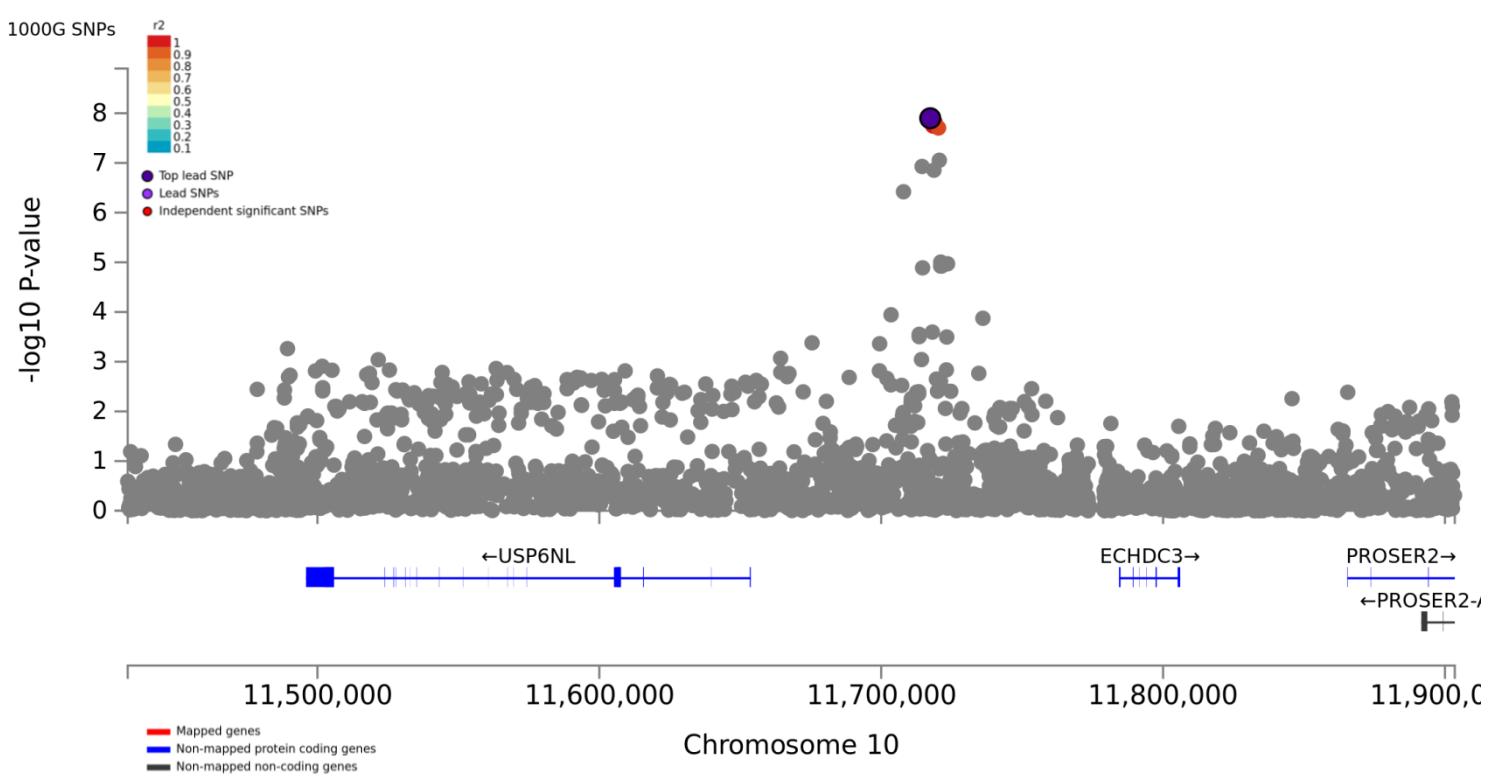
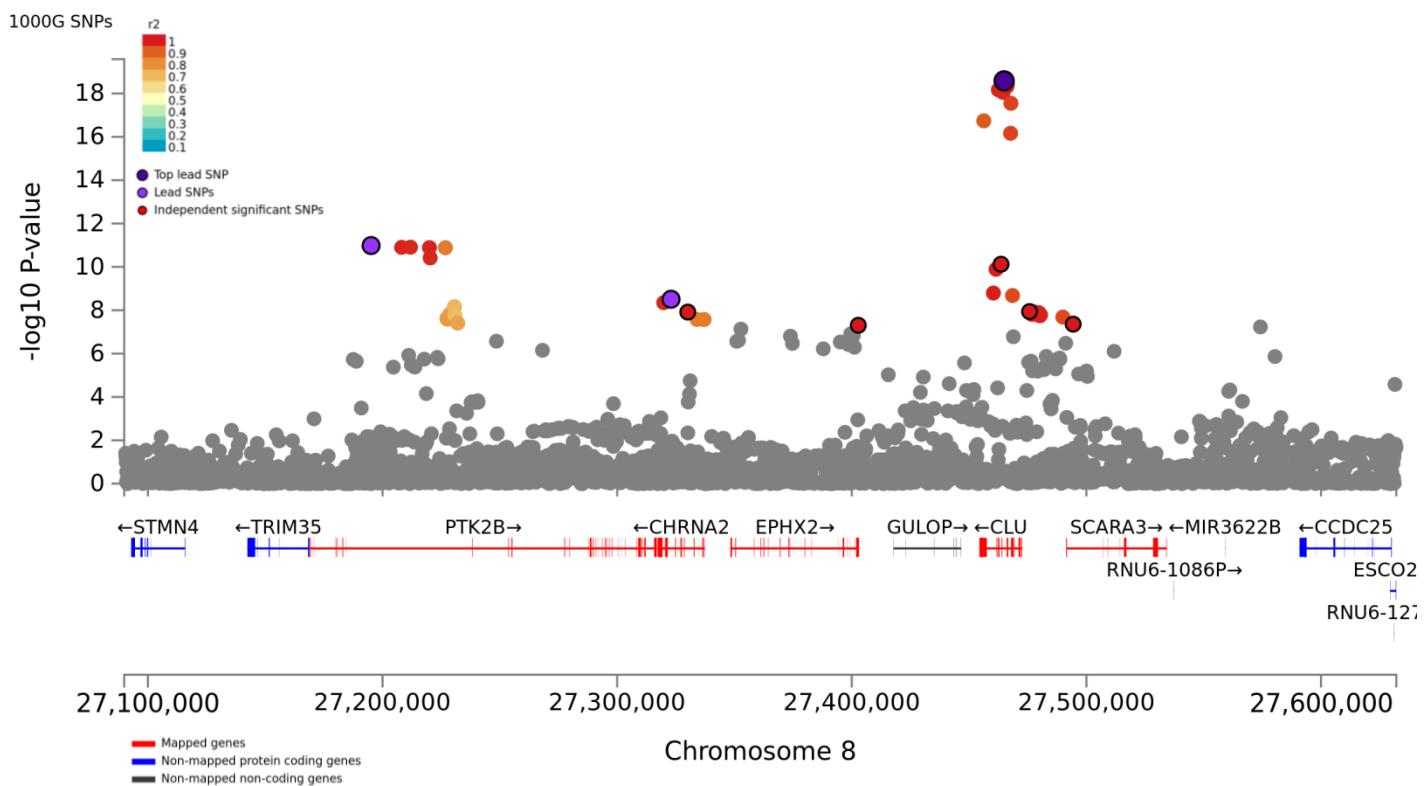


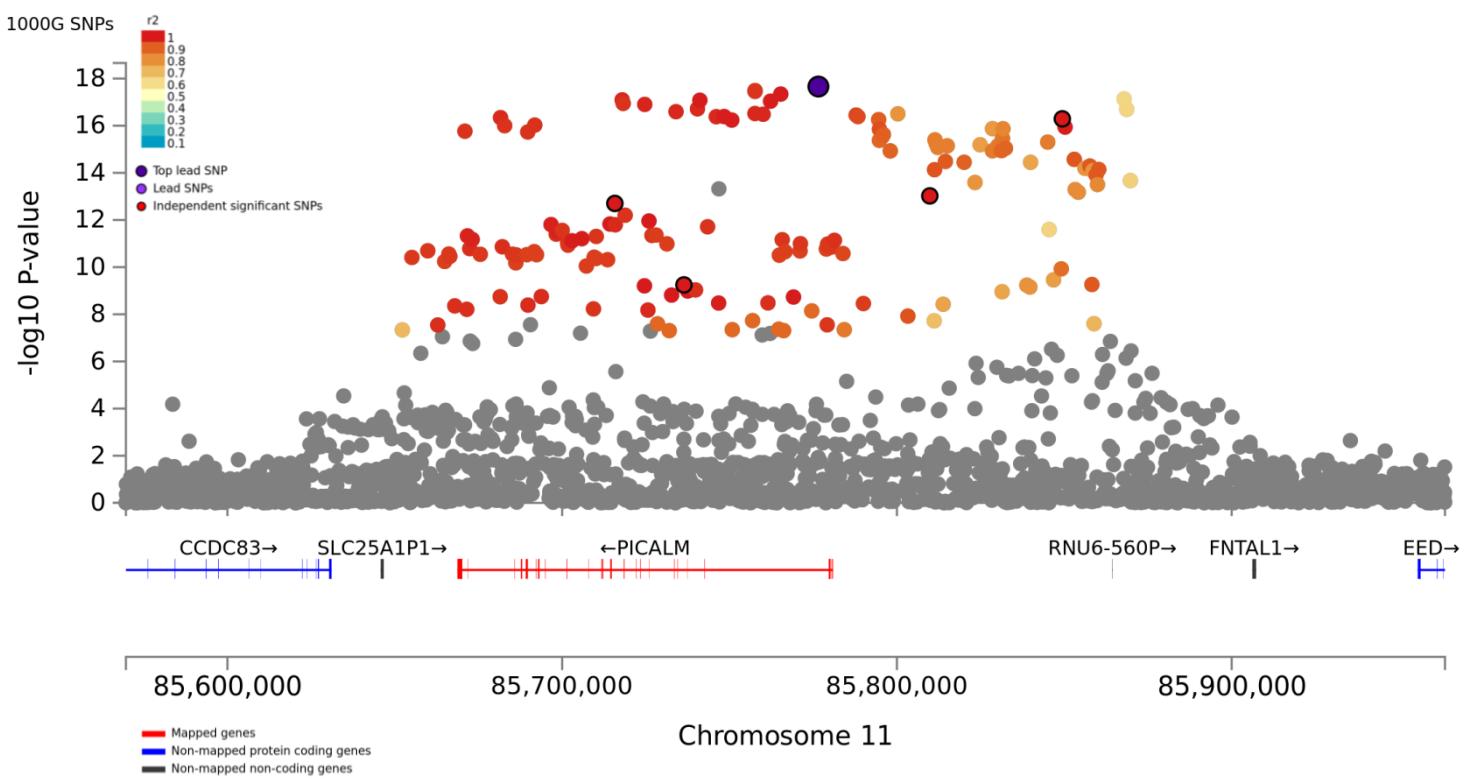
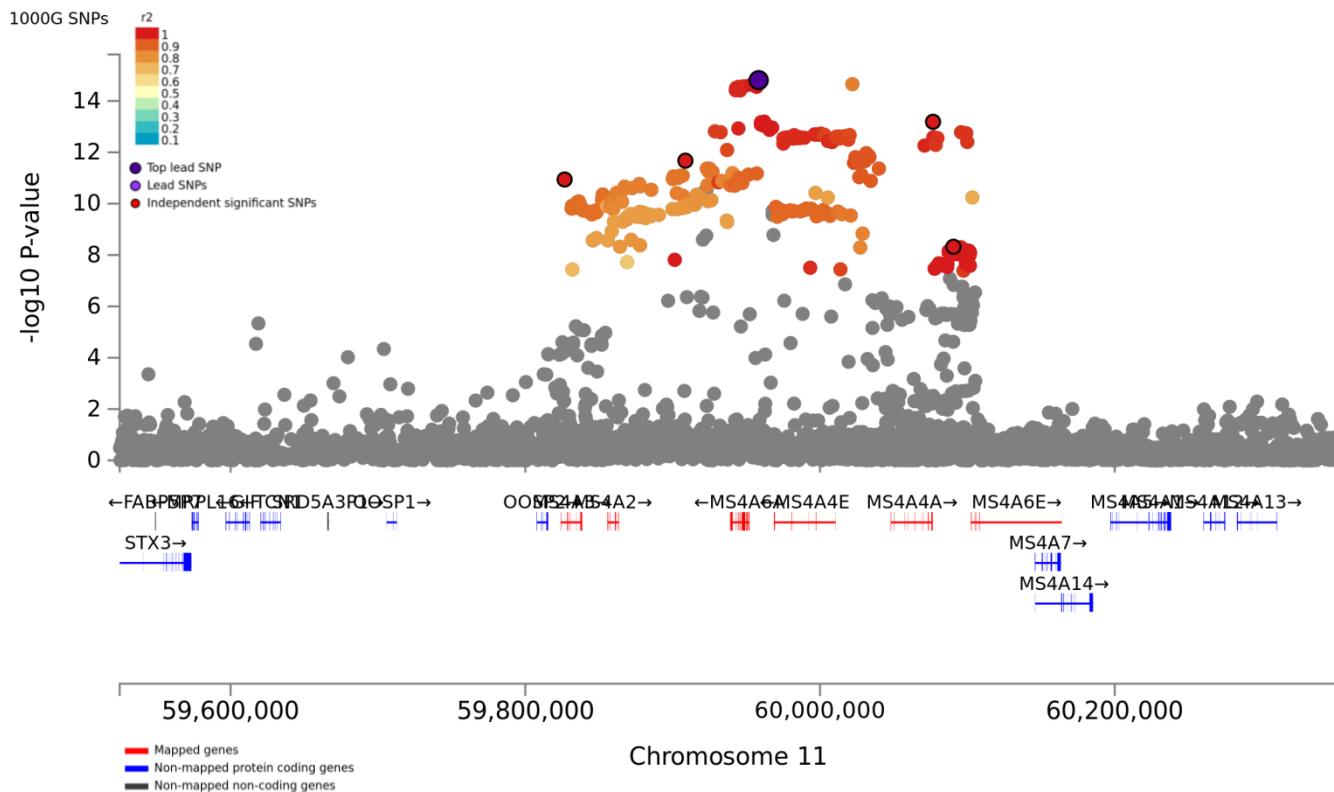


