

# 1 Common genetic risk variants identified in the SPARK cohort implicate 2 *DDHD2* as a novel autism risk gene

3  
4 Nana Matoba<sup>1,2</sup>, Dan Liang<sup>1,2</sup>, Huaigu Sun<sup>1,2</sup>, Nil Aygün<sup>1,2</sup>, Jessica C. McAfee<sup>1,2</sup>, Jessica E.  
5 Davis<sup>3-8</sup>, Laura M. Raffield<sup>1</sup>, Huijun Qian<sup>9</sup>, Joseph Piven<sup>10</sup>, Yun Li<sup>1,11,12</sup>, Sriam Kosuri<sup>3-8</sup>, Hyejung  
6 Won<sup>1,2,\*</sup>, Jason L. Stein<sup>1,2,\*</sup>

- 7  
8 1. Department of Genetics, University of North Carolina at Chapel Hill, North Carolina, USA  
9 2. UNC Neuroscience Center, University of North Carolina at Chapel Hill, North Carolina, USA  
10 3. Department of Chemistry and Biochemistry, University of California, Los Angeles, Los  
11 Angeles, California, USA  
12 4. UCLA-DOE Institute for Genomics and Proteomics, University of California, Los Angeles,  
13 Los Angeles, California, USA  
14 5. Molecular Biology Institute, University of California, Los Angeles, Los Angeles, California,  
15 USA  
16 6. Quantitative and Computational Biology Institute, University of California, Los Angeles, Los  
17 Angeles, California, USA  
18 7. Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, University  
19 of California, Los Angeles, Los Angeles, California, USA  
20 8. Jonsson Comprehensive Cancer Center, University of California, Los Angeles, Los Angeles,  
21 California, USA  
22 9. Department of Statistics and Operations Research, University of North Carolina at Chapel  
23 Hill, North Carolina, USA  
24 10. Department of Psychiatry and the Carolina Institute for Developmental Disabilities,  
25 University of North Carolina at Chapel Hill, North Carolina, USA  
26 11. Department of Biostatistics, University of North Carolina at Chapel Hill, North Carolina, USA  
27 12. Department of Computer Science, University of North Carolina at Chapel Hill, North  
28 Carolina, USA

29 \* These authors equally supervised.

30  
31 Corresponding authors: Hyejung Won or Jason L. Stein

32 Hyejung Won:

33 Address: 7202B Mary Ellen Jones Building, 116 Manning Drive, Chapel Hill, NC 27599

34 Phone: +1 (919) 966-4069

35 Email: [hyejung\\_won@med.unc.edu](mailto:hyejung_won@med.unc.edu);

36 Jason L. Stein:

37 Address: 7202A Mary Ellen Jones Building, 116 Manning Drive, Chapel Hill, NC 27599

38 Phone: +1 (919) 843-5541

39 Email: [jason\\_stein@med.unc.edu](mailto:jason_stein@med.unc.edu)

40  
41  
42 **Short title:** Common genetic risk variants for autism in SPARK

43 **Keywords:** Autism spectrum disorder, GWAS, common variant, MPRA, DDHD2

## 44 Abstract

45 **Background:** Autism spectrum disorder (ASD) is a highly heritable neurodevelopmental disorder.  
46 Large genetically informative cohorts of individuals with ASD have led to the identification of three  
47 common genome-wide significant (GWS) risk loci to date. However, many more common genetic  
48 variants are expected to contribute to ASD risk given the high heritability. Here, we performed a  
49 genome-wide association study (GWAS) using the Simons Foundation Powering Autism  
50 Research for Knowledge (SPARK) dataset to identify additional common genetic risk factors and  
51 molecular mechanisms underlying risk for ASD.

52 **Methods:** We performed an association study on 6,222 case-pseudocontrol pairs from SPARK  
53 and meta-analyzed with a previous GWAS. We integrated gene regulatory annotations to map  
54 non-coding risk variants to their regulated genes. Further, we performed a massively parallel  
55 reporter assay (MPRA) to identify causal variant(s) within a novel risk locus.

56 **Results:** We identified one novel GWS locus from the SPARK GWAS. The meta-analysis  
57 identified four significant loci, including an additional novel locus. We observed significant  
58 enrichment of ASD heritability within regulatory regions of the developing cortex, indicating that  
59 disruption of gene regulation during neurodevelopment is critical for ASD risk. The MPRA  
60 identified one variant at the novel locus with strong impacts on gene regulation (rs7001340), and  
61 expression quantitative trait loci data demonstrated an association between the risk allele and  
62 decreased expression of *DDHD2* (DDHD domain containing 2) in both adult and pre-natal brains.

63 **Conclusions:** By integrating genetic association data with multi-omic gene regulatory  
64 annotations and experimental validation, we fine-mapped a causal risk variant and demonstrated  
65 that *DDHD2* is a novel gene associated with ASD risk.

66

## 67 Introduction

68  
69 Autism spectrum disorder (ASD) is a common neurodevelopmental disorder characterized by  
70 characteristic social deficits as well as ritualistic behaviors (1). Because ASD is highly heritable  
71 (~50-80%) (2–6), a number of studies have been conducted to identify both rare and common  
72 genetic variants contributing to risk for ASD. While previous studies have successfully identified  
73 rare *de novo* presumed loss of function mutations leading to risk for ASD (7–13), these *de novo*  
74 variants do not explain the large heritability and therefore are missing an important component of  
75 ASD risk.

76  
77 To identify common inherited genetic risk factors, genome-wide association studies (GWAS) have  
78 now accumulated over 18,000 individuals with ASD and have begun discovering genome-wide  
79 significant (GWS) loci that explain some of the inherited risks for ASD (14). The three GWS ASD  
80 susceptibility loci discovered previously explain in total only 0.13% of the liability for autism risk,  
81 whereas all common variants are estimated to explain 11.8% of liability (14). Therefore, there are  
82 more common risk variants to be discovered, which requires larger sample sizes to provide  
83 sufficient power to detect risk variants of small effect (15–17). The newly established genetic  
84 cohort, SPARK (Simons Foundation Powering Autism Research for Knowledge)  
85 (<https://sparkforautism.org/>) is planning to collect and analyze data from 50,000 individuals with  
86 ASD (18). SPARK has recently released genotype data for over 8,000 families or singletons with  
87 ASD, which we utilize here to increase the power of ASD GWAS.

88  
89 Once we identify GWS loci, the critical next step is to understand their biological impact. This is  
90 especially challenging because most GWAS identified loci for neurodevelopmental disorders and  
91 other traits are located in poorly annotated non-coding regions with presumed gene regulatory  
92 function (19). In addition, most loci are comprised of multiple single nucleotide polymorphisms

93 (SNPs) that are often inherited together, which makes it difficult to identify the true causal  
94 variant(s) and their regulatory effects (20,21). To overcome these problems, various experimental  
95 validation tools have been developed (22–24). One of these tools, a massively parallel reporter  
96 assay (MPRA), simultaneously evaluates allelic effects on enhancer activity for many variants. In  
97 this assay, exogenous DNA constructs, harboring risk and protective alleles at an associated  
98 variant, drive the expression of a barcoded transcript. Differences in barcode counts between the  
99 risk and protective alleles indicate the regulatory function of that variant (23,24). This assay thus  
100 demonstrates the regulatory potential of individual SNPs and provides evidence of the causal  
101 variants within an associated locus.

102

103 Though fine-mapping approaches can suggest causal variants at a locus, they cannot identify  
104 target genes affected by those variants. Several approaches are designed to link variants to  
105 genes they regulate including expression quantitative trait loci (eQTL) (25–27) as well as  
106 chromatin interaction (via Hi-C) assays (28–30). Recently, we developed Hi-C coupled MAGMA  
107 (H-MAGMA) which predicts genes associated with the target phenotype by integrating long-range  
108 chromatin interaction with GWAS summary statistics (31). Together with existing eQTL resources  
109 in the adult and fetal cortex (32,33), it is possible to link variants associated with risk for ASD to  
110 target genes and functional pathways.

111

112 In this study, we increase the sample size of existing ASD GWAS by adding 6,222 cases-  
113 pseudocontrol pairs from the genetically diverse SPARK project. Our analysis identified five loci  
114 associated with risk for ASD including two novel loci. For one novel locus identified, we used an  
115 MPRA to identify the causal variant within the locus. Further, we integrated multi-level functional  
116 genomic data obtained from the developing brain, including eQTLs, chromatin interactions, and  
117 regulatory elements, to identify *DDHD2* as a candidate gene involved in ASD etiology at the  
118 MPRA-validated locus.

## 119 Methods and Materials

120 This study (analysis of this publicly available dataset) was reviewed by the Office of Human  
121 Research Ethics at UNC, which has determined that this study does not constitute human  
122 subjects research as defined under federal regulations [45 CFR 46.102 (d or f) and 21 CFR  
123 56.102(c)(e)(l)] and does not require IRB approval.

### 124 SPARK dataset

125 SPARK participants who received any of the following diagnoses: autism spectrum disorder [ASD],  
126 Asperger syndrome, autism/autistic disorder and pervasive developmental disorder-not otherwise  
127 specified (PDD-NOS) were recruited. The samples were enriched for affected individuals whose  
128 parents were also available to participate. Participants registered for SPARK online at  
129 [www.SPARKforAutism.org](http://www.SPARKforAutism.org) or at 25 clinical sites across the country by completing questionnaires  
130 on medical history and social communication as described here: <https://www.sfari.org/spark-phenotypic-measures/>.

132 In the current study, participants were drawn from the SPARK 27K release (20190501 ver.)  
133 through SFARIBase (<https://www.sfari.org/resource/sfari-base/>), which included 27,290  
134 individuals (who were genotyped with a SNP array and/or whole-exome sequencing [WES]) with  
135 phenotype information such as sex, diagnosis, and cognitive impairment. The data included  
136 probands and their family members if applicable (e.g. 3,192 quads (2,798 families with unaffected  
137 siblings, 394 with multiple affected siblings), 2,486 trios, and 2,448 duos) (Supplementary Figure  
138 S1). Twenty families in this release overlapped with the Simon's Variations in Individuals Project  
139 (SVIP) cohort and were subsequently removed for the genome-wide association analysis  
140 (Supplementary Figure S2) since the SVIP cohort has targeted probands with 16p11.2 deletions.  
141 We also obtained whole-exome sequencing (WES) data to estimate the imputation accuracy.  
142 Details on genotyping and whole-exome sequencing data, and pre-imputation quality control are  
143 provided in Supplementary Methods.

## 144 Genotype phasing and imputation

145 Phasing was performed using EAGLE v2.4.1 (34)  
146 (<https://data.broadinstitute.org/alkesgroup/Eagle/>) within SPARK samples. Before making  
147 pseudocontrols, we removed two individuals, one each from two pairs of monozygotic twins with  
148 Identity-By-Descent (PI\_HAT)>0.9, by selecting the individual with lower call rates. We then  
149 defined pseudocontrols by PLINK 1.9 (35) ([www.cog-genomics.org/plink/1.9/](http://www.cog-genomics.org/plink/1.9/)) for trios by  
150 selecting the alleles not inherited from the parents to the case (36). We re-phased all SPARK  
151 samples that passed our QC measures with pseudocontrols. Imputation was performed on the  
152 Michigan imputation server (37) (<https://imputationserver.sph.umich.edu/index.html>). Since  
153 SPARK participants are genetically diverse, we imputed genotypes using the Trans-Omics for  
154 Precision Medicine (TOPMed) Freeze 5b (<https://www.nhlbiwgs.org/>) reference panel which  
155 consists of 125,568 haplotypes from multiple ancestries (38,39). Imputation accuracy relative to  
156 WES was assessed using a similar approach to previous work (40) (Supplementary Figure S4)  
157 as described in Supplementary Methods.