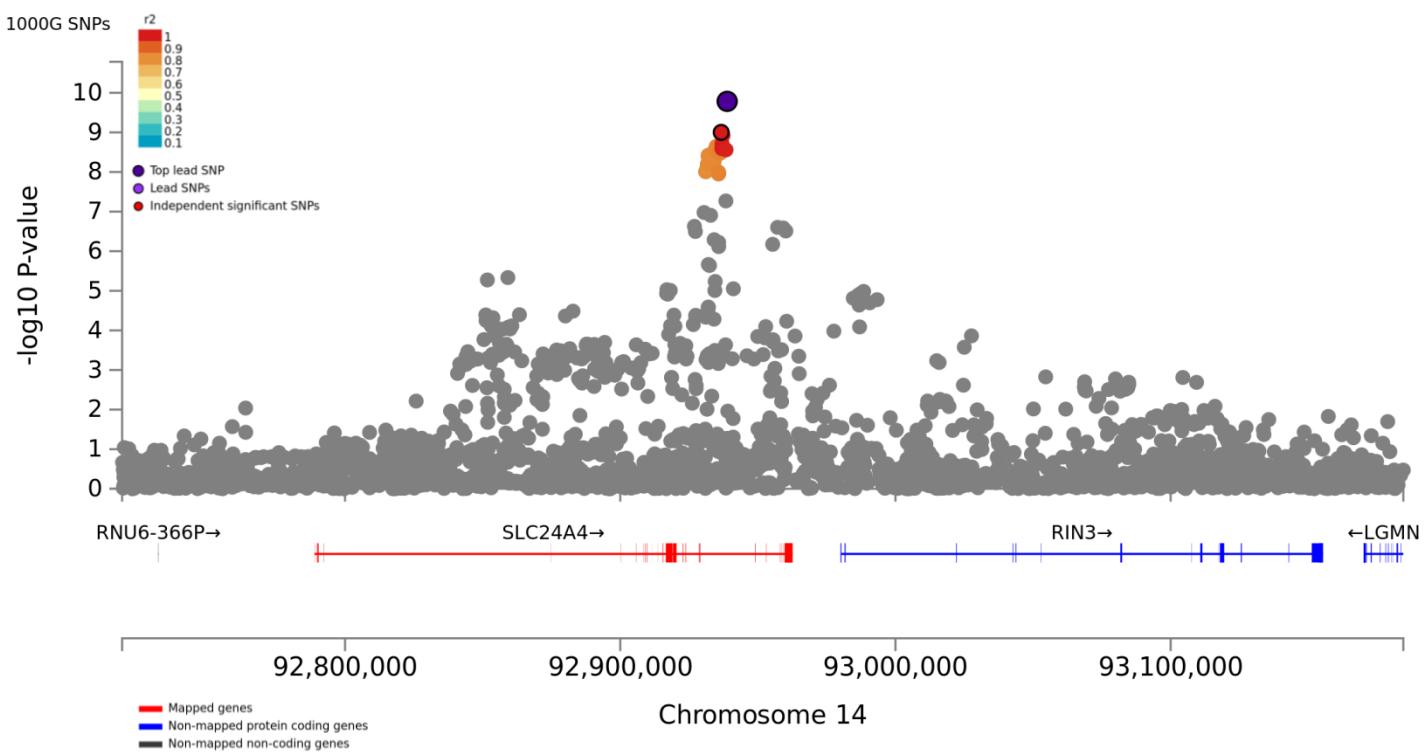
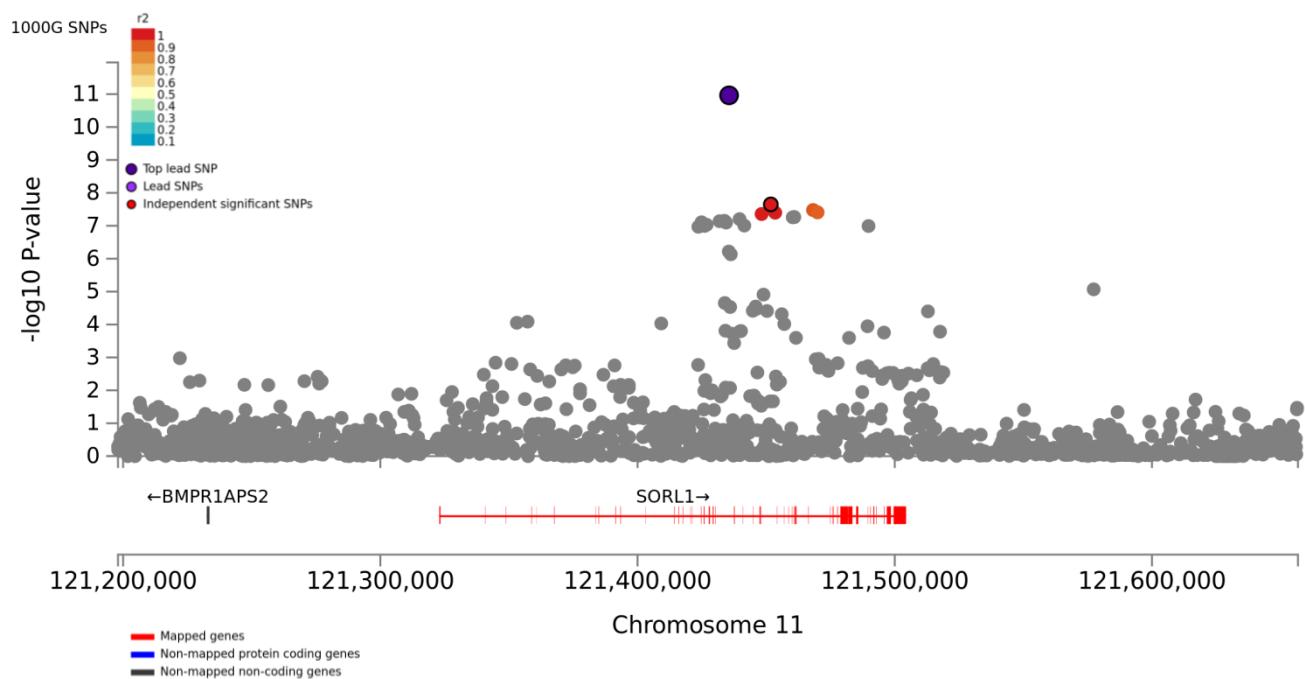


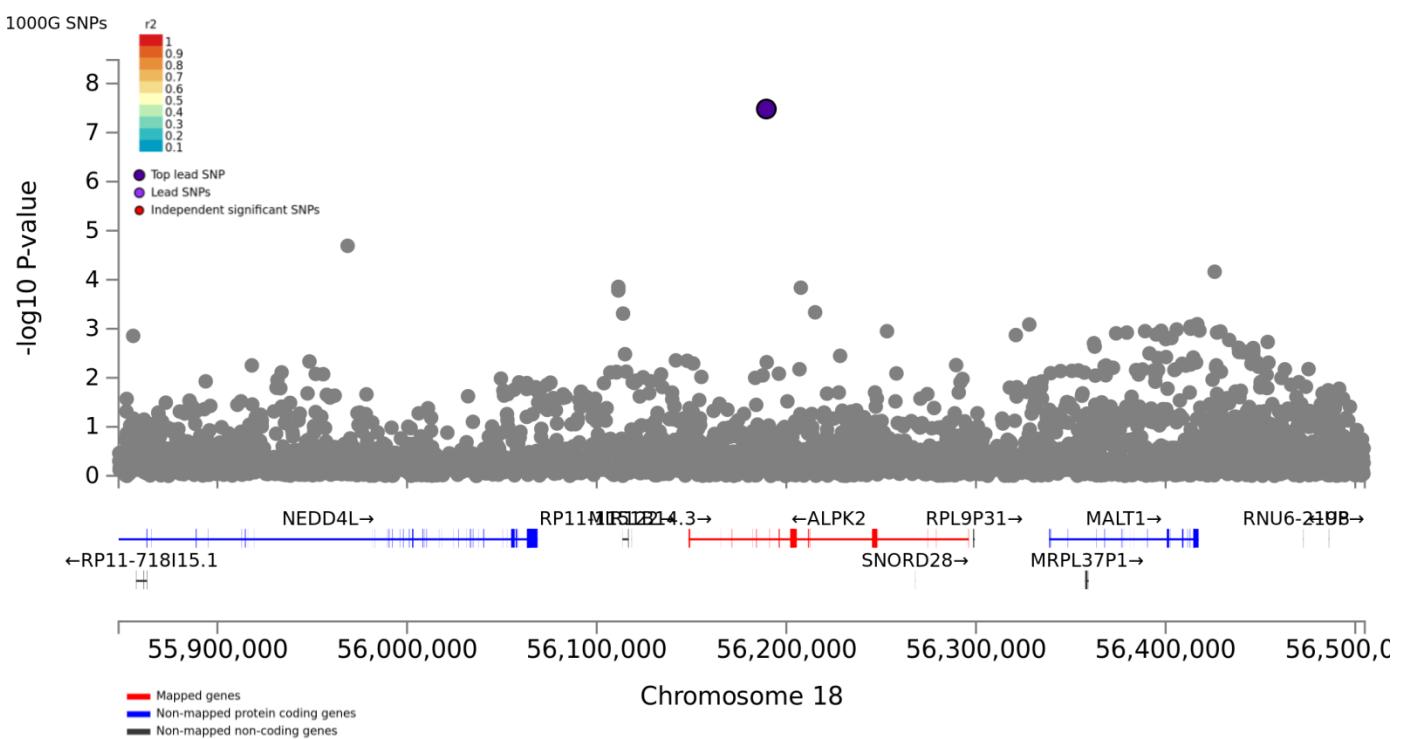
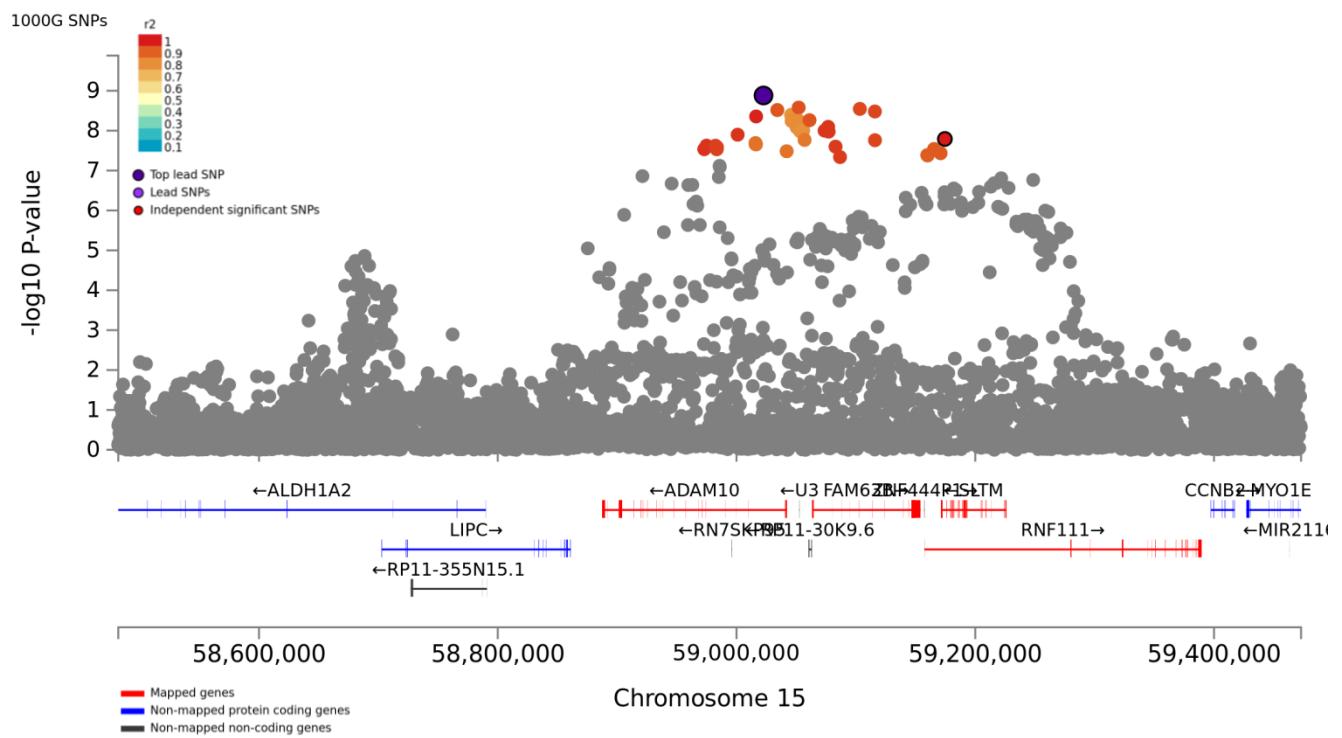


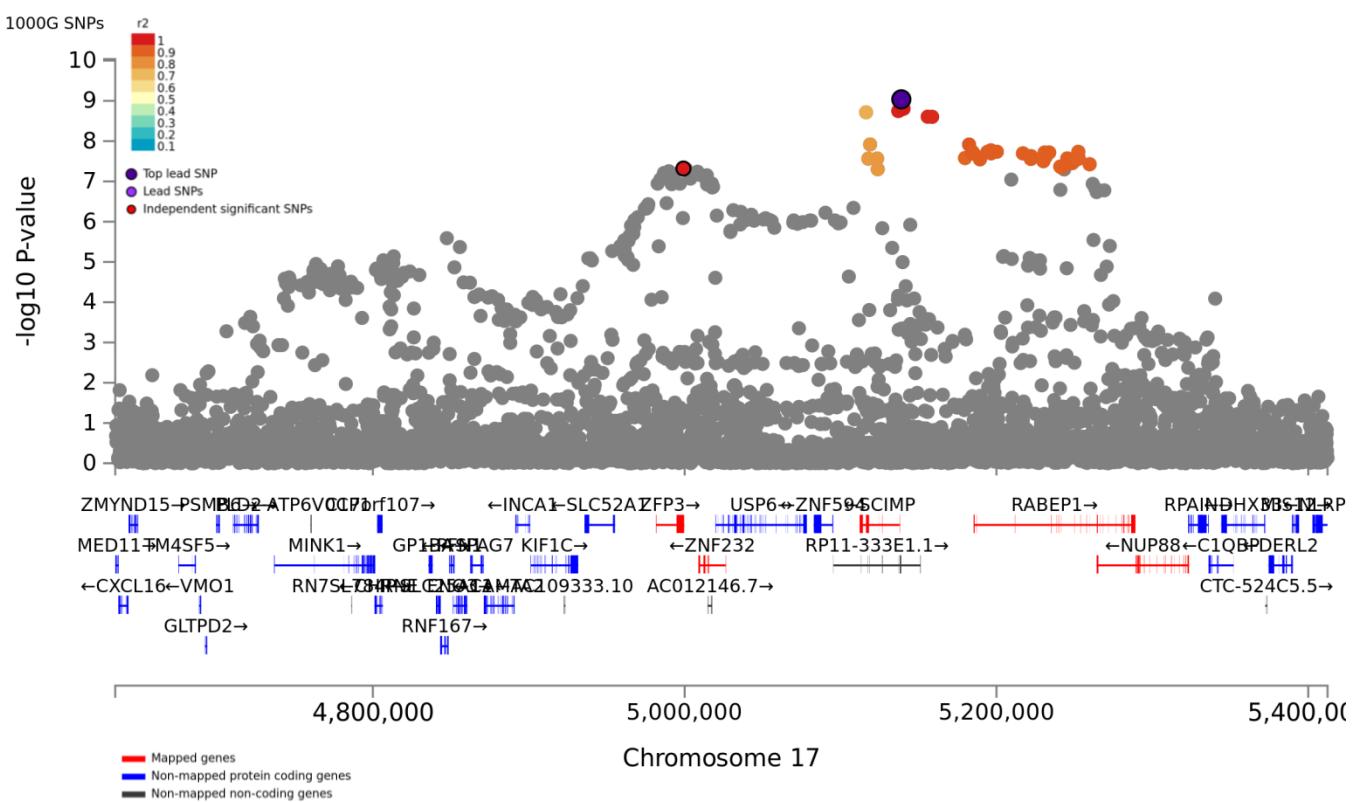
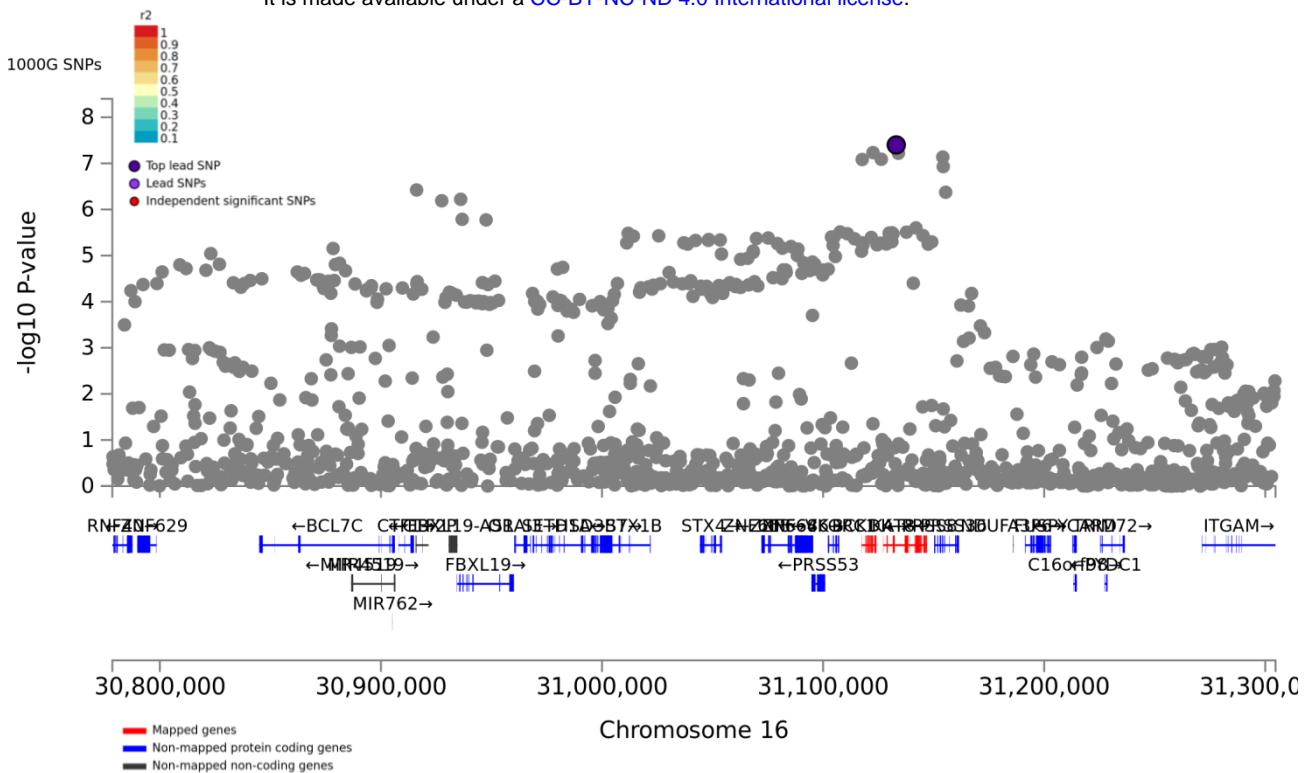