## 158 Genome-wide association analysis and Meta-analysis with iPSYCH-PGC 159 study

160 We tested association within the SPARK all case-pseudocontrol pairs (full dataset;  
161 Supplementary Table S1) using PLINK2 generalized linear model (--glm) for SNPs with MAF  $\geq$   
162 0.01 and imputation quality score from minimac4 ( $R^2 > 0.5$ ) (Supplementary Figure S4). In this  
163 model, we did not include any covariates since cases and pseudocontrols are matched on  
164 environmental variables and genetic ancestry. We performed secondary GWAS analyses by  
165 subsetting to only specific ancestry groups. We called ancestry using multidimensional scaling  
166 (MDS) analysis with 988 HapMap3 individuals and one random case from each trio  
167 (Supplementary Figure S3, Supplementary Table S2). Ancestry of individuals from SPARK was  
168 called as European, African or East Asian ancestries if they were within 5 standard deviations of

169 defined HapMap3 population (CEU/TSI; YRI/LWK; or CHB/CHD/JPT, respectively) centroids in  
170 MDS dimensions 1 and 2. Population-specific GWASs were carried out using the same  
171 association model as described above for the SPARK all ancestries dataset. Meta-analyses with  
172 iPSYCH-PGC study (14) were performed by METAL (release 2018-08-28) (41). Additional  
173 information for iPSYCH-PGC summary statistics is provided in Supplementary Methods.

#### 174 Investigation of pleiotropic effects for ASD loci

175 The pleiotropic effects of identified loci were investigated for phenotypes available in the  
176 NHGRI/EBI GWAS catalog (downloaded October 22, 2019) (42) (Supplementary Methods).

#### 177 Linkage Disequilibrium Score regression analysis

178 LD Score regression (LDSC) (v1.0.0) (43,44) was used to estimate genome-wide SNP based  
179 heritability, heritability enrichment of tissue/cell-type specific epigenetic states, and genetic correlation  
180 across phenotypes for GWAS meta-analysis result (Supplementary Methods). Prior to the analyses,  
181 we filtered SNPs to those found in HapMap3 and converted to LDSC input files (.sumstats.gz) using  
182 `munge_sumstats.py`. The pre-computed LD scores for Europeans were obtained from  
183 [https://data.broadinstitute.org/alkesgroup/LDSCORE/eur\\_w\\_ld\\_chr.tar.bz2](https://data.broadinstitute.org/alkesgroup/LDSCORE/eur_w_ld_chr.tar.bz2). For all LDSC analyses,  
184 we used individuals from European ancestry as described in the “Genome-wide association analysis  
185 (GWAS)” section above.

#### 186 Estimating polygenic risk score

187 Polygenic risk scores (PRSs) were calculated based on the iPSYCH-PGC study (14) using  
188 PRSice-2 (45) (<https://www.prsice.info/>). Details on generation of PRS, sex-stratified and family-  
189 type PRS, and parental origin PRS analyses are provided in Supplementary Methods.

190

## 191 H-MAGMA

192 SNP to Ensembl gene annotation was carried out by Hi-C coupled MAGMA (H-MAGMA)  
193 (<https://github.com/thewonlab/H-MAGMA>) by leveraging chromatin-interaction generated from  
194 fetal brain Hi-C (46) as previously described (31). Details on H-MAGMA and functional analyses  
195 of H-MAGMA genes are provided in Supplementary Methods.

## 196 Construction of a Massively Parallel Reporter Assay (MPRA) Library

197 Because the novel SPARK associated locus (chr8:38.19M - chr8:38.45M) was also detected in a  
198 previous schizophrenia GWAS which is better powered, we obtained credible set SNPs for the  
199 locus based on schizophrenia GWAS results (47) (see Supplementary Methods). Ninety-eight  
200 credible set SNPs were detected in this locus. We obtained 150bp sequences that flank each  
201 credible set SNP with the SNP at the center (74bp + 75bp). Because each SNP has risk and  
202 protective alleles, this resulted in 196 total alleles to be tested. We seeded HEK293 cells (ATCC®  
203 CRL-11268™) in 6 wells (total 6 replicates) to be 70-90% confluent at transfection. We used  
204 lipofectamine 2000 (Invitrogen cat#11668) with our final MPRA library following the  
205 manufacturer's instructions. Additional information for construction of MPRA library is available in  
206 Supplementary Methods. MPRA data was analyzed by an mpra package in R (48,49)  
207 (<https://github.com/hansenlab/mpra>) with more details in Supplementary Methods.

## 208 Functional annotation of rs7001340 locus with multi-omic datasets

209 To investigate the target genes affected by allelic variation at rs7001340, we used two expression  
210 quantitative loci (eQTL) datasets derived from fetal cortical brain tissue and adult dorsolateral  
211 prefrontal cortex (32). We also used chromatin accessibility profiles from primary human neural  
212 progenitor cells and their differentiated neuronal progeny (unpublished data from Stein lab).  
213 Further information is provided in Supplementary Methods.

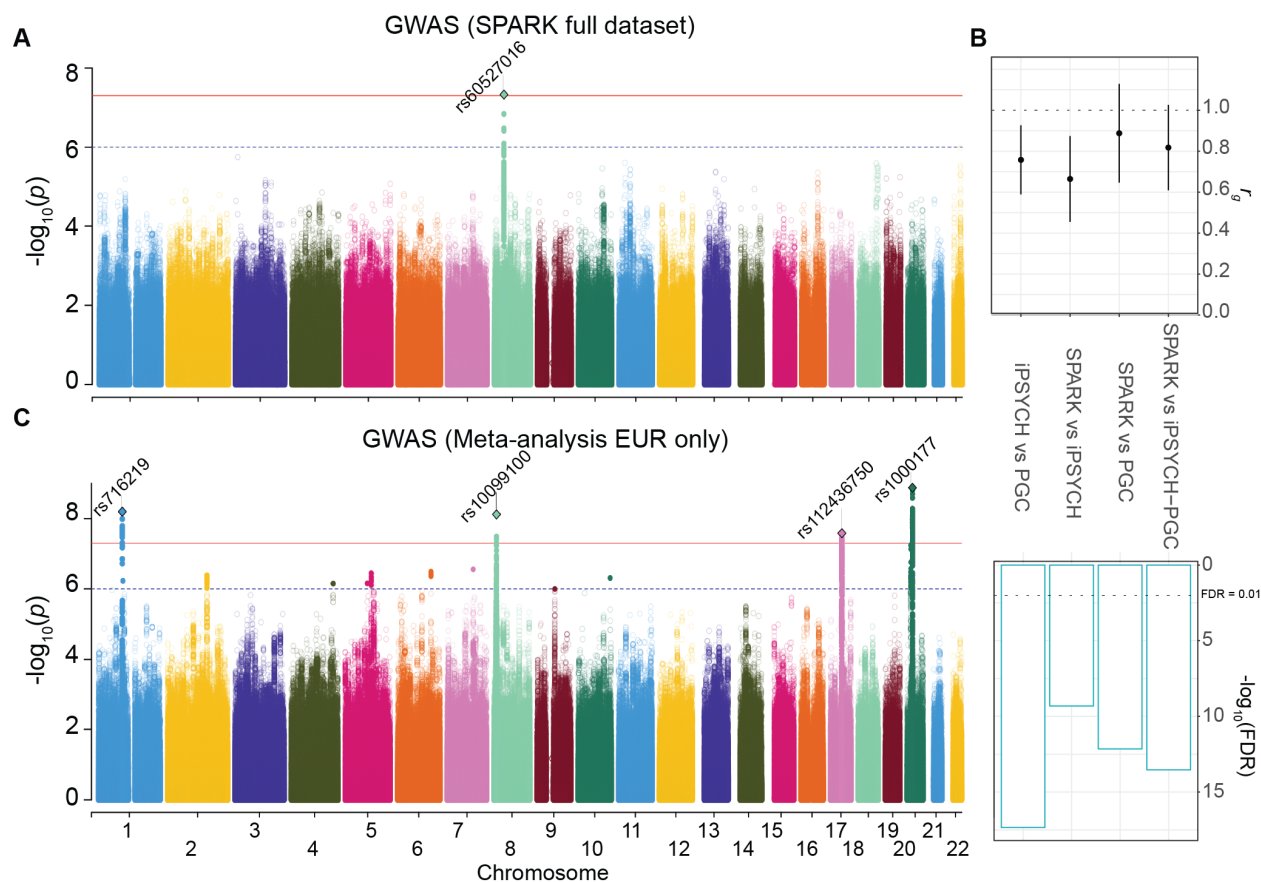
214



## 215 Results

### 216 GWAS in SPARK dataset identified a new locus associated with ASD risk

217 We obtained genotype and clinical diagnosis of ASD via self- or parent-report from 27,290  
218 individuals who participated in the SPARK project (18). The majority of data comprised families,  
219 including those where both parents and multiple children were genotyped (quads;  $N=3,192$   
220 families), where both parents and one child were genotyped (trios;  $N=2,486$  families), or where  
221 one parent and one child were genotyped (duos;  $N=2,448$  families) (Supplementary Figure S1).  
222 Only 68 individuals were ascertained without family members (singletons). After genotyping  
223 quality control (Supplementary Methods), 375,918 variants from 26,883 individuals were retained.  
224 Because the SPARK dataset did not genotype unrelated controls, we created pseudocontrols  
225 from the alleles not transmitted from parents to probands (36). Case-pseudocontrol design  
226 requires genotyping of both parents, so singletons and duos were excluded from the analysis.  
227 Due to the diverse ancestry in the cohort (Supplementary Figure S2, Supplementary Table S1),  
228 genotypes of all individuals including pseudocontrols were imputed to a diverse reference panel  
229 (TOPMed Freeze 5b reference panel consisting of 125,568 haplotypes). After imputation quality  
230 control (Methods; Supplementary Figure S3), 90,051,896 autosomal SNPs were tested for  
231 association in 6,222 case-pseudocontrol pairs (SPARK full dataset) consisting of 4,956 males  
232 and 1,266 females from multiple ancestries including European ( $N = 4,535$ ), African ( $N = 37$ ),  
233 East Asian ( $N = 83$ ) and other ancestries/admixed individuals ( $N = 1,567$ ) (Supplementary Figure  
234 S1, Supplementary Table S2). We observed no inflation of test statistics ( $\lambda_{GC} = 1.00$ )  
235 (Supplementary Figure S3), indicating population stratification was well-controlled when using this  
236 case-pseudocontrol design. We identified two SNPs at one locus (index SNP: rs60527016-C; OR  
237 = 0.84,  $P = 4.70 \times 10^{-8}$ ) at genome-wide significance ( $P < 5.0 \times 10^{-8}$ ) (Figure 1A, Table 1), which  
238 were supported by the previous largest ASD GWAS (14) (OR = 0.95,  $P = 0.0047$ ) derived from  
239 the PGC and iPSYCH cohorts (Supplementary Figure S4).



240

241 **Figure 1. Genome-wide association of ASD in the SPARK dataset**

242 (A) GWAS result from SPARK full dataset ( $N_{\text{case+pseudocontrol}} = 12,444$ ). (B) Genetic correlations across ASD  
 243 GWAS. From left to right, iPSYCH versus PGC (50), SPARK EUR versus iPSYCH, SPARK EUR versus  
 244 PGC and SPARK EUR versus iPSYCH-PGC study (14). (C) GWAS results from the meta-analysis (SPARK  
 245 European population and iPSYCH-PGC,  $N_{\text{max\_case+control}} = 55,420$ ). For Manhattan plots (A, C), the x-axes  
 246 indicate the chromosomal position and y-axes indicate the significance of associations. The blue and red  
 247 lines denote the significance threshold at suggestive ( $P < 1 \times 10^{-6}$ ) and significant ( $P < 5 \times 10^{-8}$ ) levels. SNPs  
 248 at with  $P < 1 \times 10^{-6}$  are shown as a filled circle. Rs number indicates index SNPs from independent loci (1  
 249 MB apart from each other) at  $P < 1 \times 10^{-8}$ ). Index SNPs at  $P < 5 \times 10^{-8}$  are shown as diamonds.

250

251

252 **Table 1. Genome-wide significant loci associated with ASD risk**

253 Genome-wide significant and suggestive loci in any of the GWAS analyses and meta-analysis of SPARK

254 European ancestries and iPSYCH+PGC participants are shown.

SNP <sup>1</sup>	Position (hg38)	EA	OA	SPARK			iPSYCH+PGC		Meta(EUR) <sup>2</sup>	
				EAF	OR(95%CI)	P	OR(95%CI)	P	OR(95%CI)	P
<b>Genome-wide significant loci (<math>P &lt; 5 \times 10^{-8}</math>)</b>										
rs716219	1:96104001	T	C	0.34	1.08 (1.03 - 1.14)	0.003	1.08 (1.05 - 1.11)	$3.99 \times 10^{-7}$	1.08 (1.05 - 1.11)	$6.42 \times 10^{-9}$
rs10099100	8:10719265	C	G	0.31	1.08 (1.02 - 1.14)	0.008	1.09 (1.06 - 1.12)	$1.07 \times 10^{-8}$	1.08 (1.05 - 1.11)	$7.65 \times 10^{-9}$
rs60527016	8:38442106	C	T	0.21	0.84 (0.79 - 0.90)	$4.70 \times 10^{-8}$	0.95 (0.92 - 0.99)	0.00466	0.93 (0.91 - 0.96)	$3.05 \times 10^{-6}$
rs112436750	17:45887763	A	AT	0.21	1.07 (1.01 - 1.14)	0.027	1.09 (1.05 - 1.12)	$1.23 \times 10^{-6}$	1.09 (1.06 - 1.12)	$2.62 \times 10^{-8}$
rs1000177	20:21252560	T	C	0.24	1.08 (1.02 - 1.15)	0.014	1.10 (1.07 - 1.14)	$3.32 \times 10^{-9}$	1.09 (1.06 - 1.13)	$1.34 \times 10^{-9}$
<b>Suggestive loci (<math>5 \times 10^{-8} \leq P &lt; 1 \times 10^{-6}</math>)</b>										
rs6701243	1:98627228	A	C	0.38	0.99 (0.94 - 1.00)	0.610	0.93 (0.90 - 0.96)	$3.07 \times 10^{-7}$	0.94 (0.91 - 0.96)	$5.90 \times 10^{-7}$
rs6743102	2:158521946	G	A	0.34	0.94 (0.89 - 0.99)	0.021	0.94 (0.91 - 0.97)	$8.99 \times 10^{-6}$	0.94 (0.91 - 0.96)	$4.07 \times 10^{-7}$
rs33966416	4:170285452	CA	C	0.50	0.95 (0.90 - 1.00)	0.038	0.94 (0.91 - 0.96)	$2.73 \times 10^{-6}$	0.94 (0.92 - 0.96)	$6.99 \times 10^{-7}$
rs4916723	5:88558577	A	C	0.40	1.10 (1.00 - 1.10)	0.062	1.07 (1.04 - 1.10)	$1.92 \times 10^{-6}$	1.07 (1.04 - 1.09)	$6.90 \times 10^{-7}$
rs416223	5:104655775	C	A	0.40	1.00 (0.96 - 1.10)	0.730	1.07 (1.04 - 1.10)	$3.84 \times 10^{-7}$	1.07 (1.04 - 1.09)	$3.56 \times 10^{-7}$
rs67248478	6:134711094	C	T	0.34	0.94 (0.90 - 1.10)	0.032	0.94 (0.91 - 0.96)	$3.22 \times 10^{-6}$	0.94 (0.91 - 0.96)	$3.22 \times 10^{-7}$
rs76569799	9:73565191	C	T	0.15	1.10 (0.99 - 1.10)	0.076	1.09 (1.05 - 1.13)	$3.90 \times 10^{-6}$	1.08 (1.05 - 1.12)	$9.99 \times 10^{-7}$
rs4750990	10:128689762	T	C	0.36	1.00 (0.98 - 1.10)	0.250	1.07 (1.04 - 1.10)	$1.37 \times 10^{-6}$	1.07 (1.04 - 1.09)	$4.89 \times 10^{-7}$
rs2224274	20:14780101	C	T	0.43	1.00 (0.97 - 1.10)	0.310	1.07 (1.04 - 1.1)	$2.86 \times 10^{-7}$	1.07 (1.05 - 1.10)	$5.56 \times 10^{-8}$

255 <sup>1</sup>Index SNPs from loci that survived genome-wide significance in any of the GWASs including meta-analysis.

256 <sup>2</sup>Meta-analysis of SPARK European ancestries and iPSYCH+PGC

257 EA: Effect allele, OA: Other allele; EAF: Effect allele frequency in SPARK full dataset

258

259 **Replication of genetic risk factors for ASD**

260 Given the phenotypic heterogeneity of ASD and potential technical differences such as

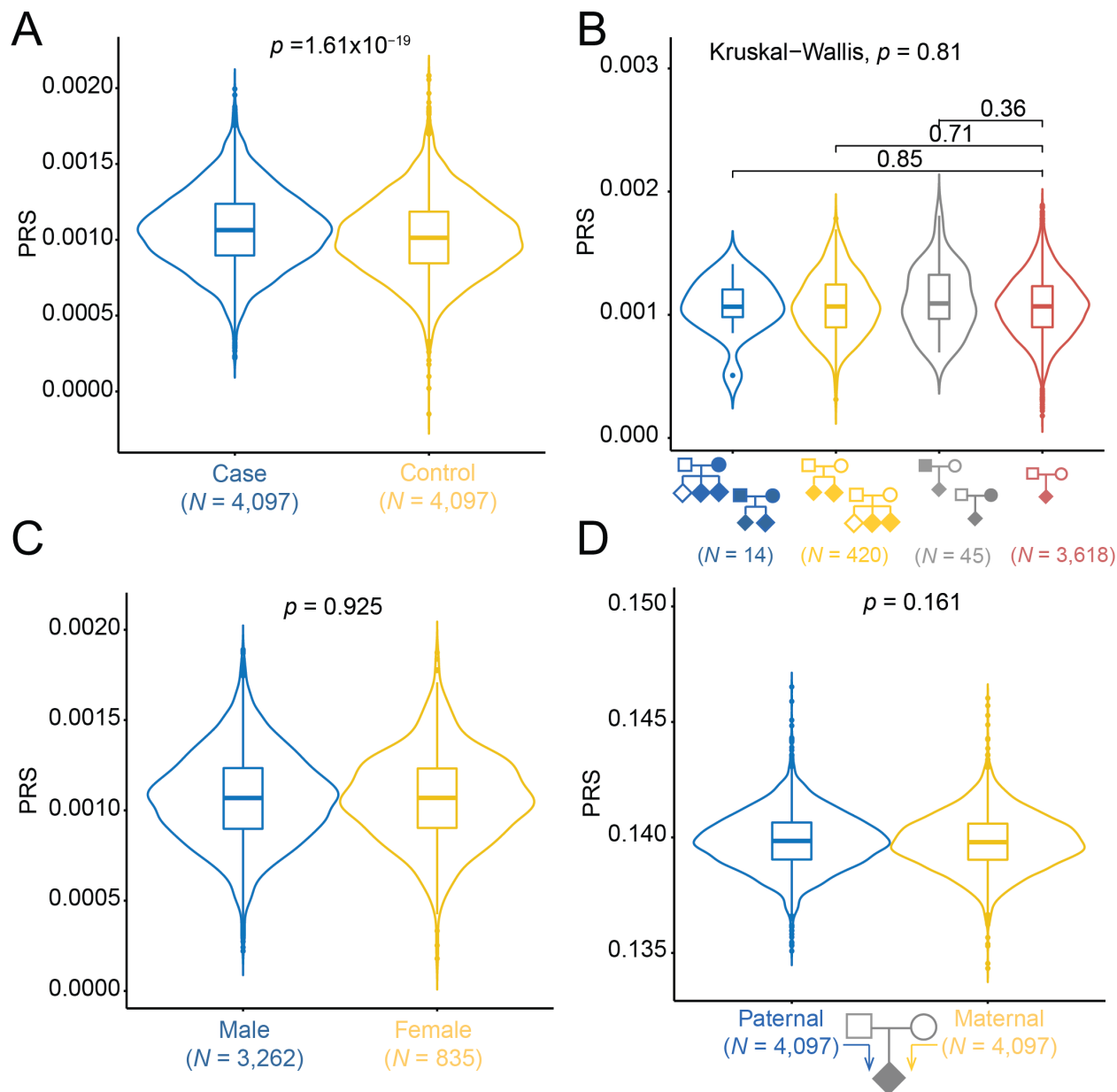
261 genotyping platforms or data processing, we assessed the replication of genetic risk factors

262 across cohorts by comparing previous major ASD studies including PGC and iPSYCH cohort

263 (14,50) with the SPARK dataset subset to individuals of European descent (EUR) (Figure 1B).  
264 Although each study included multiple ASD subtypes including ASD from DSM5, Asperger's,  
265 autism/autistic disorder, and Pervasive Developmental Disorder - Not otherwise specified (PDD-  
266 NOS) from DSM IV, and approaches differed across these samples from requiring community  
267 diagnosis to best-estimate diagnosis based on standardized assessment, we obtained high genetic  
268 correlations between the SPARK EUR dataset and the largest iPSYCH-PGC GWAS ( $r_g = 0.82$ ;  $P$   
269  $= 5.27 \times 10^{-14}$ ), suggesting the genetic risk factors for autism are largely shared among different  
270 ASD GWAS and are generalizable despite differences in diagnostic criteria and batch effects.  
271 We next performed meta-analysis with SPARK EUR samples and iPSYCH-PGC samples ( $N_{case}$   
272  $= 18,381$  and  $N_{control} = 27,969$ ) to maximize power. The meta-analysis identified four additional  
273 loci associated with risk for ASD (Supplementary Figure S5-8). These included three previously  
274 reported loci (14) and one novel locus on chromosome 17, where a gene-based test from the  
275 iPSYCH-PGC study has previously shown association with risk for ASD (14) (Figure 1C, Table 1,  
276 Supplementary Figure S7). This novel locus was also reported to be associated with more than  
277 60 phenotypes including neuroticism (51–55), educational attainment (56) and intracranial volume  
278 (57) (index SNPs  $r^2 > 0.8$  in 1KG EUR) (Supplementary Table S3) indicating highly pleiotropic  
279 effects at this locus. The SNP based heritability in SPARK EUR samples was estimated ( $h^2_G$ ) to  
280 be 0.117 (s.e. = 0.0082) for population prevalence of 0.012 (14,58) which was comparable with  
281 the previous report ( $h^2_G = 0.118$ ; s.e. = 0.010) (14).  
282 The generalization of effects across ancestries for the five index SNPs identified (Table 1) was  
283 examined (Supplementary Figure S2, S3, and Supplementary Table S4). The association results  
284 from the cross-ancestry dataset were mainly driven by the European population, as expected  
285 given the larger sampling from this population. We found that some regions showed population-  
286 differentiated allele frequencies. For example, rs10099100 was more common in European and  
287 African populations (MAF = 0.33, 0.39 from tested samples, respectively) than in East Asians

288 (MAF = 0.02 from tested samples), necessitating a further investigation of genetic risk factors for  
289 ASD in populations of diverse ancestry (59,60).