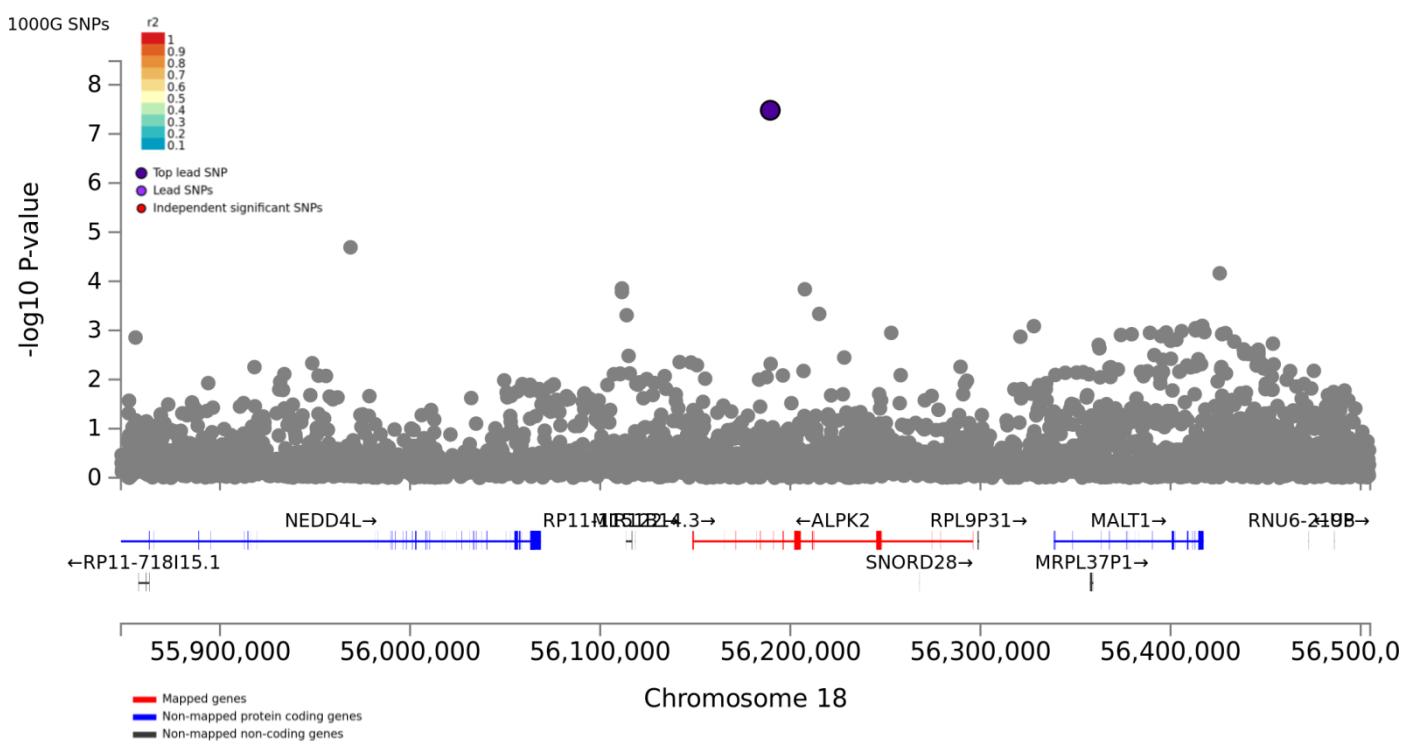
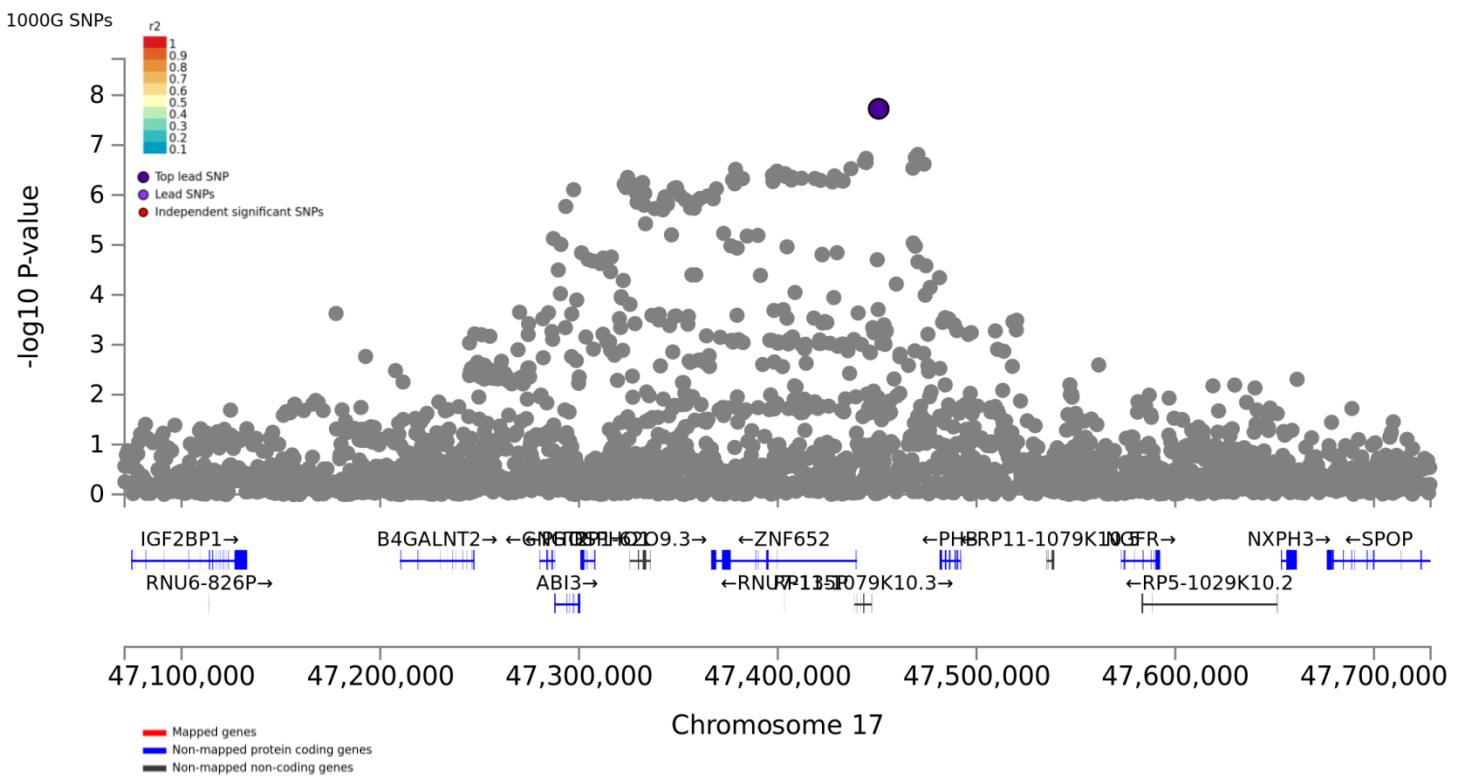


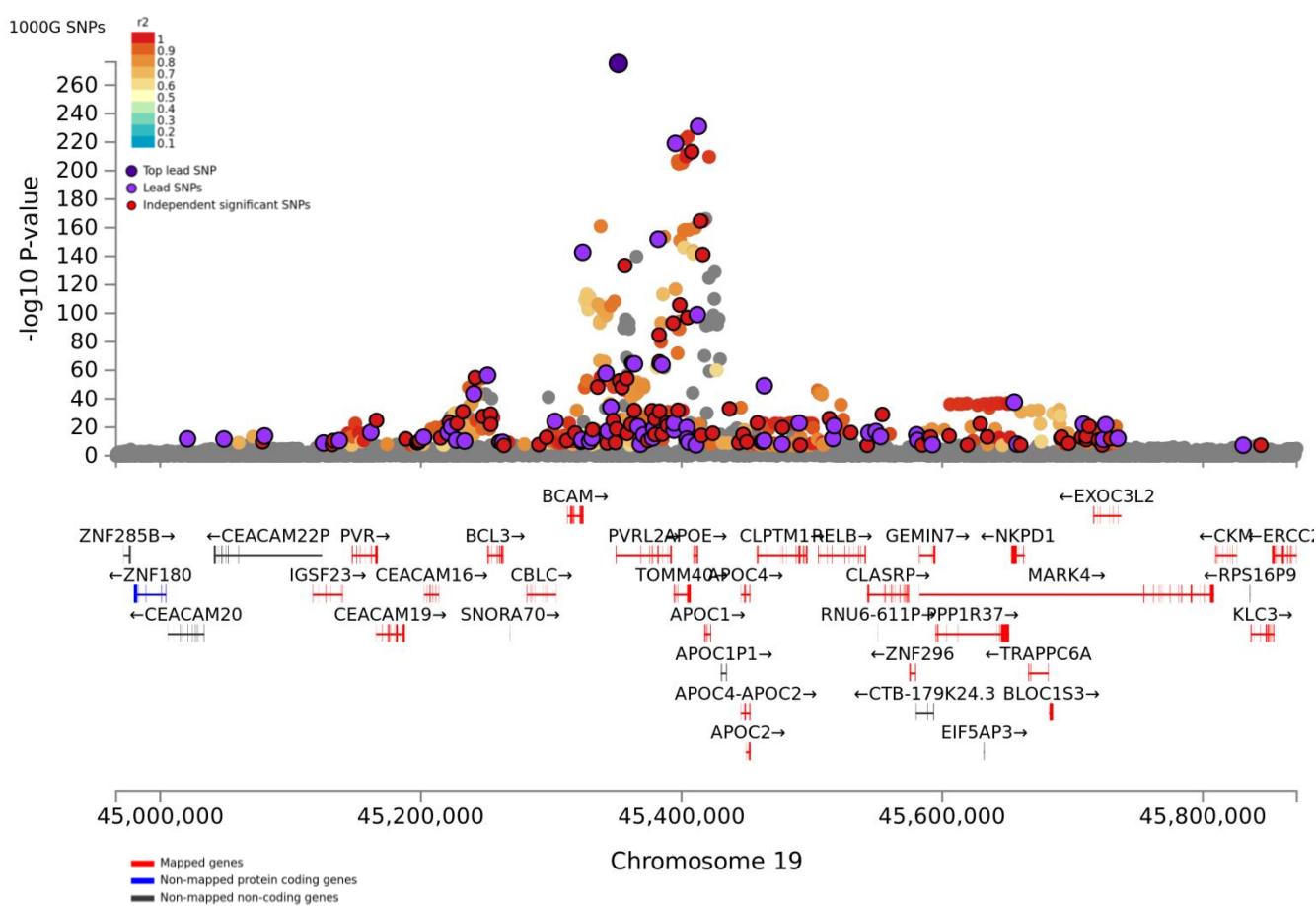
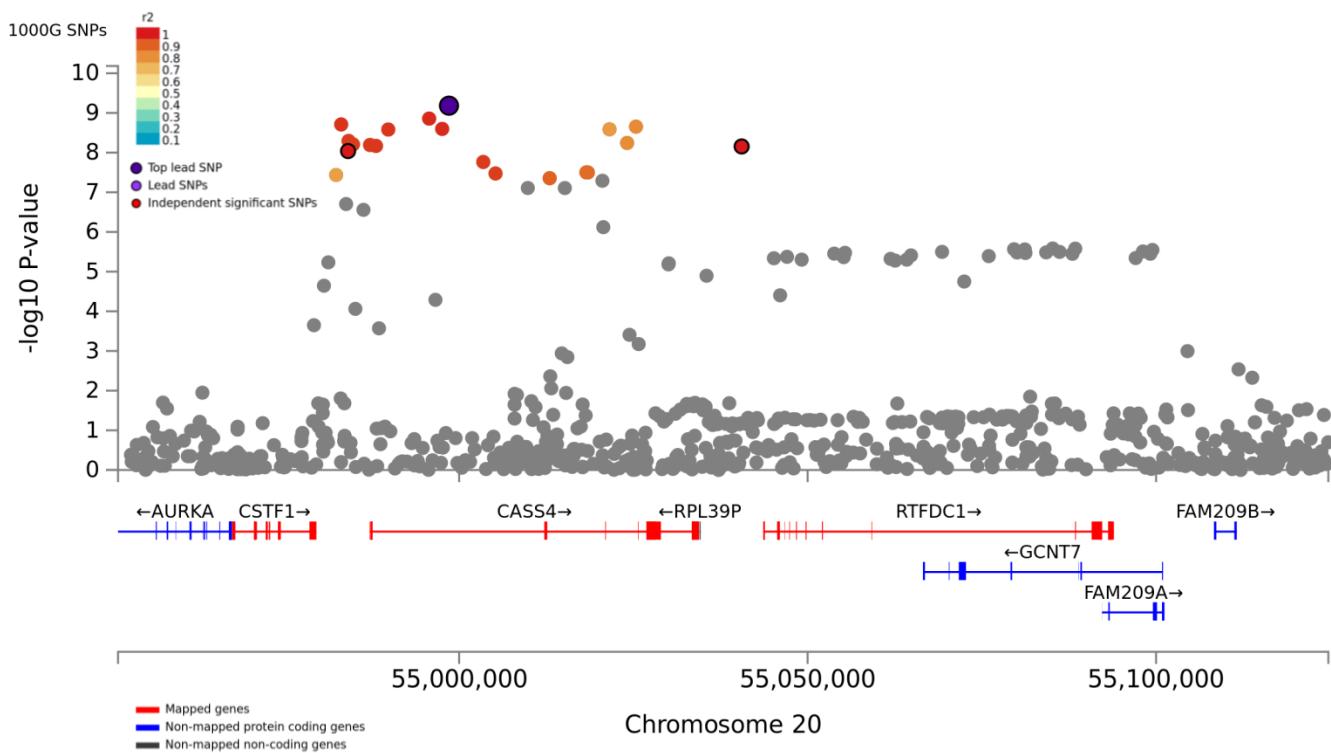


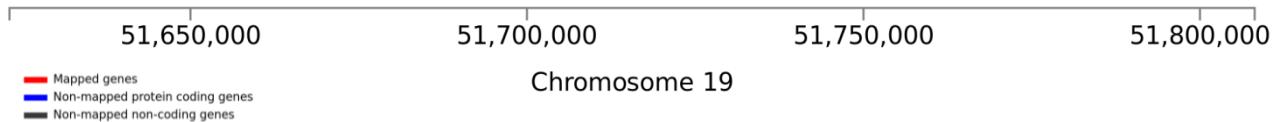
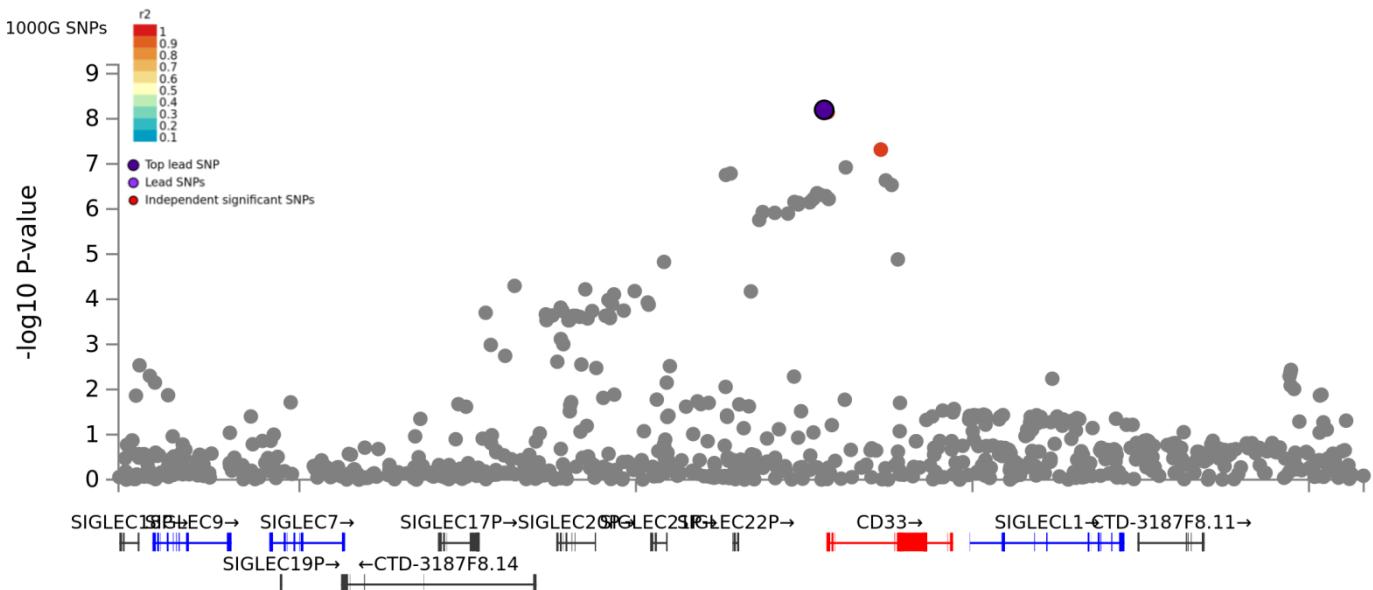
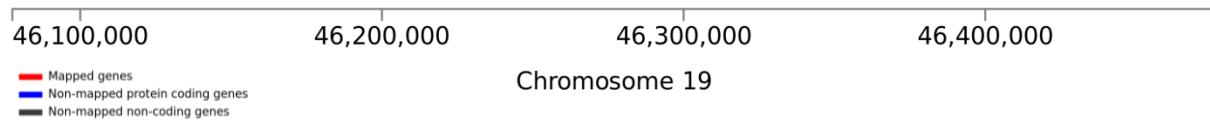
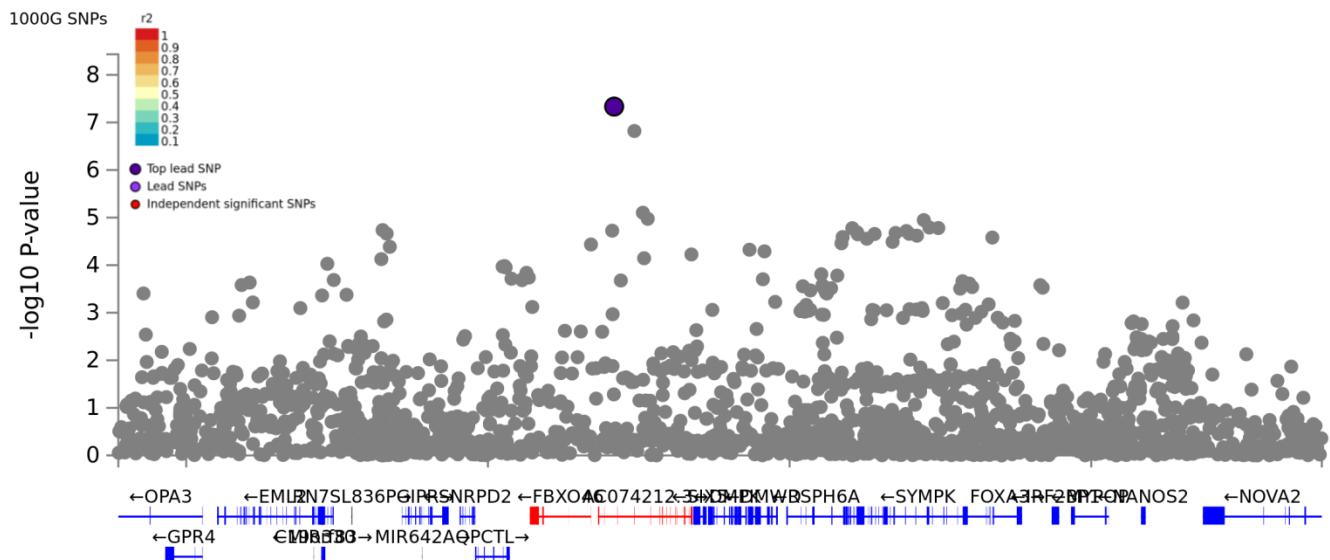