290 The generalization of genetic effects on risk for ASD was also confirmed by polygenic risk scores  
291 (PRSs) derived from the iPSYCH-PGC GWAS that showed higher scores in SPARK EUR cases  
292 ( $N = 4,097$ ) compared to pseudocontrols ( $N = 4,097$ ) ( $P = 1.61 \times 10^{-19}$ ) (Figure 2A).



293

294 **Figure 2. Comparison of polygenic risk scores between subgroups.**

295 **(A)** Common variant risk burden is higher in cases compared to pseudocontrols. **(B)** Comparison of PRS  
296 across family types (from left to right, families with multiple affected children with affected parent(s), multiple  
297 affected children with unaffected parents, one affected child with affected parent(s), and one affected child  
298 with unaffected parents) shows no evidence for a higher common variant burden in multiplex families. **(C)**  
299 Comparison of PRS between male and female probands shows no evidence of enrichment of common  
300 variants impacting risk for ASD in females. **(D)** There is no evidence for a difference in the transmission of  
301 common variant risk burden from mother versus father.

## 302 Investigation of common variant burden impacting risk for ASD

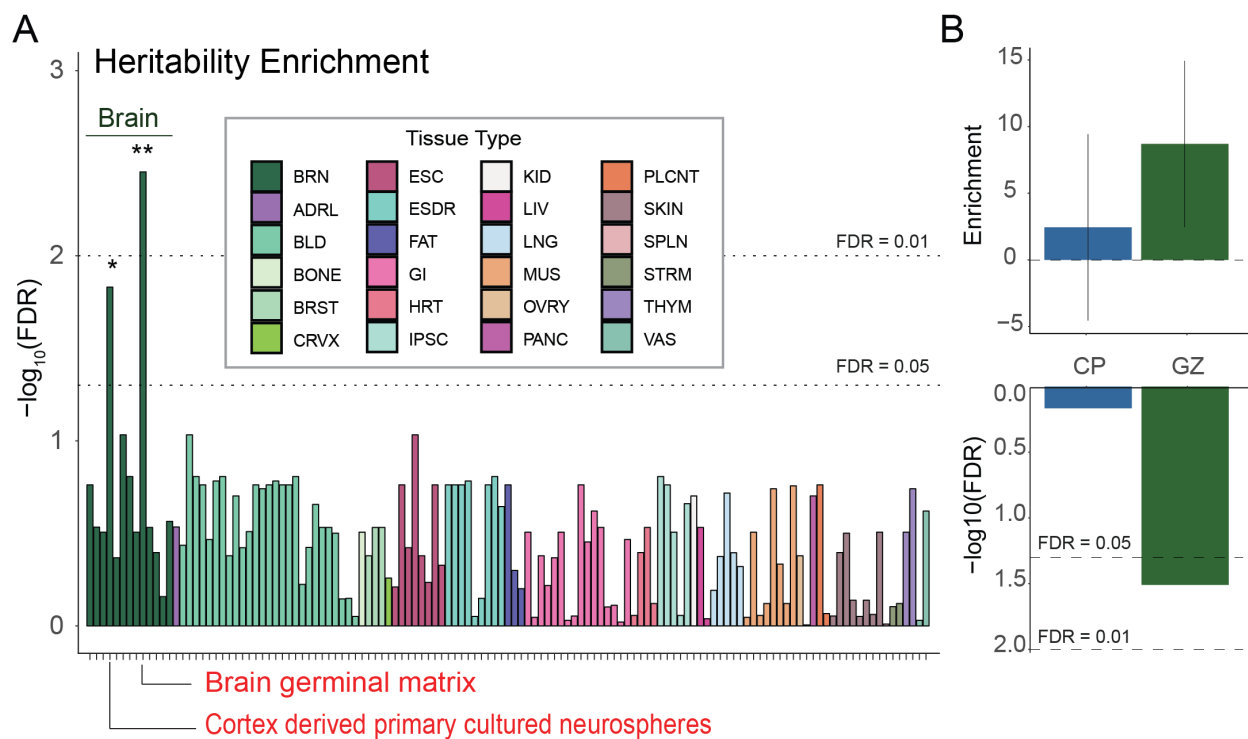
303 We next used PRSs to compare common variant risk burden among family types, sex, and parent  
304 of origin (Figure 2B-D). Because ASD families with multiple affected siblings were shown to have  
305 different segregation patterns compared with simplex families that have a higher burden of *de*  
306 *novo* mutations (12,61,62), we compared the distribution of PRSs across four family types (Figure  
307 2B, Supplementary Table S1). Our results showed no evidence for a difference in common variant  
308 burden impacting risk for ASD in multiplex families as compared to simplex families. We note that  
309 multiplex/simplex status was indicated by either enrollment or self-report in a questionnaire and  
310 may be underestimated due to missing survey data.

311 As the prevalence of ASD is higher in males than in females (OR = 4.20) (63), and previous  
312 studies have reported that females with ASD have a higher burden of *de novo* variants (9,64–66),  
313 we also investigated the potential contribution of common variants to the female protective effect  
314 by comparing PRS between sexes. We did not find any evidence that ASD common variant risk  
315 burden differs in females compared to males (Figure 2C).

316 A previous study hypothesized that a new mutation in a mother, who is less susceptible to  
317 developing autism because of the female protective effect, may be more likely to transmit risk  
318 factors to their children with ASD (67). We, therefore, examined the over-transmission of common  
319 variant risk for ASD from mother to offspring. We found no evidence of the over-transmission of  
320 common variant risk burden from either mothers or fathers to their affected children (Figure 2D).

### 321 Contribution of cortical development to risk for ASD

322 Previous studies suggest an important role of brain development in ASD (14,68). To characterize  
 323 tissue types relevant to risk for ASD, we next evaluated heritability enrichment within active  
 324 enhancer or promoter regions in different tissues (69) (Figure 3A, Table S4). Significant  
 325 enrichment of heritability was observed in regulatory elements of brain germinal matrix as well as  
 326 primary cultured neurospheres from the fetal cortex (FDR = 0.004 and 0.015, respectively),  
 327 suggesting that disruption of gene regulation in these tissues increases the risk for ASD. We  
 328 further examined SNP heritability in the developing cortex using differentially accessible peaks  
 329 between the neuron-enriched cortical plate and the progenitor-enriched germinal zone (70)  
 330 (Figure 3B). We found significant enrichment in peaks more accessible in the germinal zone (FDR  
 331 = 0.008), but not in the cortical plate, providing further evidence for genetically mediated  
 332 alterations in cortical development playing a crucial role in ASD etiology.



333  
 334 **Figure 3. Partitioned heritability enrichment of tissues implicates cortical development in ASD risk**

335 **(A)** Heritability enrichment in active enhancer and promoter regions present in different tissues shows the  
336 critical role of developing the brain in ASD etiology. **(B)** Heritability enrichment in differential chromatin  
337 accessibility from the developing fetal cortical wall. (upper) The x-axis represents tissue types and the y-  
338 axis indicates heritability enrichment. The error bar shows a 95% confidence interval. (lower) The x-axis  
339 represents tissue types and the y-axis indicates statistical significance as  $-\log_{10}(\text{FDR})$ . BRN: Brain, ADRL:  
340 Adrenal, BLD: Blood, BRST: Breast, CRVX: Cervix, ESDR: ESC\_derived, GI: GI\_duodenum, GI\_colon,  
341 GI\_rectum, GI\_stomach, GI\_intestine, GI\_colon, GI\_rectum, GI\_duodenum and GI\_esophagus, HRT:  
342 Heart, KID: Kidney, LIV: Liver, LNG: Lung, MUS: Muscle, OVRY: Ovary, PANC: Pancreas, PLCNT:  
343 Placenta, SPLN: Spleen, STRM: Stromal connective, THYM: Thymus, VAS: Vascular, CP: Peaks more  
344 accessible in cortical plate, GZ: Peaks more accessible in germinal zone. \* FDR < 0.05, \*\* FDR < 0.01.

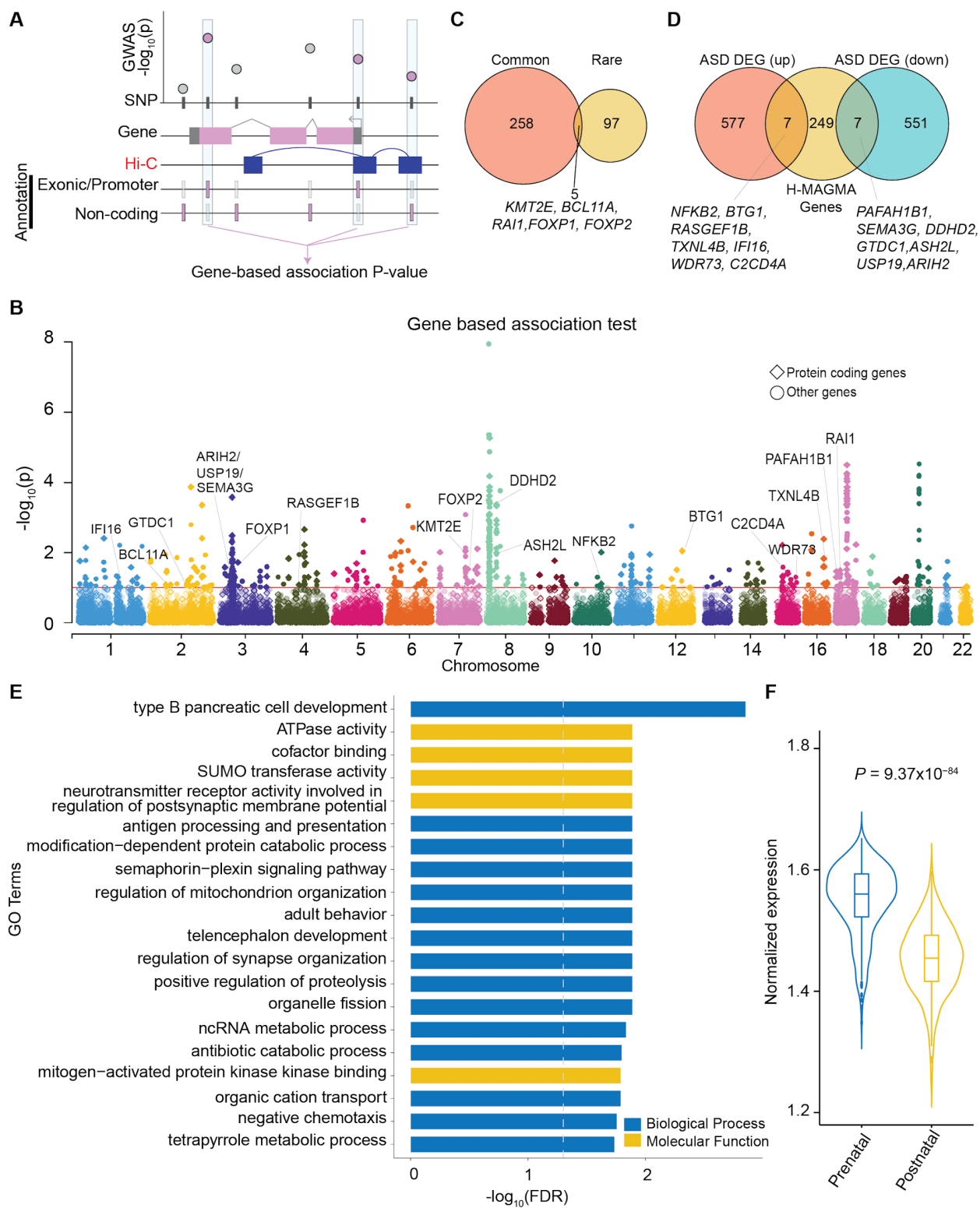
345

## 346 H-MAGMA identified genes and pathways impacting risk for ASD

347 To identify genes associated with risk for ASD from meta-analysis (EUR only), we applied Hi-C  
348 coupled MAGMA (H-MAGMA) (31), which aggregates SNP-level P-values into a gene-level  
349 association statistic with an additional assignment of non-coding SNPs to their chromatin-interacting  
350 target genes generated from fetal brain Hi-C (46) (Figure 4A). We identified 567 genes associated  
351 with ASD (FDR < 0.1), including 263 protein coding genes (Figure 4B, Table S5). Five genes  
352 implicated from common variant evidence (*KMT2E*, *RAI1*, *BCL11A*, *FOXP1*, and *FOXP2*) also  
353 harbored an excess of rare variants associated with ASD (11). This overlap between rare and common  
354 ASD risk variants was more than expected by chance (hypergeometric  $P = 0.01$ ; Figure 4C),  
355 corroborating the previous result that common and rare variation converges on the same genes and  
356 pathways (31,71). We also found that 14 H-MAGMA genes were also differentially expressed in the  
357 post-mortem cortex between individuals with ASD and neurotypical controls (up-regulated in ASD:  
358 *NFKB2*, *BTG1*, *RASGEF1B*, *TXNL4B*, *IFI16*, *WDR73* and *C2CD4A*; down-regulated in ASD:  
359 *PAFAH1B1*, *SEMA3G*, *DDHD2*, *GTDC1* *ASH2L*, *USP19* and *ARIH2*; FDR < 0.05) (72) (Figure 4D).  
360 Rank-based gene ontology enrichment analysis (73) suggested that ASD risk genes were enriched in  
361 188 terms including telencephalon development and regulation of synapse organization (Figure 4E,  
362 Supplementary Table S7).



363 Since heritability enrichment analyses suggested genetically mediated impacts on cortical  
364 development contribute to ASD risk (Figure 3), we explored whether the expression level of ASD  
365 risk genes from H-MAGMA is different between prenatal and postnatal cortex. As shown  
366 previously (14,31), we found ASD risk genes exhibited higher expression in the prenatal cortex  
367 as compared to the postnatal cortex ( $P = 2.77 \times 10^{-62}$ ) (Figure 4F) (74). In particular, the expression  
368 level of ASD risk genes was highest between 20 and 30 post-conception weeks (Supplementary  
369 Figure S9). Taken together, our results demonstrate common risk variants for ASD play an  
370 important role in the developing cortex.



371

372 **Figure 4. H-MAGMA identified 263 protein-coding genes linked to ASD**

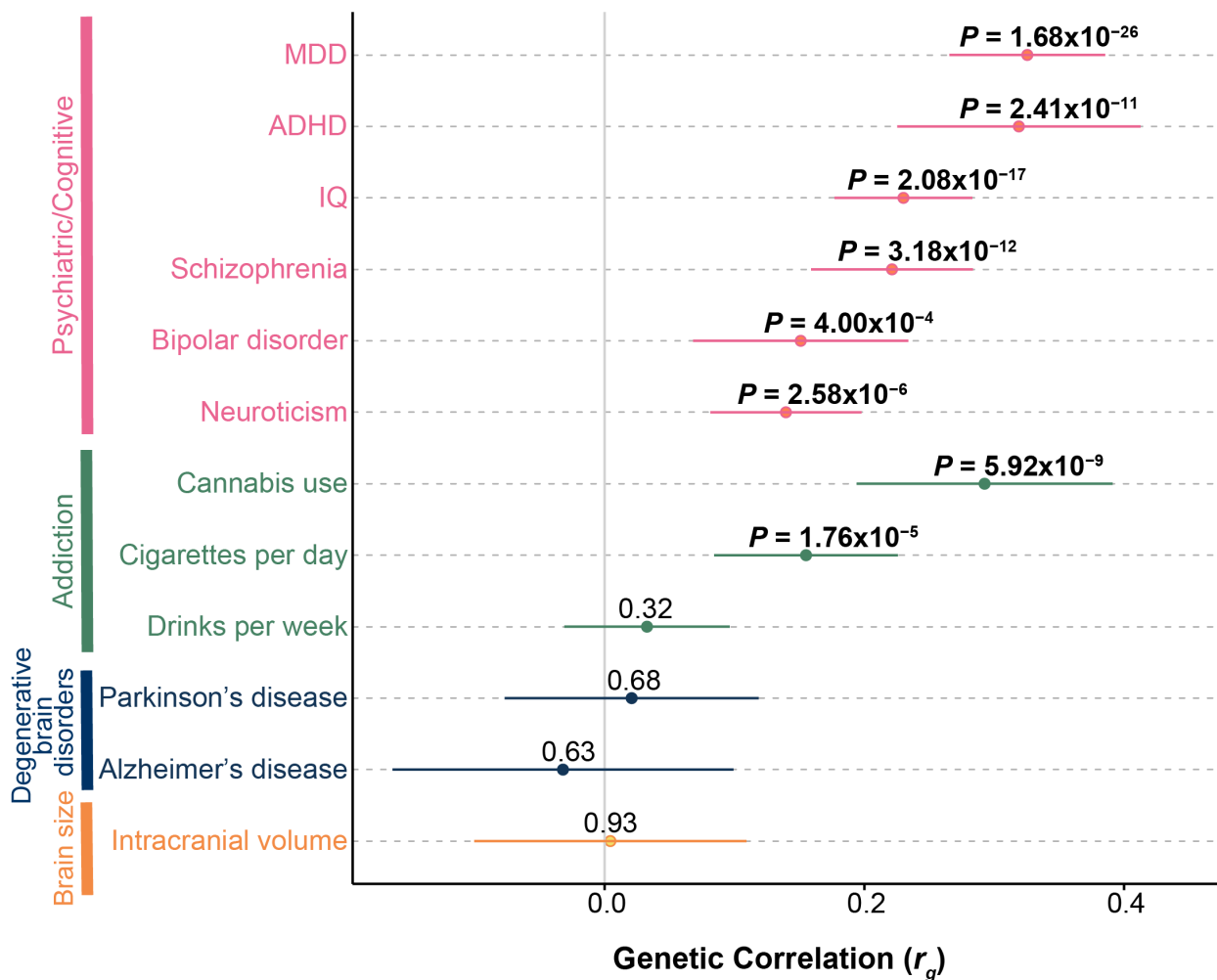
373 **(A)** Schematic diagram of H-MAGMA. SNP based association P-values were aggregated to gene-based P-values using positional information as well as chromatin interaction in the fetal brain. **(B)** Gene-based

374

375 association result from H-MAGMA. The x-axis indicates the start position of genes (hg19). **(C)** Overlap of  
376 ASD risk genes harboring common and rare variants (11). **(D)** Overlapped genes with differential  
377 expression from post-mortem brains in individuals with ASD patients and neurotypical controls (72). **(E)**  
378 Gene ontologies enriched for ASD linked genes (top 20). **(F)** Developmental expression pattern of ASD risk  
379 genes (74).

## 380 Genetic correlation between ASD and 12 brain and behavioral phenotypes

381 Both epidemiological studies and genetic studies suggested the phenotypic comorbidity (75–78)  
382 or genetic correlation (14,79) of ASD with various brain and behavioral phenotypes. Thus, we  
383 evaluated the pleiotropic effect of ASD risk SNPs with twelve other brain and behavioral  
384 phenotypes (47,54,57,80–87) (Figure 5, Supplementary Table S8). We observed a novel genetic  
385 correlation between ASD and cigarettes per day ( $r_g = 0.16$ ,  $P = 8.80 \times 10^{-5}$ ), indicating a partially  
386 shared genetic basis between risk for ASD and addictive smoking behavior. We also replicated  
387 positive genetic correlations previously detected for seven phenotypes (FDR < 0.05) (14),  
388 providing further support for a shared genetic basis of multiple neuropsychiatric disorders (79,88).



389

390 **Figure 5. Genetic correlation of ASD against twelve brain and behavioral phenotypes**

391 The x-axis represents an estimate of the genetic correlation ( $r_g$ ). Error bars represent the 95% confidence  
 392 interval. P-values at FDR < 0.05 are shown in bold. MDD: Major depressive disorder, ADHD: Attention-  
 393 Deficit/Hyperactivity Disorder.

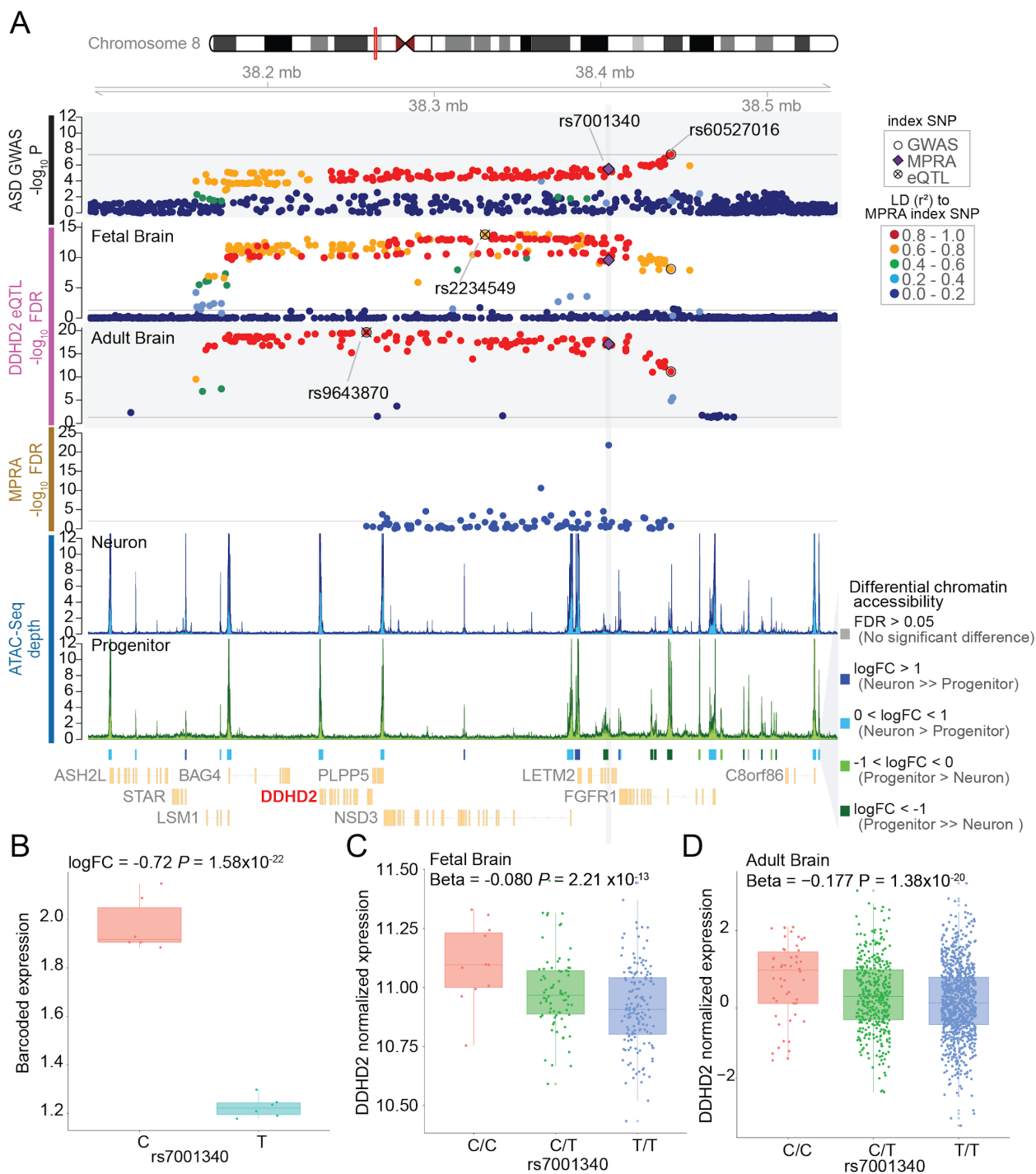
394

395 **Functional validation to fine-map causal variants and prioritize genes**

396 Interestingly, the novel locus identified by the SPARK full dataset (rs60527016 at chr8:38.19M -  
 397 chr8:38.45M, Figure 1A, 6A) was also identified as a pleiotropic locus in a recent cross-disorder  
 398 meta-analysis on eight psychiatric disorders (79). This locus was not only associated with ASD  
 399 but also with schizophrenia, bipolar disorder and obsessive-compulsive disorder (OCD),  
 400 suggesting that understanding the regulatory mechanism at this locus may reveal the basis for