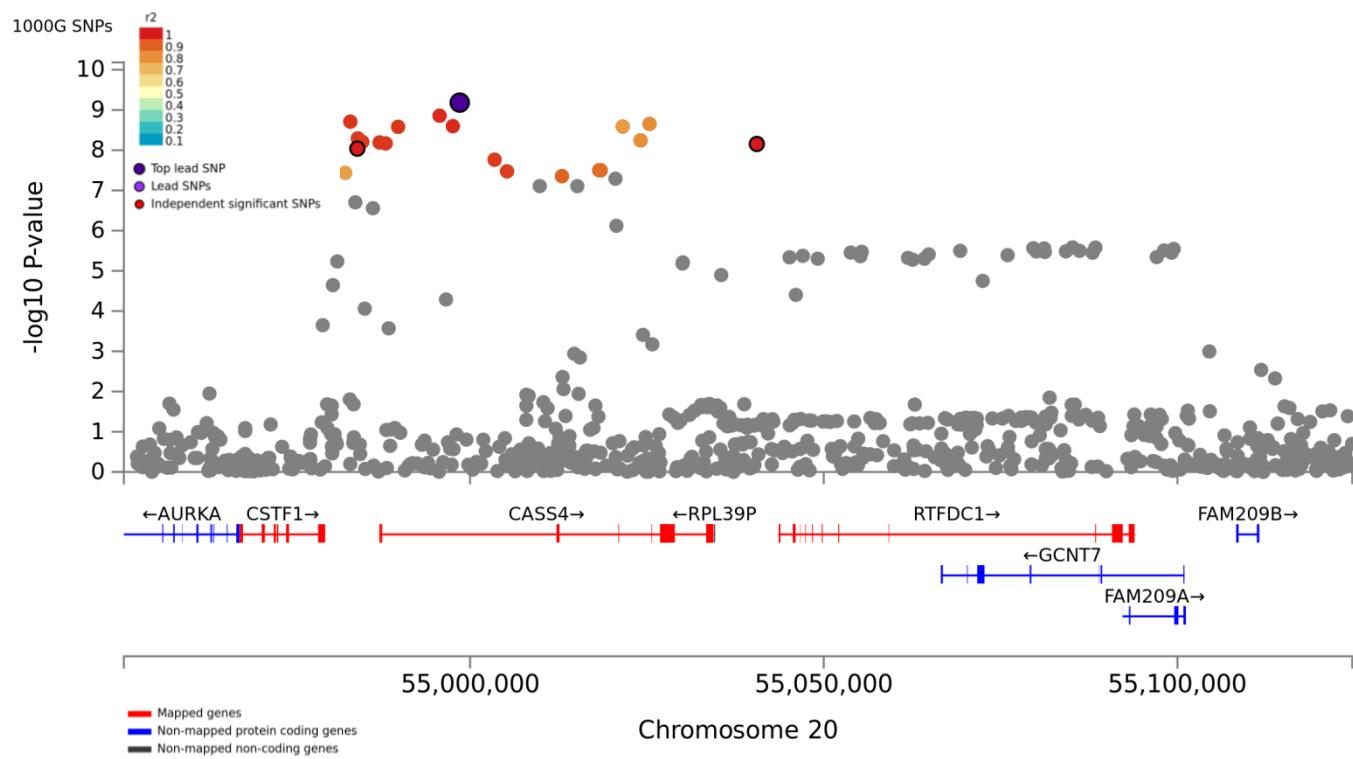








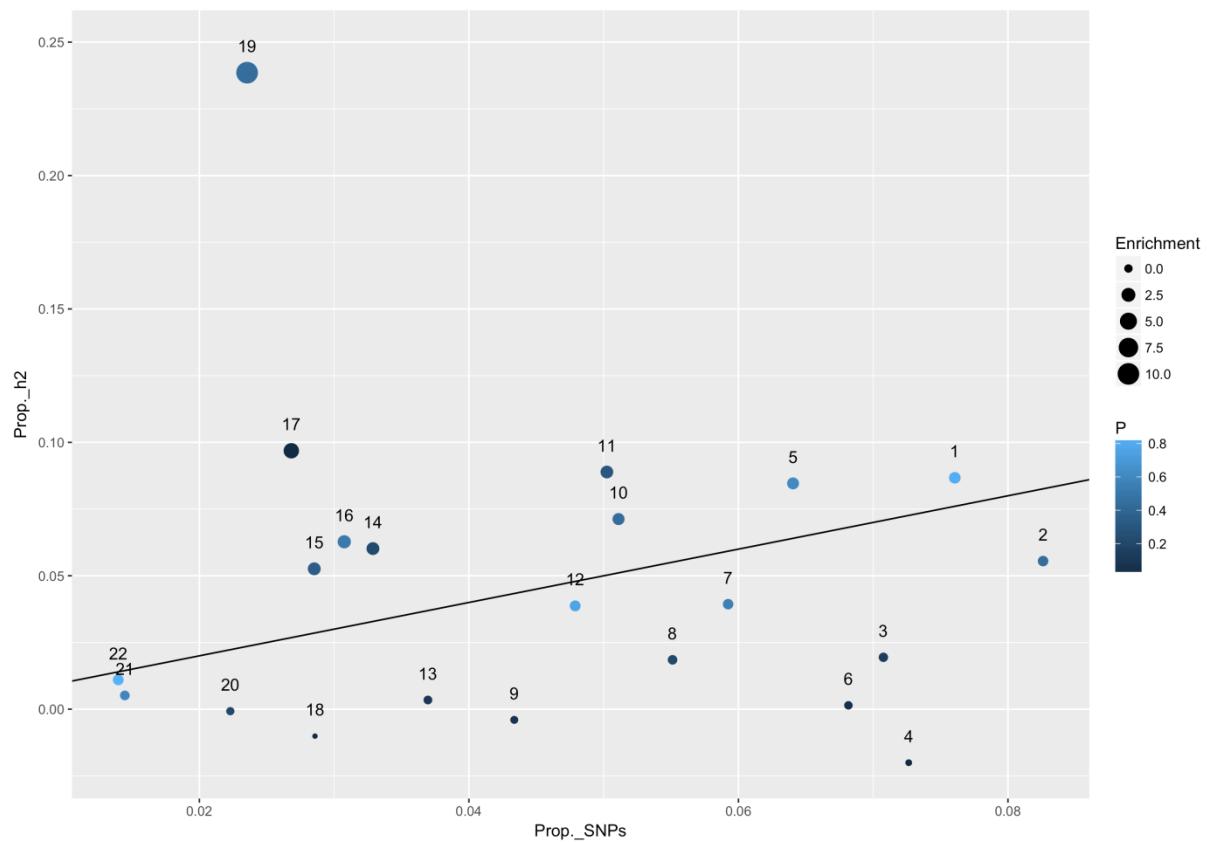




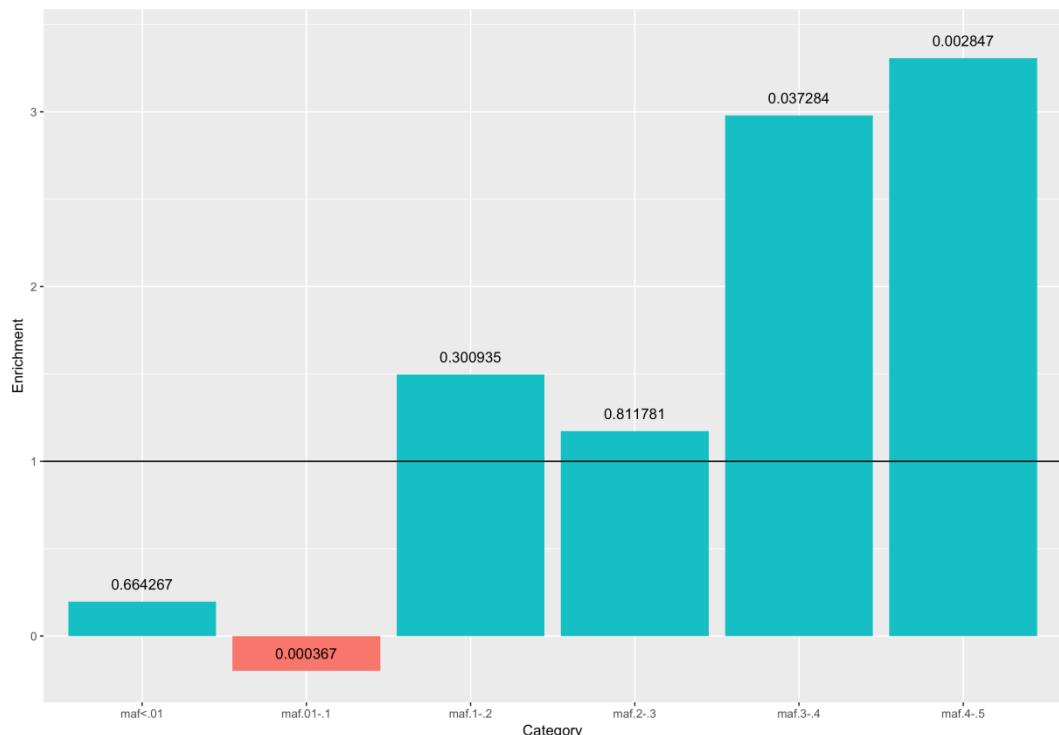
It is made available under a CC-BY-NC-ND 4.0 International license.

**Supplementary Figure 3. Partitioned heritability results for the meta-analysis.** Variants were binned by chromosome or minor allele frequency and tested for a significant over- or underrepresentation as to what is expected by chance. A) Enrichment results for heritability calculations where variants have been partitioned per chromosome. B) Enrichment results for heritability calculations where variants have been partitioned into multiple categories based on minor allele frequency.