401 pleiotropic effects across psychiatric disorders. The index SNP (rs60527016) was located within  
402 a 300 kb LD block ( $r^2 > 0.5$  in SPARK full dataset) that contains seven genes (Figure 6A). To  
403 prioritize causal variants within this locus, we performed a massively parallel reporter assay  
404 (MPRA) (23,24) on 98 credible SNPs in this region in HEK 293 cells. MPRA measures barcoded  
405 transcriptional activity driven by each allele in a high-throughput fashion (Supplementary Figure  
406 S13). Surprisingly, SNP rs7001340 exhibited the strongest allelic difference in barcoded  
407 expression ( $P = 1.51 \times 10^{-24}$ ) even though it is 37 kb away from the GWAS index SNP ( $r^2 = 0.85$   
408 with the index SNP in SPARK full dataset) (Figure 6A, B, Supplementary Table S9),  
409 demonstrating the regulatory potential of this SNP and suggesting its causal role in psychiatric  
410 disorders, including ASD. While MPRA was performed in HEK cells, the SNP was located in a  
411 regulatory element with higher chromatin accessibility in human neural progenitors compared with  
412 postmitotic neurons (Figure 6A) (unpublished data from Stein lab), indicating its regulatory  
413 potential in the developing brain. The risk allele (T) at this SNP was associated with  
414 downregulation of barcoded expression in MPRA (Figure 6B), and was predicted to disrupt two  
415 transcription factor binding motifs (89) (TBX1 and SMARCC1) (Supplementary Figure S14),  
416 providing a possible mechanism of action of this variant. We next investigated potential target  
417 genes impacted by regulatory changes at this SNP by using eQTL data from fetal (33) and adult  
418 brain tissues (32). Expression levels of three eGenes were significantly associated with this SNP  
419 (*DDHD2* from the fetal brain and *DDHD2*, *LSM1*, *LETM2* from the adult brain) (Figure 6A). Of  
420 these three genes, two genes (*DDHD2*, *LETM2*) showed the direction of the effect expected from  
421 the MPRA result (risk allele downregulates the eGene) (Supplementary Figure S15). It is of note  
422 that *DDHD2* was identified in both fetal and adult brain eQTL datasets (beta = -0.080,  $P =$   
423  $2.212 \times 10^{-13}$ ; beta = -0.177,  $P = 1.38 \times 10^{-20}$ , respectively; Figure 6C, D). Notably, *DDHD2* was also  
424 significantly downregulated in the post-mortem cortex of individuals with autism (logFC = -0.28,  
425 FDR = 0.013), providing an added layer of evidence supporting its role in ASD risk (72). *DDHD2*  
426 also was identified by H-MAGMA (Figure 4F), and a copy number variation (CNV) containing

427 *DDHD2* (deletions) was found in proband-sibling pairs with discordant social-behavior  
428 phenotypes (90). Collectively, by integrating existing multi-level functional genomic resources and  
429 an experimental fine-mapping approach using MPRA, we suggest *DDHD2* as a strong candidate  
430 gene impacting risk for ASD.  
431



432

433 **Figure 6. Identification of putative causal variant and gene impacting risk for ASD**

434 **(A)** Annotated locus plot near rs60527016 ASD risk index variant, from top panel to bottom, ASD  
 435 associations within SPARK full dataset ( $n = 6,222$  case-pseudocontrol pairs), eQTL for DDHD2 in fetal  
 436 brains ( $n = 235$ ) and adult brain ( $n = 1,387$ ), MPRA expression ( $n=6$ ), ATAC-seq averaged depth in neuron  
 437 ( $n = 61$ ) and progenitor ( $n = 73$ ). Differential open chromatin accessibility peaks from ATAC-seq, and gene

438 model (NCBI Refseq). LD was calculated to rs7001340 within SPARK parents of cases, fetal brain donors,  
439 or 1KG EUR and colored accordingly. **(B)** The barcoded expression level of mRNA based from each allele  
440 at rs7001340 from the MPRA experiment. **(C)** The expression level of DDHD2 by rs7001340 genotypes in  
441 the fetal brain. **(D)** The expression level of DDHD2 by rs7001340 genotypes in adult brain. Individuals with  
442 allele dosage (0-0.1 as C/C, 0.9-1.1 as C/T, 1.9-2.0 as T/T) are shown. For **(B)** to **(D)**, ASD risk allele for  
443 rs7001340 is T and protective allele is C.

## 444 Discussion

445 In this study, we increased sample sizes for ASD GWAS to  $N_{\text{case}(\text{max})} = 24,063$ ,  $N_{\text{control}(\text{max})} = 34,191$   
446 and identified five loci associated with risk for ASD (four from European only meta-analysis, one  
447 locus from SPARK project alone), including two new loci (marked by index SNPs rs60527016 and  
448 rs112436750). These loci have pleiotropic effects on multiple psychiatric disorders including  
449 schizophrenia (for rs60527016 and rs112436750), bipolar disorder, and OCD (for rs60527016).

450  
451 Using a PRS derived from a previous study (14), we found enrichment of risk variants in SPARK  
452 cases, indicating the contribution of common genetic risk factors to ASD is consistent across  
453 cohorts. However, despite several hypotheses that rare variants associated with risk for ASD are  
454 enriched in certain subgroups of individuals with ASD, such as in females compared to males  
455 (female protective effect) (9,64–66,91), multiplex families compared to simplex families (12,61,62),  
456 or maternal alleles compared to paternal alleles (10,67), we do not find evidence to support the  
457 increased burden of common risk variants within those subgroups. This result indicates potential  
458 rare and common variant differences in contribution to subgroup risk for ASD. However, given the  
459 small sample size of specific subgroups ( $N = 835$  in female whereas  $N = 3,262$  in male,  $N = 14$   
460 for families with multiple affected children vs  $N = 3,618$  with one affected children), our study may  
461 have limited power to identify the differences among subgroups. Thus, a larger sample size would  
462 be warranted to compare the difference in the role of common variants within these categories.

463



464 Identifying locations in the genome associated with risk for ASD does not in itself lead to insights  
465 into what tissues or developmental time points are crucial for the etiology of ASD. Here, by  
466 integrating SNP association statistics with existing annotations of regulatory elements active  
467 during specific developmental time periods or within specific brain regions, we found an excess  
468 of common genetic risk for ASD in the fetal brain regulatory elements (brain germinal matrix and  
469 primary cultured neurospheres from the fetal cortex), and progenitor enriched germinal zone of  
470 the developing cortex, confirming previous findings that alterations of gene regulation in the  
471 prenatal cortex play a key role in ASD etiology (14).

472

473 To further understand the specific genes leading to risk for ASD, we applied a recently developed  
474 platform, H-MAGMA (31) and identified 263 putative candidate protein-coding risk genes. H-  
475 MAGMA genes are highly expressed in the prenatal brain, similar to the enrichment of ASD risk  
476 genes with rare variations during neurodevelopment (92). This result suggests potential molecular  
477 convergence regardless of classes of mutation, which is supported by five genes (previously  
478 identified *KMT2E* and newly identified *RAI1*, *BCL11A*, *FOXP1*, and *FOXP2*) that are affected by  
479 both rare and common variation.

480

481 Since identification of a GWS locus does not elucidate the causal variant(s), we performed MPRA  
482 and identified a potential causal SNP (rs7001340) at a novel ASD locus discovered in the SPARK  
483 sample. Interestingly, the individual variant with the strongest regulatory effect (rs7001340;  $r^2=$   
484 0.85 with the index SNP in SPARK full dataset) was different from the SNP with the strongest  
485 association with ASD (rs60527016), highlighting the importance of experimental validation in  
486 identifying causal variants. It should be noted that the regulatory effects of these variants were  
487 assessed in non-neural (HEK) cells, so further validation of these effects in ASD-relevant cell  
488 types would provide increased confidence in declaring this SNP as causal. The experimentally  
489 validated regulatory SNP (rs7001340) is in the intron of *LETM2*, and is also an eQTL for *LETM2*,

490 *LSM1* (247 kb away) and *DDHD2* (173 kb away), indicating that the SNP functions as a distal  
491 regulatory element. The risk allele (T) was associated with decreased expression of barcoded  
492 transcripts in the MPRA and downregulation of *DDHD2* from eQTL in both fetal and adult brains,  
493 implying a consistent direction of the allelic effects on gene regulation. The risk allele showed the  
494 same direction of effect for *LETM2* in adult brain tissue, but was not significantly associated in  
495 fetal brain tissue (P-values = 0.33). Intriguingly, *DDHD2* was also downregulated in the cortex  
496 from individuals with ASD compared to neurotypical controls (72), providing an additional level of  
497 support for this gene as a risk factor for ASD. *DDHD2* (DDHD domain-containing protein 2), also  
498 known as KIAA0725p, encodes a phospholipase and is localized in the Golgi (93). *DDHD2* plays  
499 a role in the efficient regulation of membrane trafficking between the Golgi and cytosol (93) and  
500 is highly expressed in the brain (94–96). Mutations in *DDHD2* have been previously found in  
501 individuals with spastic paraplegia type 54 (SPG) (96–98). *Ddhd2* null mice exhibited motor and  
502 cognitive impairments (99), which are frequent comorbidities of ASD (100). We, therefore,  
503 conclude *DDHD2* is a strong candidate risk gene for ASD through multiple lines of evidence.

504  
505 There is still a large amount of common variant heritability not explained in this study indicating  
506 that further increases in sample size will be necessary to explain the common inherited  
507 component of ASD risk. While the combination of TOPMed imputation and the case-  
508 pseudocontrol study model enabled us to include individuals from multiple ancestries, the case-  
509 pseudocontrol model is lower powered compared to case-unscreened control models because a  
510 pseudocontrol might have greater liability for ASD than the average individual in the population  
511 (101). In addition, the case-pseudocontrol model cannot incorporate duos or singletons due to  
512 the lack of parental genotype information, which resulted in over 2,000 individuals with ASD with  
513 genotyping information in the SPARK project not being included in our analysis. Future studies  
514 could potentially increase power by including all cases available in SPARK and using unscreened  
515 population matched controls (102). Secondly, subsequent analyses including PRS, LDSC

516 regression, and H-MAGMA were limited to individuals from European ancestries only, because  
517 most resources and software are designed to be used only within one population, generally  
518 European ancestry (103). Including other ancestries for these analyses will be able to uncover  
519 risk factors shared or specific to existing human populations.