**A**

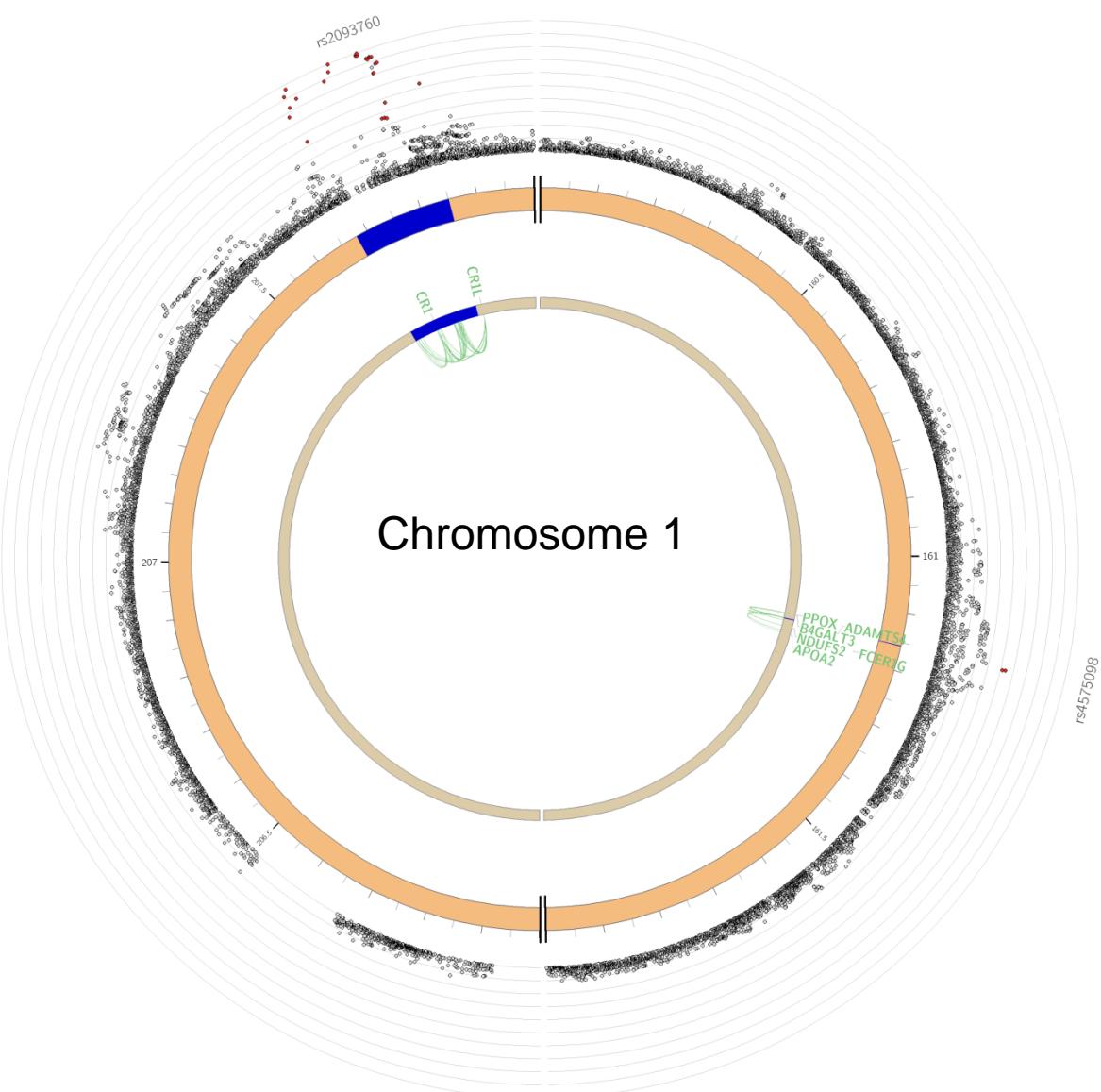


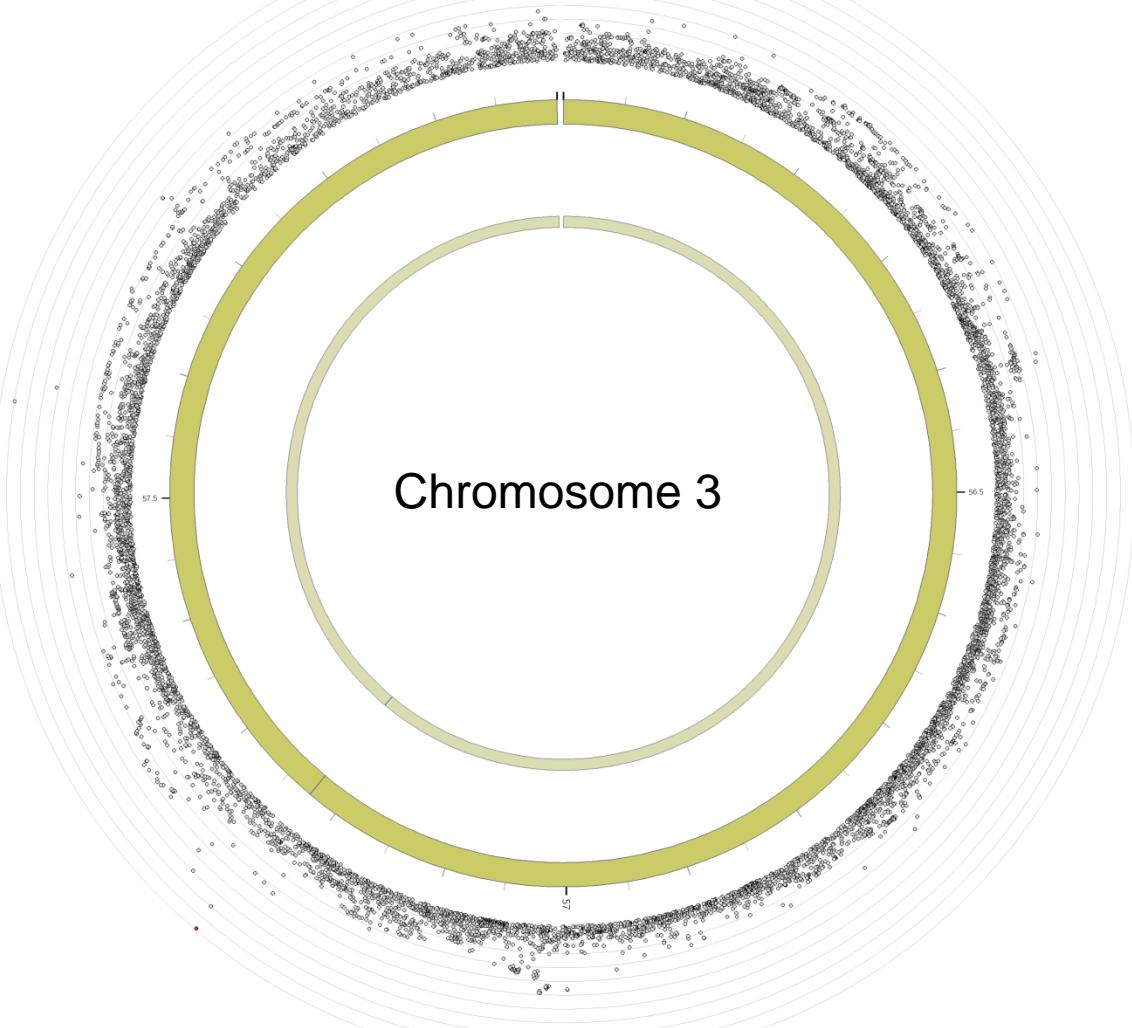
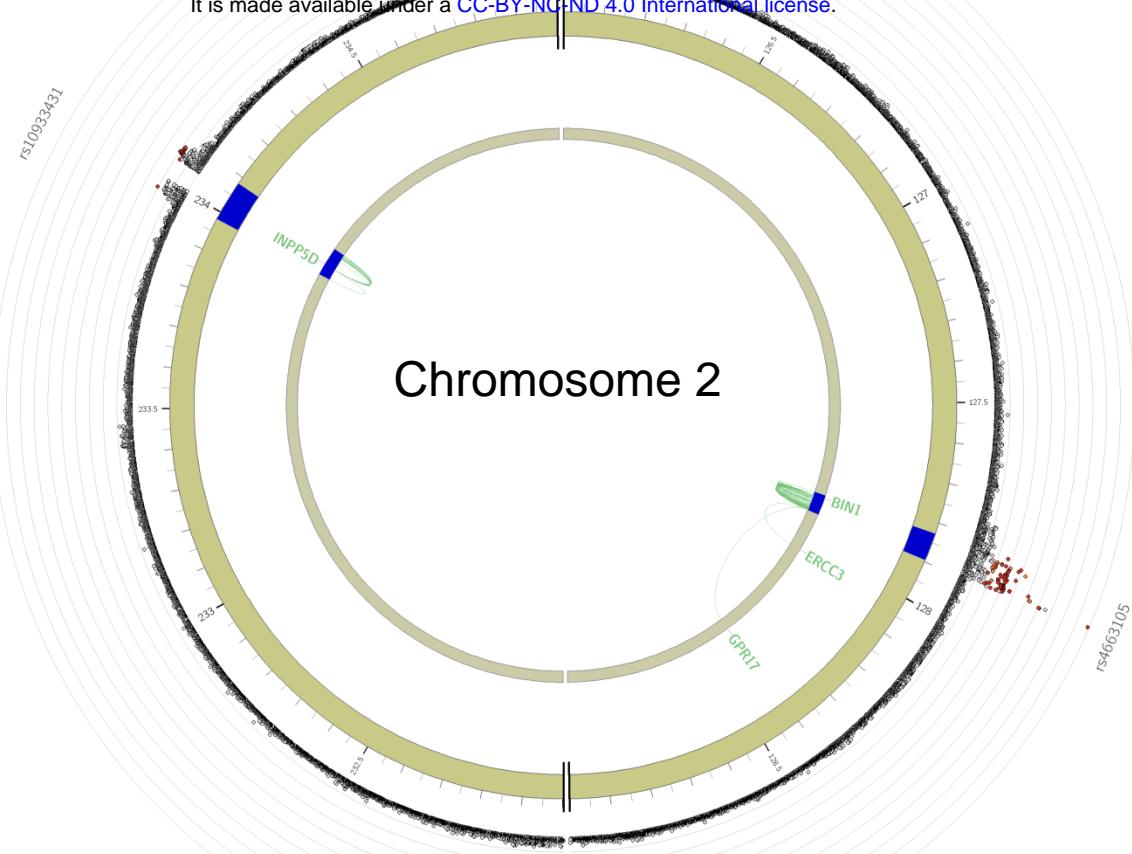
**B**

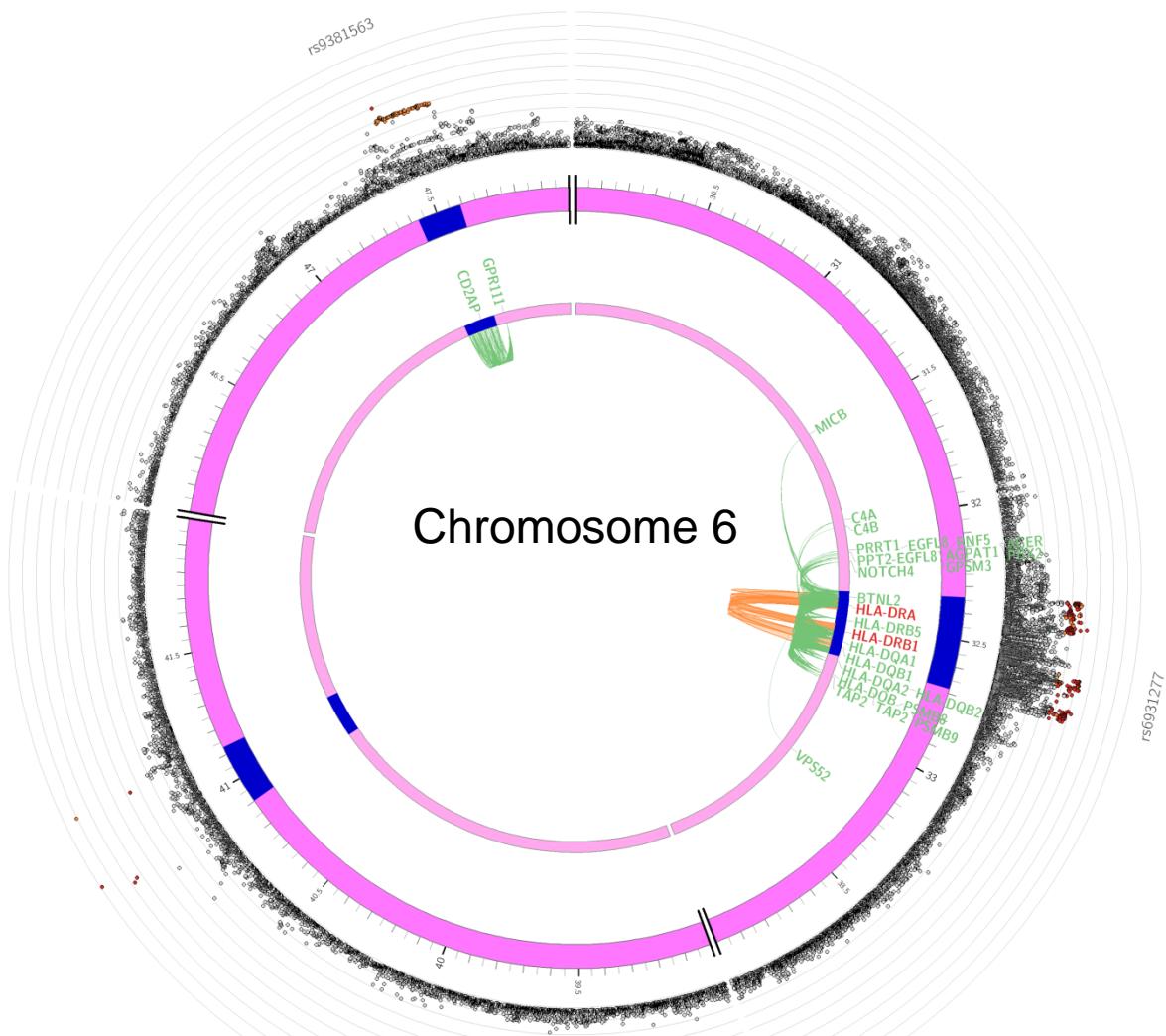
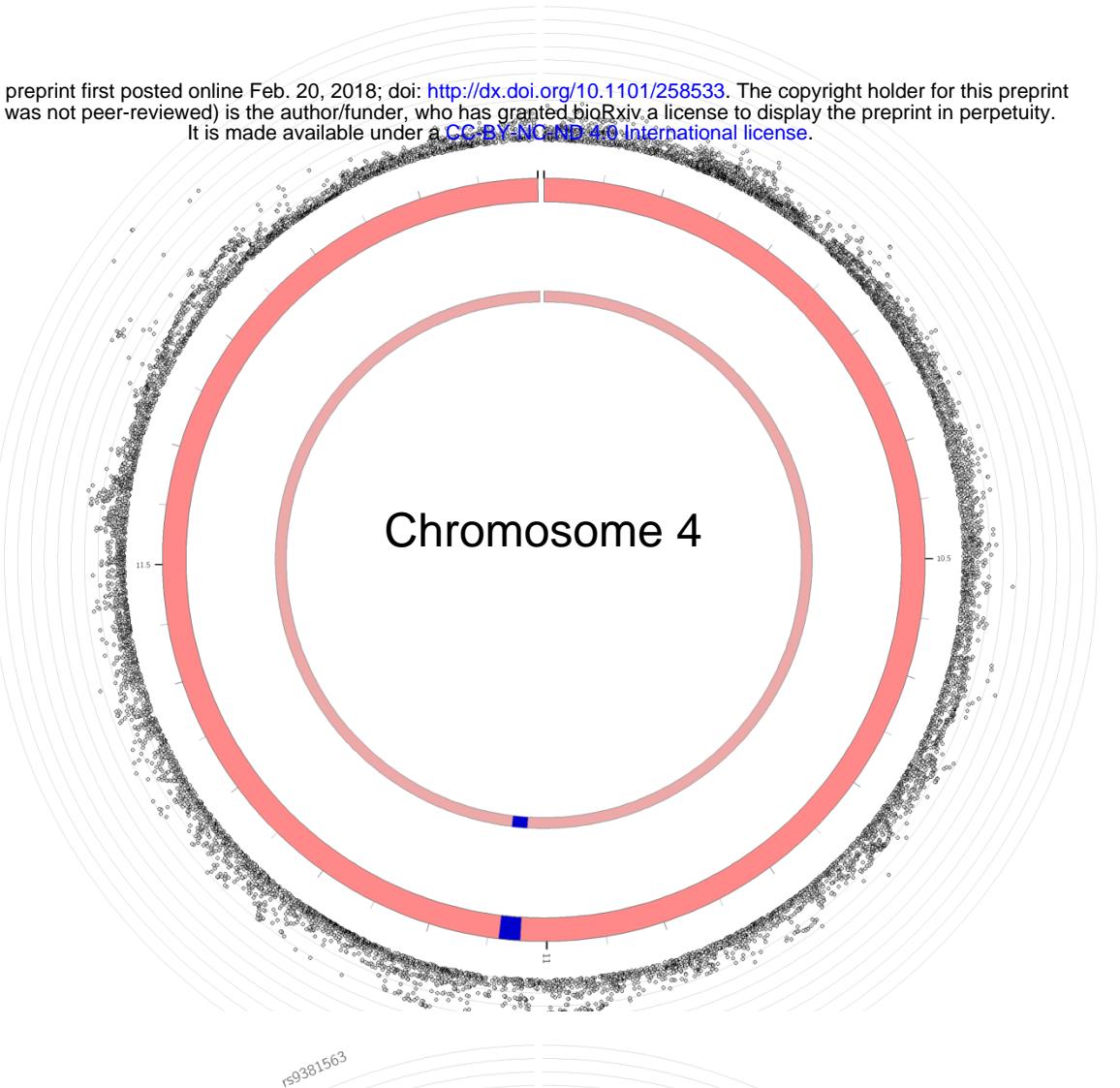


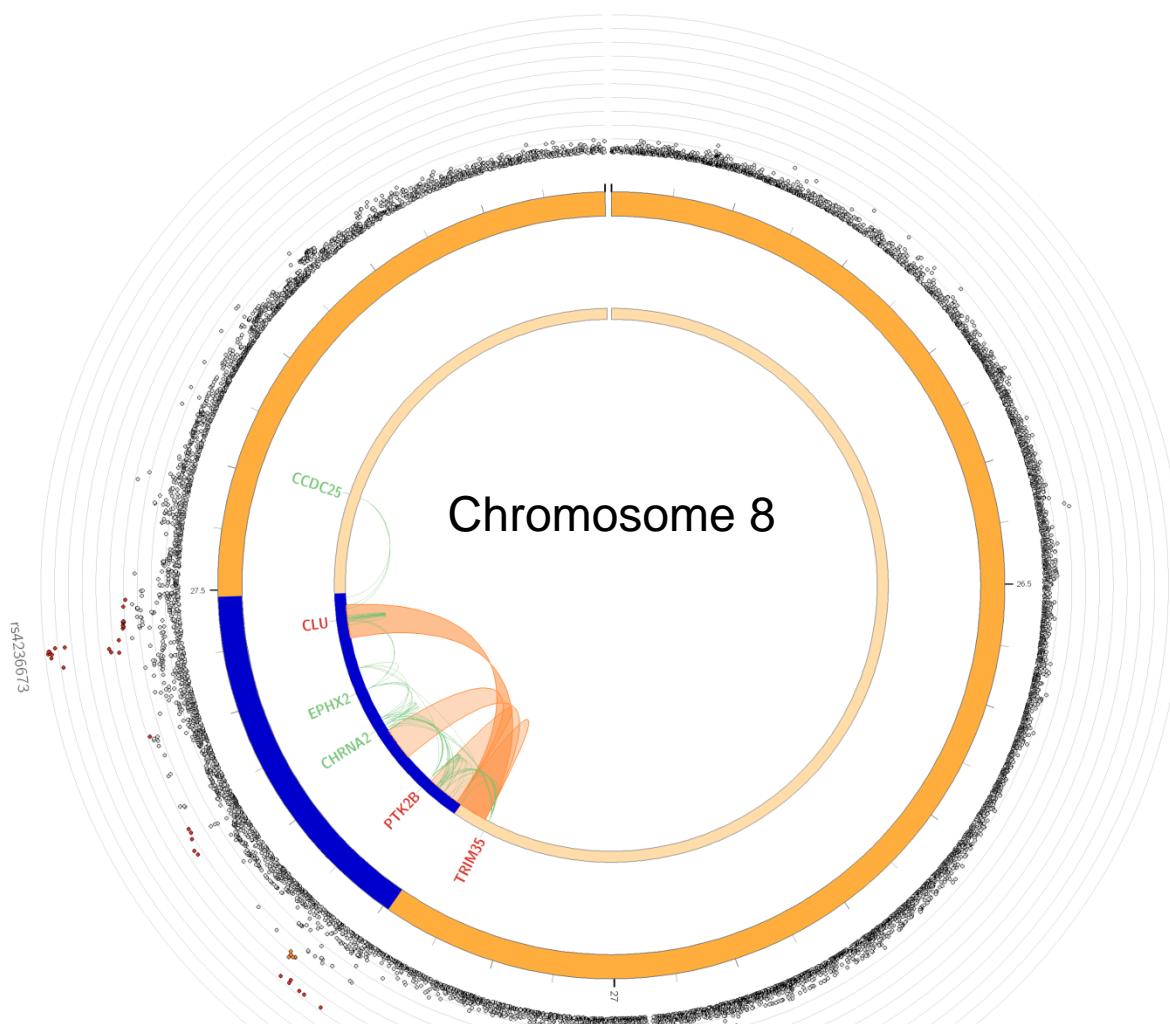
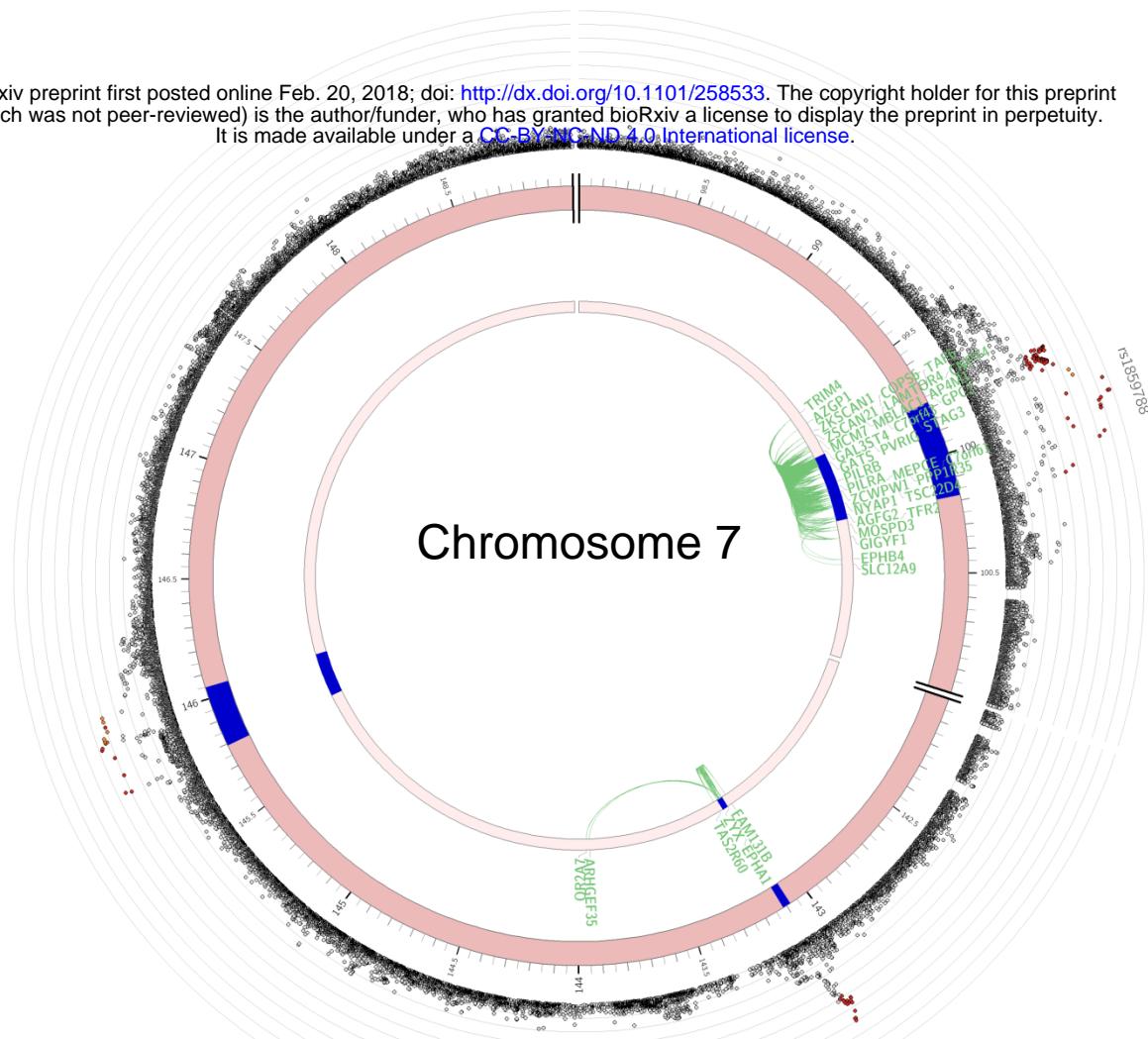
It is made available under a CC-BY-NC-ND 4.0 International license.

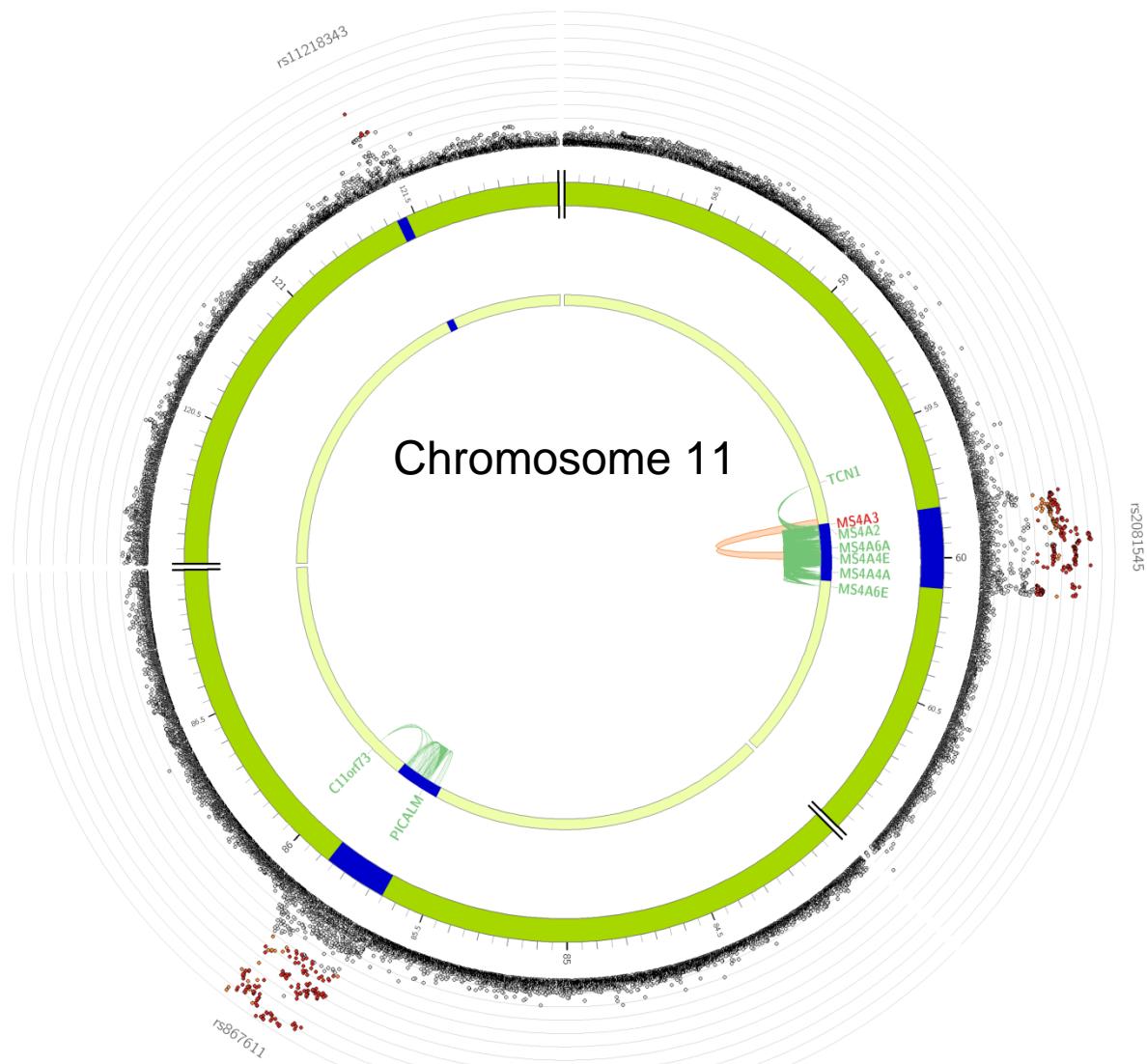
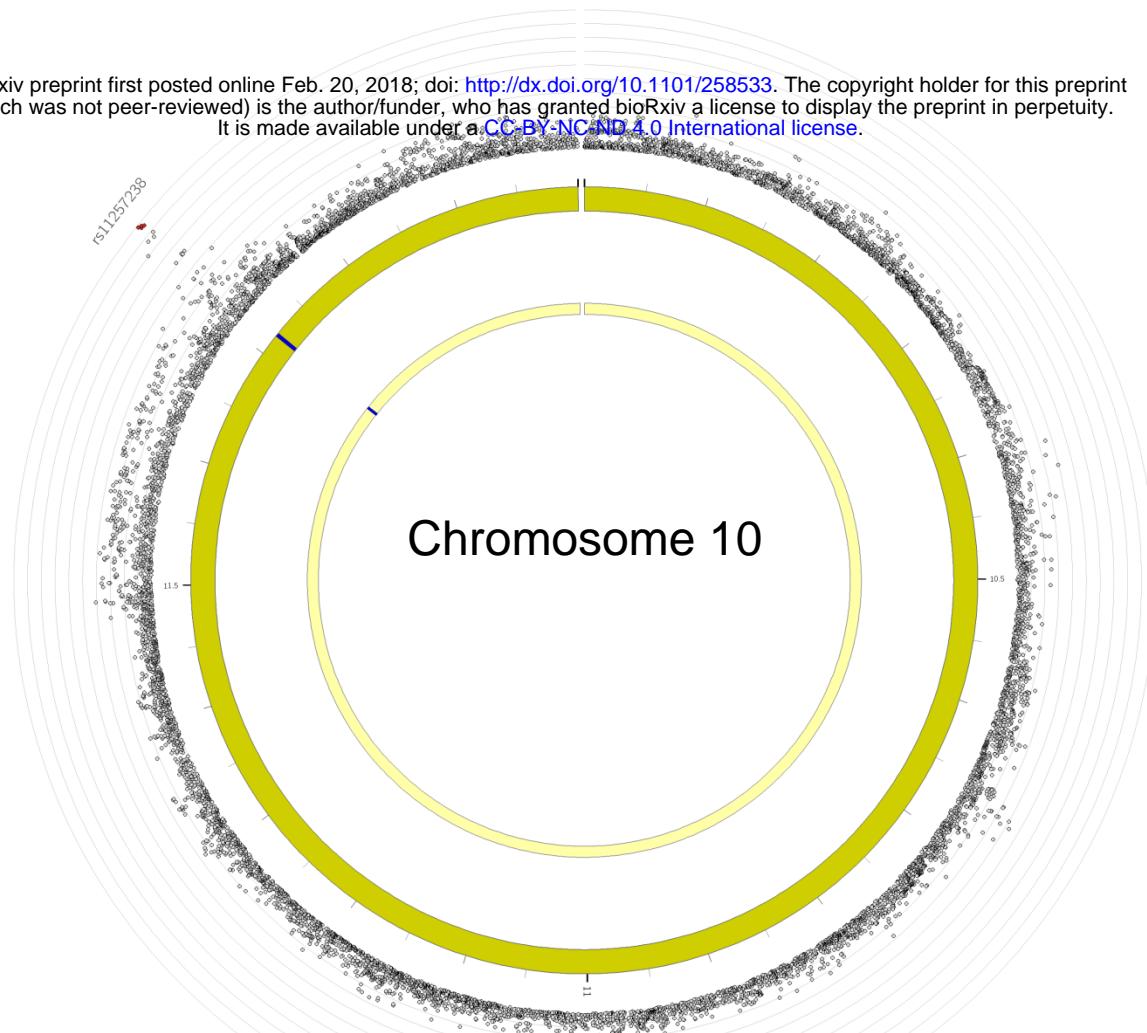
**Supplementary Figure 4. Full circos plots of chromatin interactions and eQTLs for all chromosomes with significantly associated loci.** The distinct layers and colors correspond to various features. The outer layer contains zoomed in Manhattan plots containing only SNPs with  $P < 0.05$ . SNPs in genomic risk loci are color-coded as a function of their maximum  $r^2$  to the one of the independent significant SNPs in the locus, as follows: red ( $r^2 > 0.8$ ), orange ( $r^2 > 0.6$ ), green ( $r^2 > 0.4$ ) and blue ( $r^2 > 0.2$ ). SNPs that are not in LD with any of the independent significant SNPs (with  $r^2 \leq 0.2$ ) are grey. The second layer displays the position of the genomic risk loci in blue. The third layer contains the mapped genes that are implicated by chromatin interactions and/or eQTL analysis (orange = chromatin interaction; green = eQTL; red = chromatin interaction and eQTL).

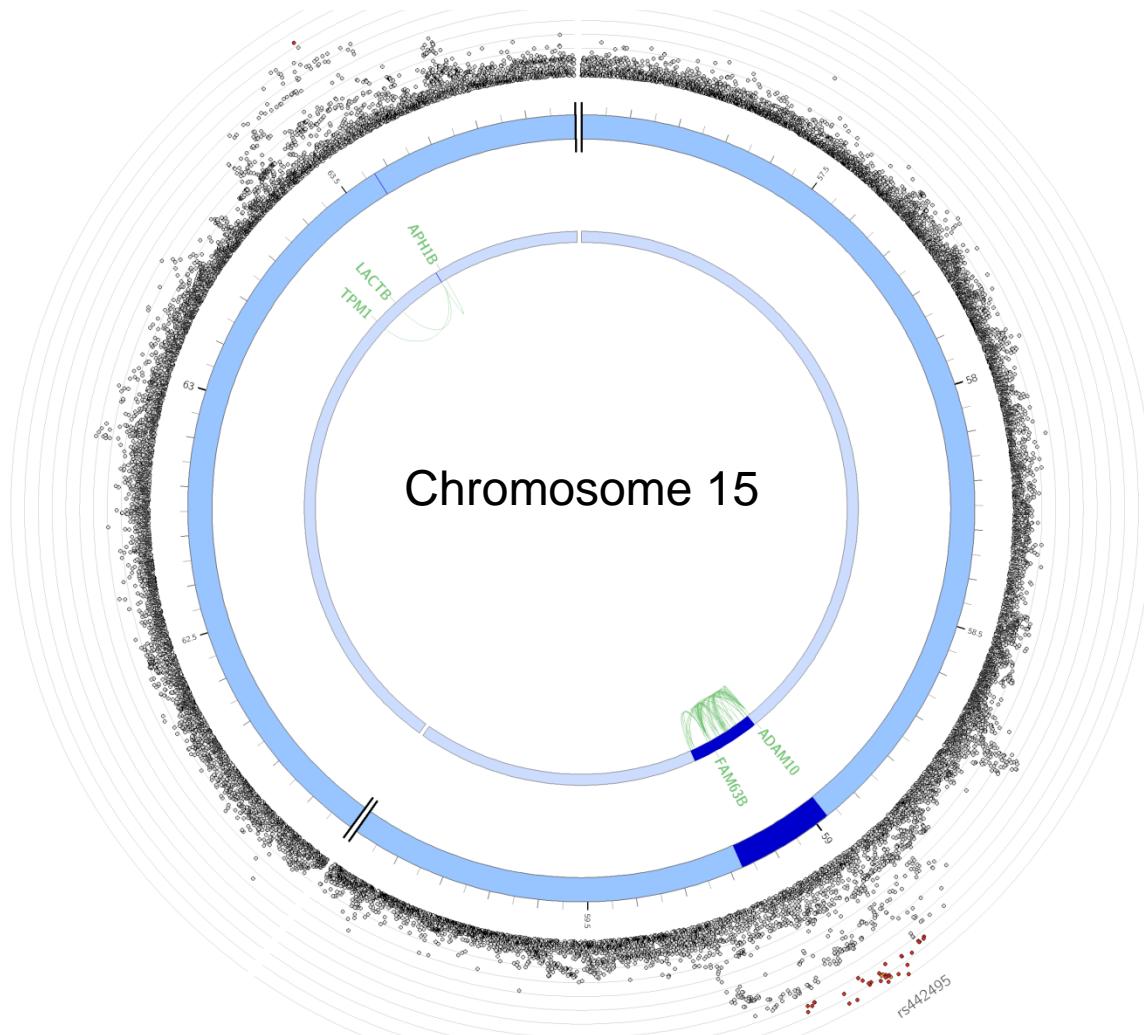
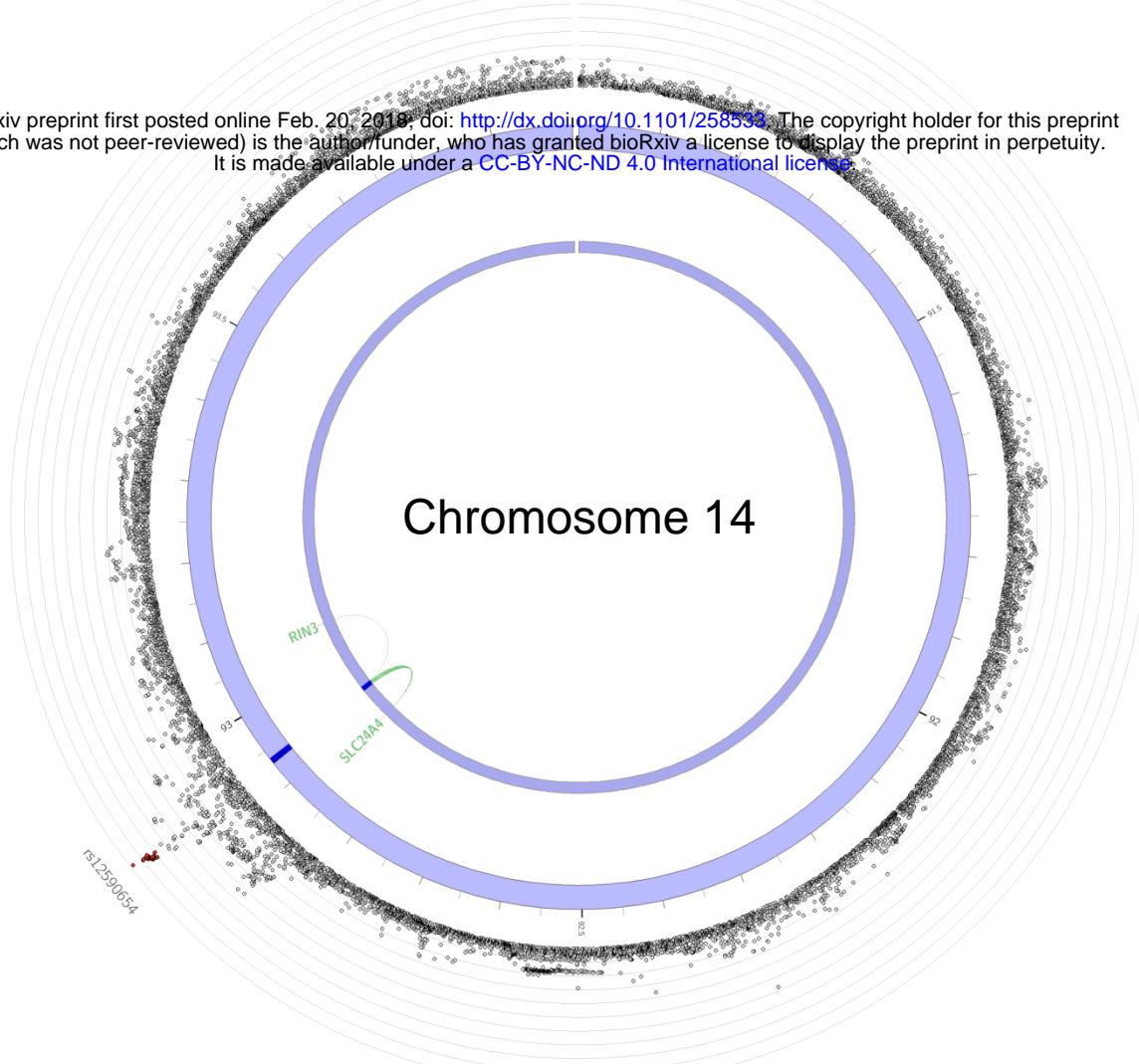