520  
521 In summary, ASD GWAS in the SPARK dataset and meta-analysis with previous GWAS identified  
522 two new susceptibility loci. By integrating existing multi-level functional genomic resources and  
523 experimental tools such as MPRA and eQTL, we highlight *DDHD2* as a novel high confidence  
524 ASD risk gene impacted by a distal common variant within a regulatory element present in neural  
525 progenitors of the developing cortex. This strategy can be broadly applied to common variant risk  
526 loci of multiple neuropsychiatric disorders to identify causal variant(s), regulatory regions, cell-  
527 types, and genes whose misregulation leads to risk for neuropsychiatric disorders.

528

## 529 Acknowledgments

530 This work was supported by a grant from the Simon Foundation (SFARI[Award #: 605259], H.W  
531 and J.L.S), NIMH (R00MH113823 and DP2MH122403 to H.W. and R01MH118349,  
532 R01MH120125, R01MH121433 to J.L.S), NIGMS (DP2GM114829 to S.K.) and the NARSAD  
533 Young Investigator Award (H.W.). LMR was supported by T32 HL129982. We are grateful to all  
534 of the families in SPARK, the SPARK clinical sites and SPARK staff. We appreciate obtaining  
535 access to genetic and phenotypic data on SFARI Base. Approved researchers can obtain the  
536 SPARK population dataset described in this study (more details available at  
537 <https://www.sfari.org/resource/spark/>) by applying at <https://base.sfari.org>.  
538 Trans-Omics in Precision Medicine (TOPMed) program imputation panel (version Freeze5)  
539 supported by the National Heart, Lung and Blood Institute (NHLBI); see [www.nhlbiwgs.org](http://www.nhlbiwgs.org).

540 TOPMed study investigators contributed data to the reference panel, which can be accessed  
541 through the Michigan Imputation Server; see <https://imputationserver.sph.umich.edu>. The panel  
542 was constructed and implemented by the TOPMed Informatics Research Center at the University  
543 of Michigan (3R01HL-117626-02S1; contract HHSN268201800002I). The TOPMed Data  
544 Coordinating Center (3R01HL-120393-02S1; contract HHSN268201800001I) provided additional  
545 data management, sample identity checks, and overall program coordination and support. We  
546 gratefully acknowledge the studies and participants who provided biological samples and data for  
547 TOPMed. Summary statistics from the GWAS conducted here will be available upon publication.  
548 Code is available at [https://github.com/thewonlab/GWAS\\_ASD\\_SPARK](https://github.com/thewonlab/GWAS_ASD_SPARK).

## 549 Disclosures

550 All authors report no financial interests or potential conflicts of interest.

## 551 References

552

- 553 1. American Psychiatric Association (2013): *Diagnostic and Statistical Manual of Mental*  
554 *Disorders (DSM-5®)*. American Psychiatric Pub.
- 555 2. Lee SH, Ripke S, Neale BM, Faraone SV, Purcell SM, Perlis RH, *et al.* (2013): Genetic  
556 relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat*  
557 *Genet* 45: 984–994.
- 558 3. Gaugler T, Klei L, Sanders SJ, Bodea CA, Goldberg AP, Lee AB, *et al.* (2014): Most genetic  
559 risk for autism resides with common variation. *Nat Genet* 46: 881–885.
- 560 4. de la Torre-Ubieta L, Won H, Stein JL, Geschwind DH (2016): Advancing the understanding  
561 of autism disease mechanisms through genetics. *Nat Med* 22: 345–361.
- 562 5. Sandin S, Lichtenstein P, Kuja-Halkola R, Hultman C, Larsson H, Reichenberg A (2017): The  
563 Heritability of Autism Spectrum Disorder. *JAMA* 318: 1182–1184.

- 564 6. Bai D, Yip BHK, Windham GC, Sourander A, Francis R, Yoffe R, *et al.* (2019): Association of  
565 Genetic and Environmental Factors With Autism in a 5-Country Cohort. *JAMA Psychiatry*.  
566 <https://doi.org/10.1001/jamapsychiatry.2019.1411>
- 567 7. Michaelson JJ, Shi Y, Gujral M, Zheng H, Malhotra D, Jin X, *et al.* (2012): Whole-genome  
568 sequencing in autism identifies hot spots for de novo germline mutation. *Cell* 151: 1431–  
569 1442.
- 570 8. O’Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, *et al.* (2012): Sporadic  
571 autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature*  
572 485: 246–250.
- 573 9. Sanders SJ, He X, Willsey AJ, Ercan-Sencicek AG, Samocha KE, Cicek AE, *et al.* (2015):  
574 Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk  
575 Loci. *Neuron* 87: 1215–1233.
- 576 10. Leppa VM, Kravitz SN, Martin CL, Andrieux J, Le Caignec C, Martin-Coignard D, *et al.*  
577 (2016): Rare Inherited and De Novo CNVs Reveal Complex Contributions to ASD Risk in  
578 Multiplex Families. *Am J Hum Genet* 99: 540–554.
- 579 11. Kyle Satterstrom F, Kosmicki JA, Wang J, Breen MS, De Rubeis S, An J-Y, *et al.* (2019,  
580 April 24): Large-scale exome sequencing study implicates both developmental and  
581 functional changes in the neurobiology of autism. *bioRxiv*. p 484113.
- 582 12. Ruzzo EK, Pérez-Cano L, Jung J-Y, Wang L-K, Kashef-Haghighi D, Hartl C, *et al.* (2019):  
583 Inherited and De Novo Genetic Risk for Autism Impacts Shared Networks. *Cell*, vol. 178. pp  
584 850–866.e26.
- 585 13. Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J, *et al.* (2012): De novo  
586 gene disruptions in children on the autistic spectrum. *Neuron* 74: 285–299.
- 587 14. Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, *et al.* (2019): Identification of  
588 common genetic risk variants for autism spectrum disorder. *Nat Genet* 51: 431–444.
- 589 15. Park J-H, Wacholder S, Gail MH, Peters U, Jacobs KB, Chanock SJ, Chatterjee N (2010):

- 590 Estimation of effect size distribution from genome-wide association studies and implications  
591 for future discoveries. *Nat Genet* 42: 570–575.
- 592 16. Gibson G (2012): Rare and common variants: twenty arguments. *Nat Rev Genet* 13: 135–  
593 145.
- 594 17. Robinson MR, Wray NR, Visscher PM (2014): Explaining additional genetic variation in  
595 complex traits. *Trends Genet* 30: 124–132.
- 596 18. SPARK Consortium (2018): SPARK: A US Cohort of 50,000 Families to Accelerate Autism  
597 Research. *Neuron* 97: 488–493.
- 598 19. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, *et al.* (2012):  
599 Systematic localization of common disease-associated variation in regulatory DNA. *Science*  
600 337: 1190–1195.
- 601 20. Tak YG, Farnham PJ (2015): Making sense of GWAS: using epigenomics and genome  
602 engineering to understand the functional relevance of SNPs in non-coding regions of the  
603 human genome. *Epigenetics Chromatin* 8: 57.
- 604 21. Schaid DJ, Chen W, Larson NB (2018): From genome-wide associations to candidate  
605 causal variants by statistical fine-mapping. *Nat Rev Genet* 19: 491–504.
- 606 22. Muerdter F, Boryn ŁM, Arnold CD (2015): STARR-seq—principles and applications.  
607 *Genomics* 106: 145–150.
- 608 23. Inoue F, Ahituv N (2015): Decoding enhancers using massively parallel reporter assays.  
609 *Genomics* 106: 159–164.
- 610 24. Davis JE, Insigne KD, Jones EM, Hastings QB, Kosuri S (2019, May 13): Multiplexed  
611 dissection of a model human transcription factor binding site architecture. *bioRxiv*. p  
612 625434.
- 613 25. Rockman MV, Kruglyak L (2006): Genetics of global gene expression. *Nat Rev Genet* 7:  
614 862–872.
- 615 26. Nica AC, Dermitzakis ET (2013): Expression quantitative trait loci: present and future. *Philos*

- 616 *Trans R Soc Lond B Biol Sci* 368: 20120362.
- 617 27. GTEx Consortium, Laboratory, Data Analysis & Coordinating Center (LDACC)—Analysis  
618 Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx  
619 (eGTEx) groups, NIH Common Fund, NIH/NCI, *et al.* (2017): Genetic effects on gene  
620 expression across human tissues. *Nature* 550: 204–213.
- 621 28. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragooczy T, Telling A, *et al.*  
622 (2009): Comprehensive mapping of long-range interactions reveals folding principles of the  
623 human genome. *Science* 326: 289–293.
- 624 29. Sanyal A, Lajoie BR, Jain G, Dekker J (2012): The long-range interaction landscape of gene  
625 promoters. *Nature* 489: 109–113.
- 626 30. Jin F, Li Y, Dixon JR, Selvaraj S, Ye Z, Lee AY, *et al.* (2013): A high-resolution map of the  
627 three-dimensional chromatin interactome in human cells. *Nature* 503: 290–294.
- 628 31. Nancy Y, Fauni H, Ma W, Won H (2019): Connecting gene regulatory relationships to  
629 neurobiological mechanisms of brain disorders. *bioRxiv*. Retrieved from  
630 <https://www.biorxiv.org/content/10.1101/681353v1.abstract>
- 631 32. Wang D, Liu S, Warrell J, Won H, Shi X, Navarro FCP, *et al.* (2018): Comprehensive  
632 functional genomic resource and integrative model for the human brain. *Science* 362.  
633 <https://doi.org/10.1126/science.aat8464>
- 634 33. Walker RL, Ramaswami G, Hartl C, Mancuso N, Gandal MJ, de la Torre-Ubieta L, *et al.*  
635 (2019): Genetic Control of Expression and Splicing in Developing Human Brain Informs  
636 Disease Mechanisms. *Cell* 179: 750–771.e22.
- 637 34. Loh P-R, Danecek P, Palamara PF, Fuchsberger C, A Reshef Y, K Finucane H, *et al.*  
638 (2016): Reference-based phasing using the Haplotype Reference Consortium panel. *Nat*  
639 *Genet* 48: 1443–1448.
- 640 35. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ (2015): Second-generation  
641 PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4: 7.

- 642 36. Cordell HJ, Barratt BJ, Clayton DG (2004): Case/pseudocontrol analysis in genetic  
643 association studies: A unified framework for detection of genotype and haplotype  
644 associations, gene-gene and gene-environment interactions, and parent-of-origin effects.  
645 *Genet Epidemiol* 26: 167–185.
- 646 37. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, *et al.* (2016): Next-generation  
647 genotype imputation service and methods. *Nat Genet* 48: 1284–1287.
- 648 38. Taliun D, Harris DN, Kessler MD, Carlson J, Szpiech ZA, Torres R, *et al.* (2019, March 6):  
649 Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *bioRxiv*. p  
650 563866.
- 651 39. Kowalski MH, Qian H, Hou Z, Rosen JD, Tapia AL, Shan Y, *et al.* (2019, July 2): Use of  
652 >100,000 NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium whole  
653 genome sequences improves imputation quality and detection of rare variant associations  
654 in admixed African and Hispanic/Latino populations. *bioRxiv*. p 683201.
- 655 40. Sariya S, Lee JH, Mayeux R, Vardarajan BN, Reyes-Dumeyer D, Manly JJ, *et al.* (2019):  
656 Rare Variants Imputation in Admixed Populations: Comparison Across Reference Panels  
657 and Bioinformatics Tools. *Front Genet* 10: 239.
- 658 41. Willer CJ, Li Y, Abecasis GR (2010): METAL: fast and efficient meta-analysis of  
659 genomewide association scans. *Bioinformatics* 26: 2190–2191.
- 660 42. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, *et al.* (2019):  
661 The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted  
662 arrays and summary statistics 2019. *Nucleic Acids Res* 47: D1005–D1012.
- 663 43. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group  
664 of the Psychiatric Genomics Consortium, *et al.* (2015): LD Score regression distinguishes  
665 confounding from polygenicity in genome-wide association studies. *Nat Genet* 47: 291–295.
- 666 44. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh P-R, *et al.* (2015):  
667 Partitioning heritability by functional annotation using genome-wide association summary