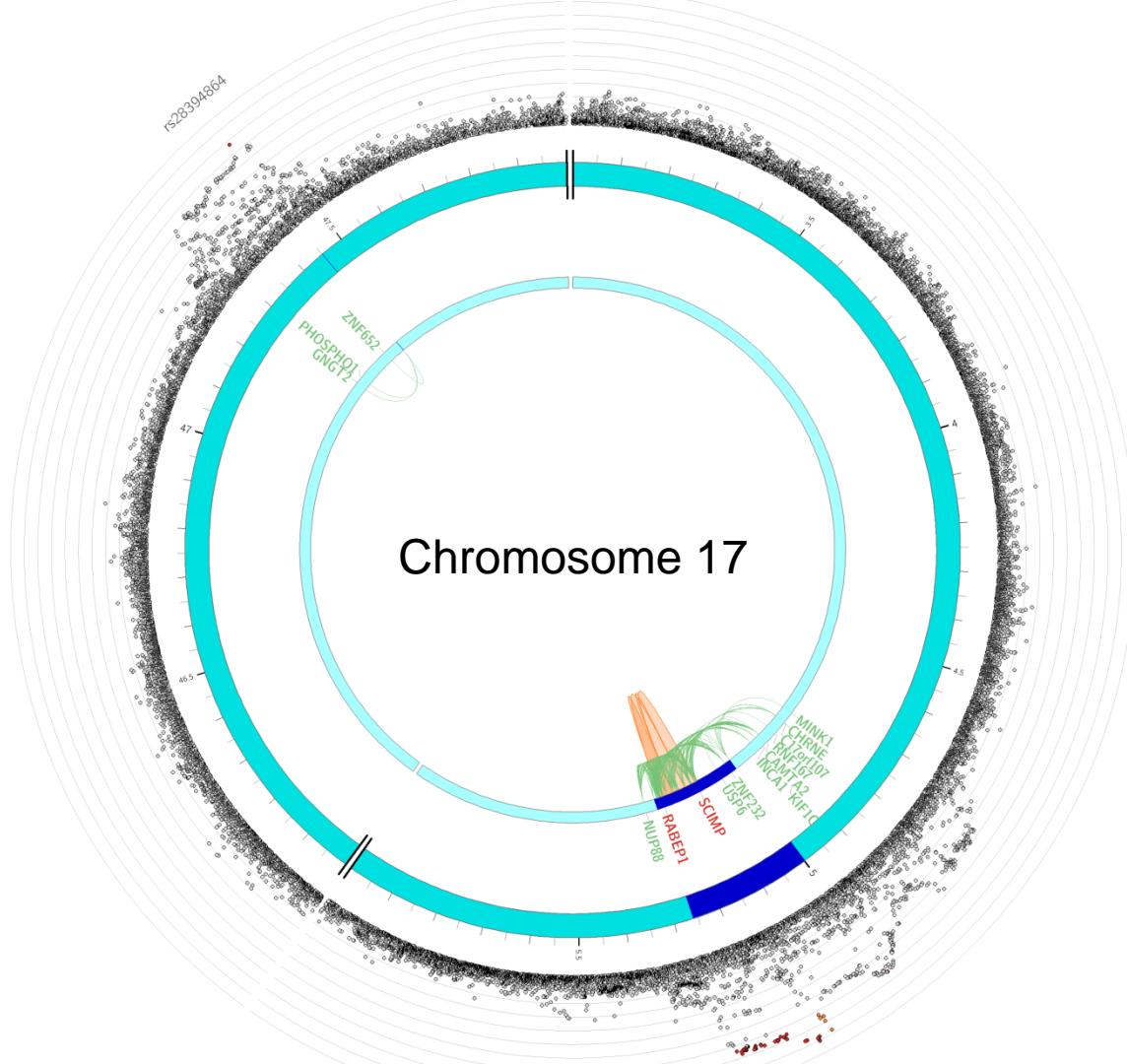
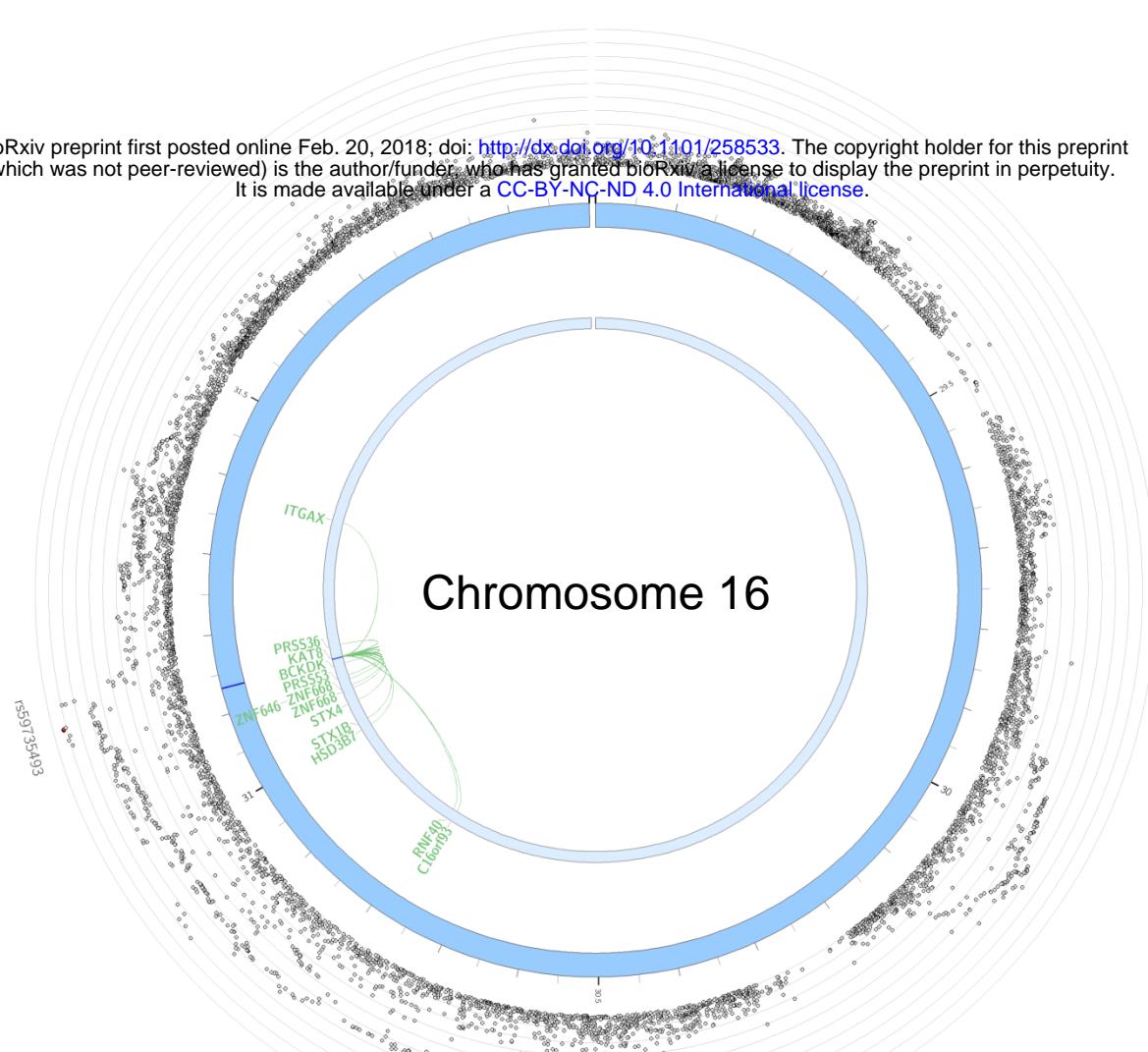


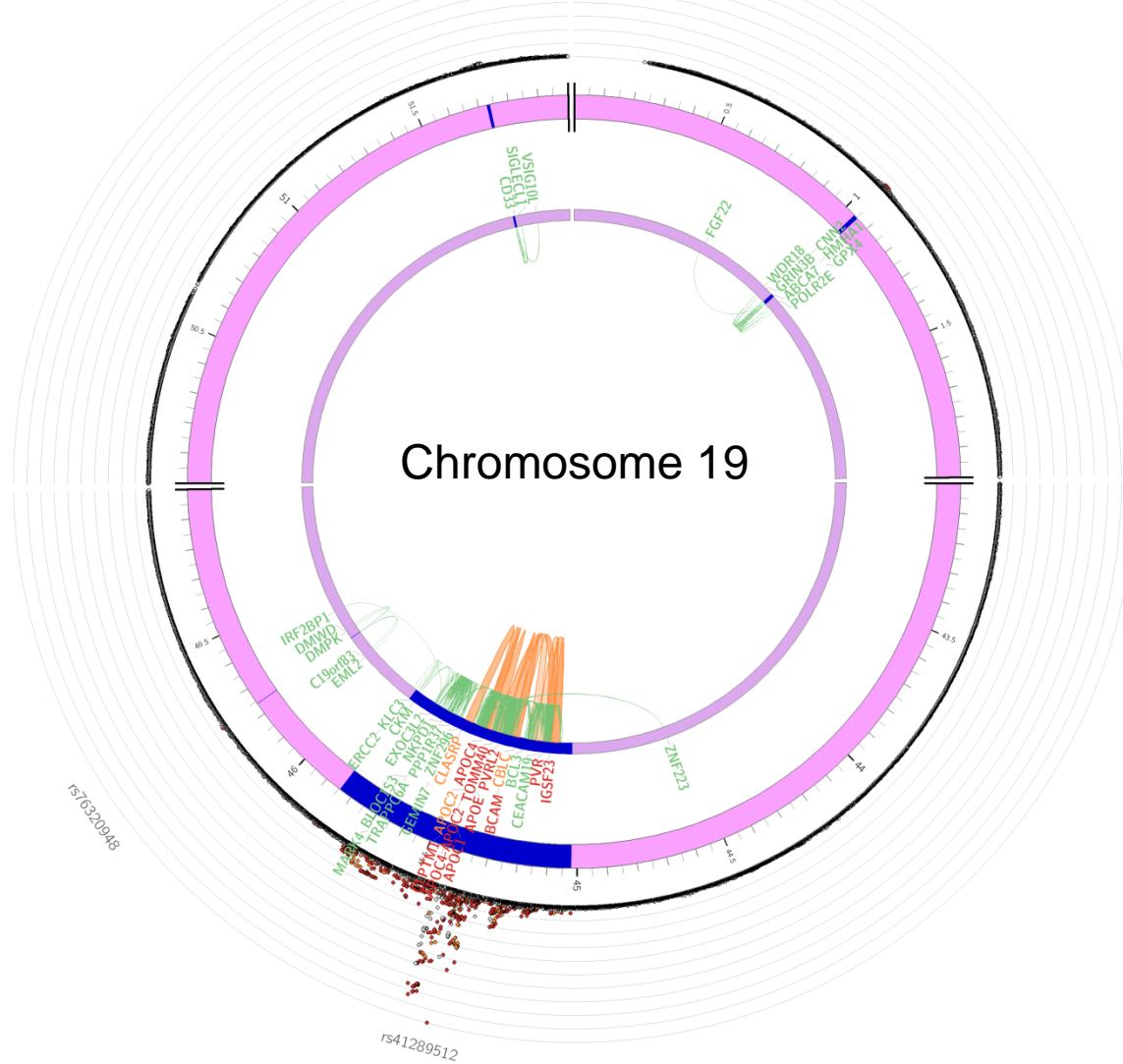
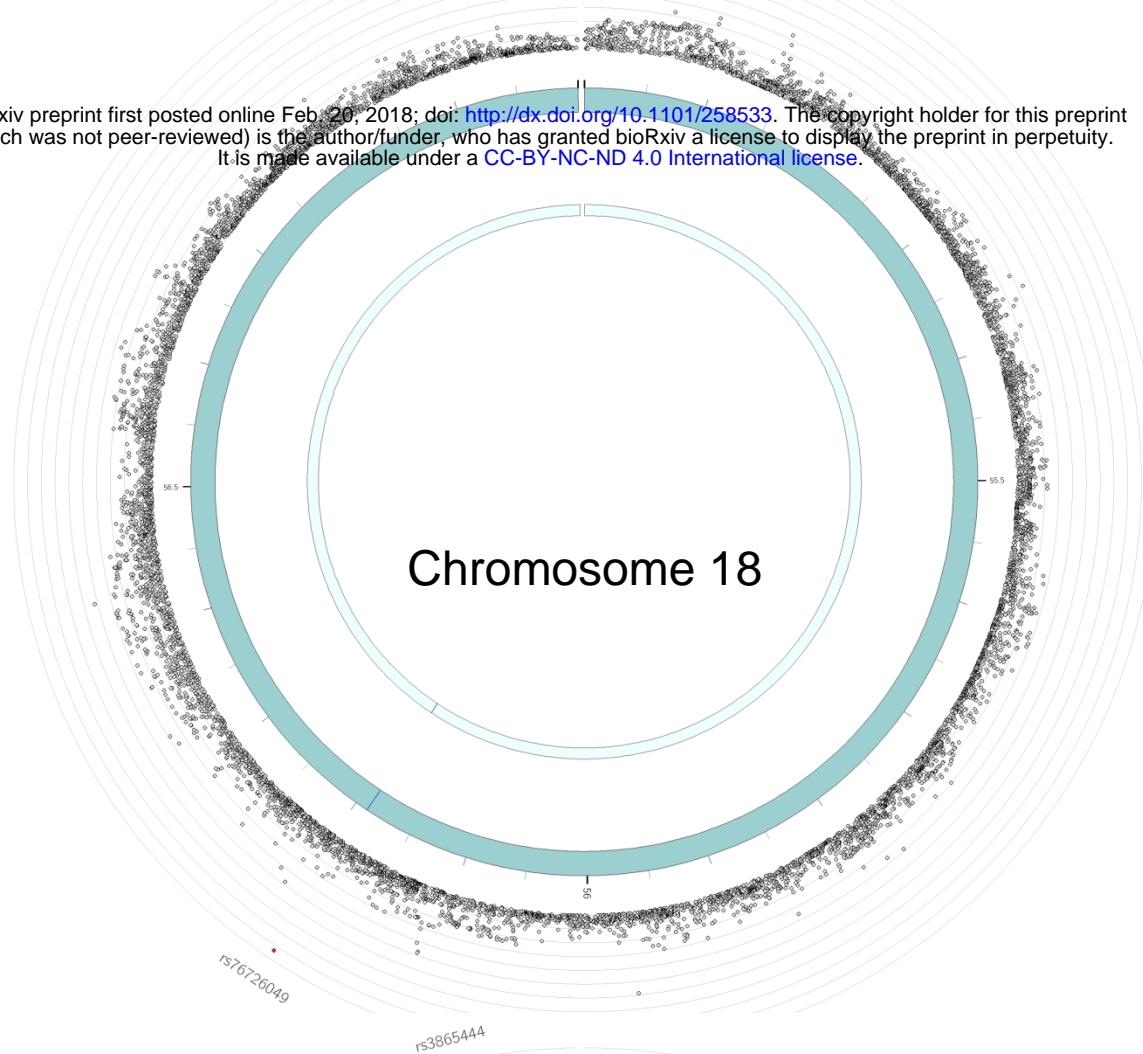


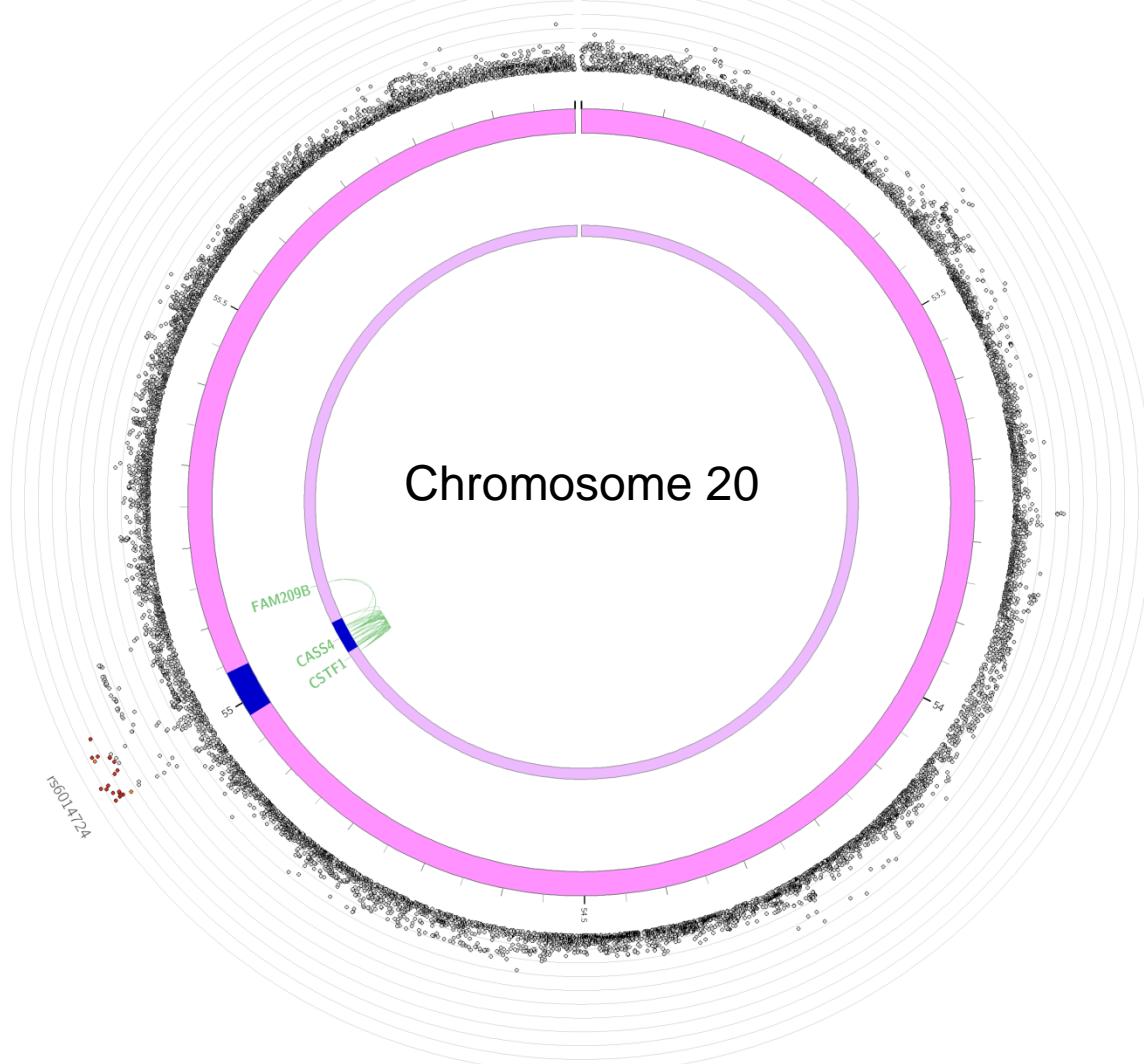




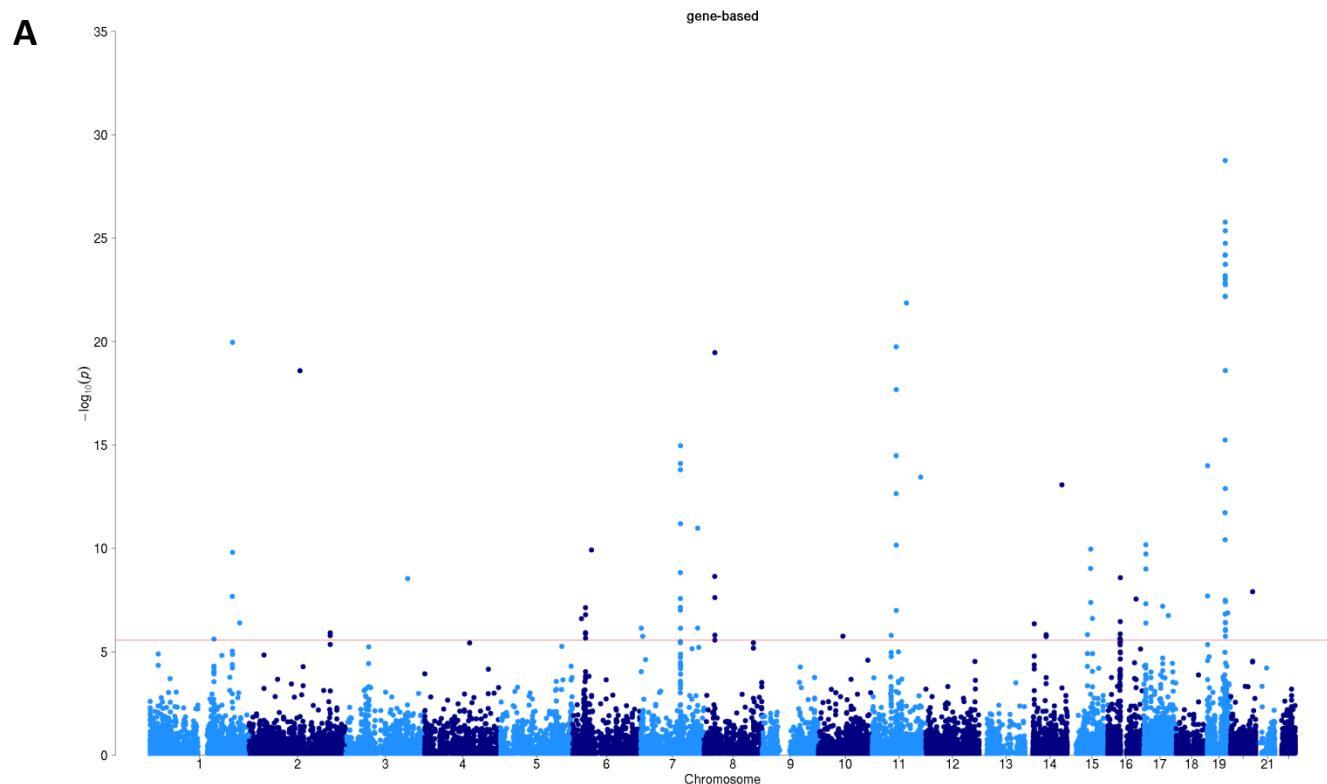






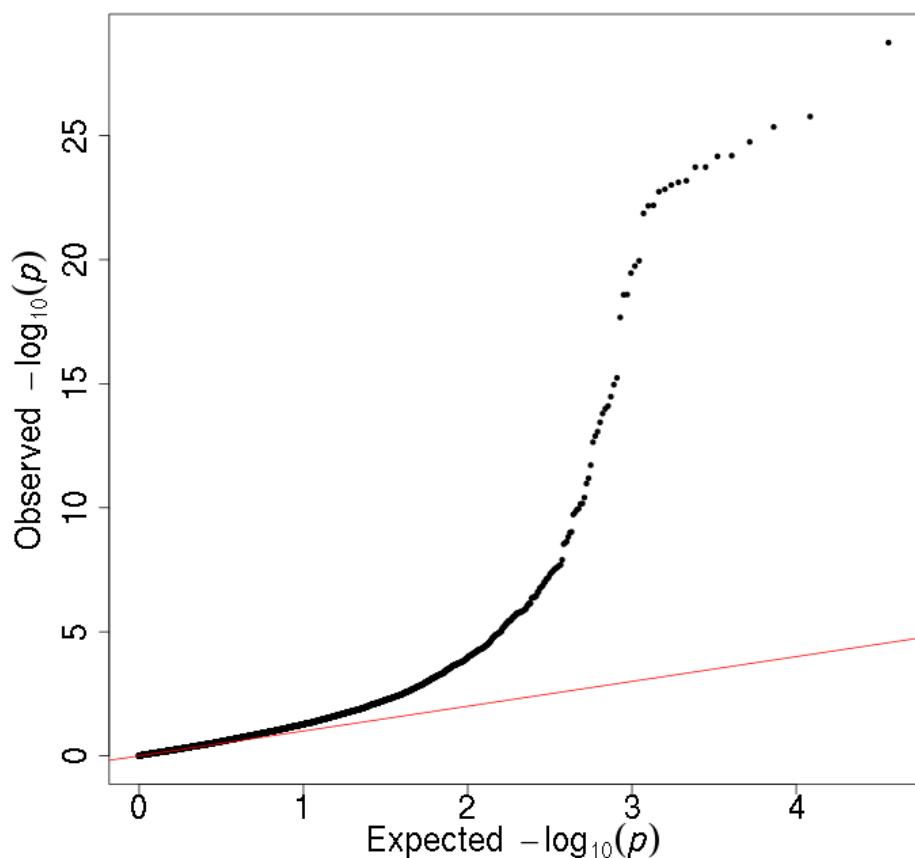


**Supplementary Figure 5. Gene-based association results with MAGMA.** A) The Manhattan plot displays all associations per gene ordered according to their genomic position (start of gene) on the x-axis and showing the strength of the association with the  $-\log_{10}$  transformed  $P$ -values on the y-axis. B) The QQ plot displays the expected  $-\log_{10}$  transformed  $p$ -values on the x-axis and the observed  $-\log_{10}$  transformed  $p$ -values on the y-axis.



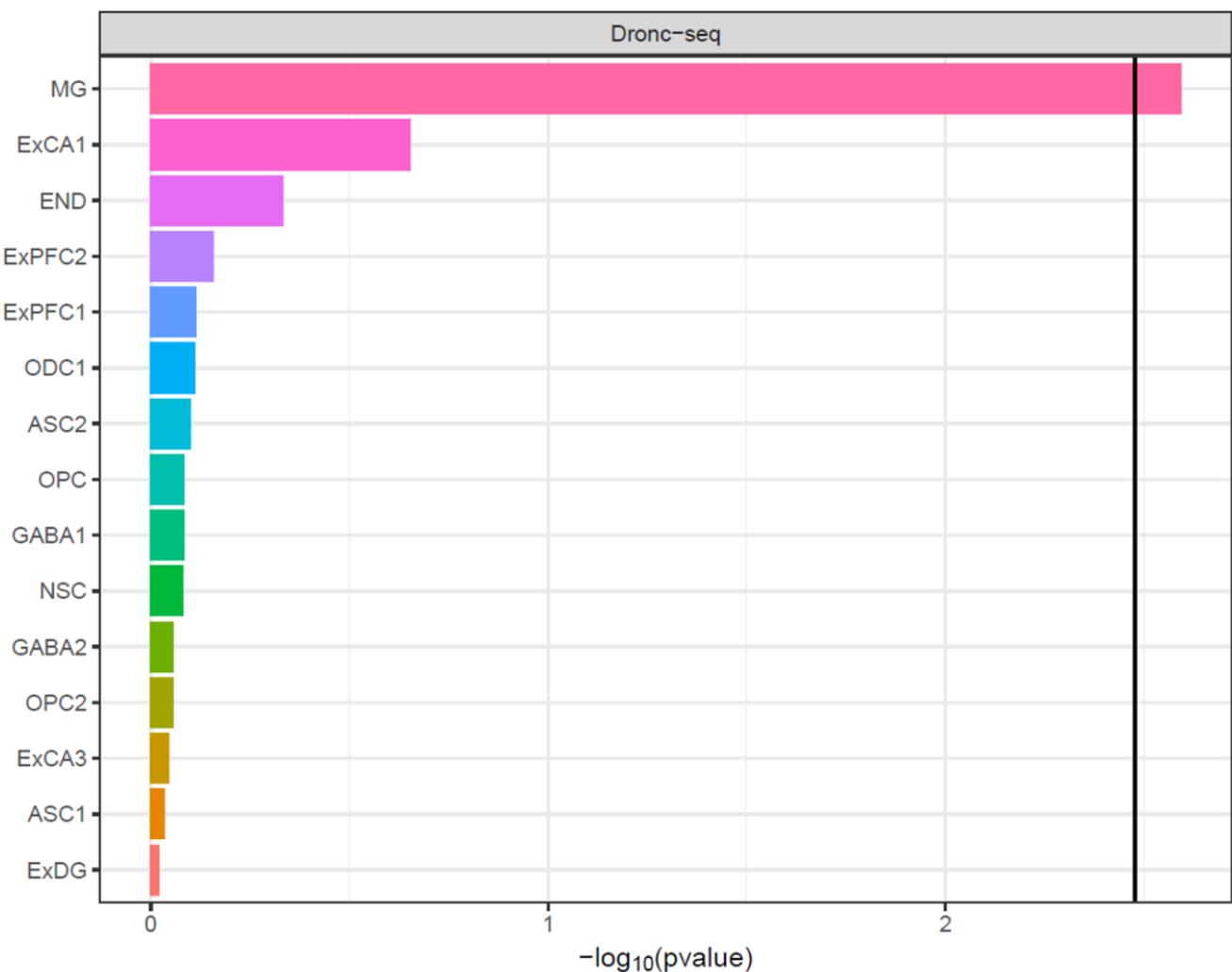
**B**

gene-based



It is made available under a CC-BY-NC-ND 4.0 International license.

**Supplementary Figure 6. Single-cell expression gene-set results of human brain tissue.** The black vertical line indicates the significance threshold correcting for number of tests within category. MG = microglia; ExCA1 = Hippocampal CA 1 pyramidal neurons; END = Endothelial cells; ExPFC2 = Prefrontal glutameric neurons 2; ExPFC1 = Prefrontal glutameric neurons 1; ODC1 = Oligodendrocytes; ASC2 = Astrocytes 2; OPC = Oligodendrocyte precursor cells 1; GABA1 = GABAergic interneurons 1; NSC = Neuronal stem cells; GABA2 = GABAergic interneurons 2; OPC2 = Oligodendrocyte precursor cells 2; ExCA3 = Hippocampal CA 3 pyramidal neurons; ASC1 = Astrocytes 1; ExDG = Dentate gyrus granule neurons.



It is made available under a [CC-BY-NC-ND 4.0 International license](#).

**Supplementary Figure 7.** Mendelian Randomization tests for the effect of correlated phenotypes on risk for Alzheimer's disease. For independent significant SNPs from each correlated phenotype, effect sizes of the SNPs for Alzheimer's disease ( $b_{zy}$ ) are shown on the x-axis and effect sizes for correlated phenotypes are on the y-axis ( $b_{zx}$ ). The dotted line represents a line with slope of ( $b_{xy}$ ) and an intercept of 0. Red dots represent outliers that were excluded for the Mendelian Randomization analysis.