- 668 statistics. *Nat Genet* 47: 1228–1235.
- 669 45. Choi SW, O'Reilly PF (2019): PRSice-2: Polygenic Risk Score software for biobank-scale  
670 data. *Gigascience* 8. <https://doi.org/10.1093/gigascience/giz082>
- 671 46. Won H, de la Torre-Ubieta L, Stein JL, Parikshak NN, Huang J, Opland CK, *et al.* (2016):  
672 Chromosome conformation elucidates regulatory relationships in developing human brain.  
673 *Nature* 538: 523–527.
- 674 47. Pardiñas AF, Holmans P, Pocklington AJ, Escott-Price V, Ripke S, Carrera N, *et al.* (2018):  
675 Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions  
676 under strong background selection. *Nat Genet* 50: 381–389.
- 677 48. Law CW, Chen Y, Shi W, Smyth GK (2014): voom: Precision weights unlock linear model  
678 analysis tools for RNA-seq read counts. *Genome Biol* 15: R29.
- 679 49. Myint L, Avramopoulos DG, Goff LA, Hansen KD (2019): Linear models enable powerful  
680 differential activity analysis in massively parallel reporter assays. *BMC Genomics* 20: 209.
- 681 50. Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium  
682 (2017): Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder  
683 highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. *Mol*  
684 *Autism* 8: 21.
- 685 51. Smith DJ, Escott-Price V, Davies G, Bailey MES, Colodro-Conde L, Ward J, *et al.* (2016):  
686 Genome-wide analysis of over 106 000 individuals identifies 9 neuroticism-associated loci.  
687 *Mol Psychiatry* 21: 749–757.
- 688 52. Luciano M, Hagenaars SP, Davies G, Hill WD, Clarke T-K, Shiralil M, *et al.* (2018):  
689 Association analysis in over 329,000 individuals identifies 116 independent variants  
690 influencing neuroticism. *Nat Genet* 50: 6–11.
- 691 53. Nagel M, Watanabe K, Stringer S, Posthuma D, van der Sluis S (2018): Item-level analyses  
692 reveal genetic heterogeneity in neuroticism. *Nat Commun* 9: 905.
- 693 54. Nagel M, Jansen PR, Stringer S, Watanabe K, de Leeuw CA, Bryois J, *et al.* (2018): Meta-

- 694 analysis of genome-wide association studies for neuroticism in 449,484 individuals  
695 identifies novel genetic loci and pathways. *Nat Genet* 50: 920–927.
- 696 55. Kichaev G, Bhatia G, Loh P-R, Gazal S, Burch K, Freund MK, *et al.* (2019): Leveraging  
697 Polygenic Functional Enrichment to Improve GWAS Power. *Am J Hum Genet* 104: 65–75.
- 698 56. Lee JJ, Wedow R, Okbay A, Kong E, Maghazian O, Zacher M, *et al.* (2018): Gene discovery  
699 and polygenic prediction from a genome-wide association study of educational attainment  
700 in 1.1 million individuals. *Nat Genet* 50: 1112–1121.
- 701 57. Adams HHH, Hibar DP, Chouraki V, Stein JL, Nyquist PA, Rentería ME, *et al.* (2016): Novel  
702 genetic loci underlying human intracranial volume identified through genome-wide  
703 association. *Nat Neurosci* 19: 1569–1582.
- 704 58. Hansen SN, Overgaard M, Andersen PK, Pamer ET (2017): Estimating a population  
705 cumulative incidence under calendar time trends. *BMC Med Res Methodol* 17: 7.
- 706 59. Peterson RE, Kuchenbaecker K, Walters RK, Chen C-Y, Popejoy AB, Periyasamy S, *et al.*  
707 (2019): Genome-wide Association Studies in Ancestrally Diverse Populations:  
708 Opportunities, Methods, Pitfalls, and Recommendations. *Cell* 179: 589–603.
- 709 60. Duncan L, Shen H, Gelaye B, Meijsen J, Ressler K, Feldman M, *et al.* (2019): Analysis of  
710 polygenic risk score usage and performance in diverse human populations. *Nat Commun*  
711 10: 3328.
- 712 61. Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, *et al.* (2007): Strong  
713 association of de novo copy number mutations with autism. *Science* 316: 445–449.
- 714 62. Ronemus M, lossifov I, Levy D, Wigler M (2014): The role of de novo mutations in the  
715 genetics of autism spectrum disorders. *Nat Rev Genet* 15: 133–141.
- 716 63. Loomes R, Hull L, Mandy WPL (2017): What Is the Male-to-Female Ratio in Autism  
717 Spectrum Disorder? A Systematic Review and Meta-Analysis. *J Am Acad Child Adolesc*  
718 *Psychiatry* 56: 466–474.
- 719 64. Levy D, Ronemus M, Yamrom B, Lee Y-H, Leotta A, Kendall J, *et al.* (2011): Rare de novo

- 720 and transmitted copy-number variation in autistic spectrum disorders. *Neuron* 70: 886–897.
- 721 65. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, *et al.* (2012):  
722 De novo mutations revealed by whole-exome sequencing are strongly associated with  
723 autism. *Nature* 485: 237–241.
- 724 66. Dong S, Walker MF, Carriero NJ, DiCola M, Willsey AJ, Ye AY, *et al.* (2014): De Novo  
725 Insertions and Deletions of Predominantly Paternal Origin Are Associated with Autism  
726 Spectrum Disorder. *Cell Rep* 9: 16–23.
- 727 67. Zhao X, Leotta A, Kustanovich V, Lajonchere C, Geschwind DH, Law K, *et al.* (2007): A  
728 unified genetic theory for sporadic and inherited autism. *Proc Natl Acad Sci U S A* 104:  
729 12831–12836.
- 730 68. Hazlett HC, Gu H, Munsell BC, Kim SH, Styner M, Wolff JJ, *et al.* (2017): Early brain  
731 development in infants at high risk for autism spectrum disorder. *Nature* 542: 348–351.
- 732 69. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, *et*  
733 *al.* (2015): Integrative analysis of 111 reference human epigenomes. *Nature* 518: 317–330.
- 734 70. de la Torre-Ubieta L, Stein JL, Won H, Opland CK, Liang D, Lu D, Geschwind DH (2018):  
735 The Dynamic Landscape of Open Chromatin during Human Cortical Neurogenesis. *Cell*  
736 172: 289–304.e18.
- 737 71. Zhou J, Park CY, Theesfeld CL, Wong AK, Yuan Y, Scheckel C, *et al.* (2019): Whole-  
738 genome deep-learning analysis identifies contribution of noncoding mutations to autism  
739 risk. *Nat Genet* 51: 973–980.
- 740 72. Parikhshak NN, Swarup V, Belgard TG, Irimia M, Ramaswami G, Gandal MJ, *et al.* (2016):  
741 Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in  
742 autism. *Nature* 540: 423–427.
- 743 73. Reimand J, Kull M, Peterson H, Hansen J, Vilo J (2007): g:Profiler—a web-based toolset for  
744 functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res* 35:  
745 W193–W200.

- 746 74. Kang HJ, Kawasaki YI, Cheng F, Zhu Y, Xu X, Li M, *et al.* (2011): Spatio-temporal  
747 transcriptome of the human brain. *Nature* 478: 483–489.
- 748 75. Simonoff E, Pickles A, Charman T, Chandler S, Loucas T, Baird G (2008): Psychiatric  
749 disorders in children with autism spectrum disorders: prevalence, comorbidity, and  
750 associated factors in a population-derived sample. *J Am Acad Child Adolesc Psychiatry* 47:  
751 921–929.
- 752 76. van Steensel FJA, Bögels SM, Perrin S (2011): Anxiety disorders in children and  
753 adolescents with autistic spectrum disorders: a meta-analysis. *Clin Child Fam Psychol Rev*  
754 14: 302–317.
- 755 77. Antshel KM, Zhang-James Y, Wagner KE, Ledesma A, Faraone SV (2016): An update on  
756 the comorbidity of ADHD and ASD: a focus on clinical management. *Expert Rev Neurother*  
757 16: 279–293.
- 758 78. Romero M, Aguilar JM, Del-Rey-Mejías Á, Mayoral F, Rapado M, Peciña M, *et al.* (2016):  
759 Psychiatric comorbidities in autism spectrum disorder: A comparative study between DSM-  
760 IV-TR and DSM-5 diagnosis. *Int J Clin Health Psychol* 16: 266–275.
- 761 79. Cross-Disorder Group of the Psychiatric Genomics Consortium (2019): Genomic  
762 Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders.  
763 *Cell* 179: 1469–1482.e11.
- 764 80. Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E, *et al.* (2019): Discovery  
765 of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat*  
766 *Genet* 51: 63–75.
- 767 81. Savage JE, Jansen PR, Stringer S, Watanabe K, Bryois J, de Leeuw CA, *et al.* (2018):  
768 Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and  
769 functional links to intelligence. *Nat Genet* 50: 912–919.
- 770 82. Stahl EA, Breen G, Forstner AJ, McQuillin A, Ripke S, Trubetsky V, *et al.* (2019): Genome-  
771 wide association study identifies 30 loci associated with bipolar disorder. *Nat Genet* 51:

- 772 793–803.
- 773 83. Howard DM, Adams MJ, Clarke T-K, Hafferty JD, Gibson J, Shirali M, *et al.* (2019):  
774 Genome-wide meta-analysis of depression identifies 102 independent variants and  
775 highlights the importance of the prefrontal brain regions. *Nat Neurosci* 22: 343–352.
- 776 84. Pasman JA, Verweij KJH, Gerring Z, Stringer S, Sanchez-Roige S, Treur JL, *et al.* (2018):  
777 GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits,  
778 and a causal influence of schizophrenia. *Nat Neurosci* 21: 1161–1170.
- 779 85. Liu M, Jiang Y, Wedow R, Li Y, Brazel DM, Chen F, *et al.* (2019): Association studies of up  
780 to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol  
781 use. *Nat Genet* 51: 237–244.
- 782 86. Jansen IE, Savage JE, Watanabe K, Bryois J, Williams DM, Steinberg S, *et al.* (2019):  
783 Genome-wide meta-analysis identifies new loci and functional pathways influencing  
784 Alzheimer’s disease risk. *Nat Genet* 51: 404–413.
- 785 87. Nalls MA, Blauwendraat C, Vallerga CL, Heilbron K, Bandres-Ciga S, Chang D, *et al.* (2018,  
786 August 9): Expanding Parkinson’s disease genetics: novel risk loci, genomic context,  
787 causal insights and heritable risk [no. biorxiv;388165v3]. *Genetics*. bioRxiv, p 223.
- 788 88. Brainstorm Consortium, Anttila V, Bulik-Sullivan B, Finucane HK, Walters RK, Bras J, *et al.*  
789 (2018): Analysis of shared heritability in common disorders of the brain. *Science* 360.  
790 <https://doi.org/10.1126/science.aap8757>
- 791 89. Coetzee SG, Coetzee GA, Hazelett DJ (2015): motifbreakR: an R/Bioconductor package for  
792 predicting variant effects at transcription factor binding sites. *Bioinformatics* 31: 3847–3849.
- 793 90. Krumm N, O’Roak BJ, Karakoc E, Mohajeri K, Nelson B, Vives L, *et al.* (2013):  
794 Transmission disequilibrium of small CNVs in simplex autism. *Am J Hum Genet* 93: 595–  
795 606.
- 796 91. Jacquemont S, Coe BP, Hersch M, Duyzend MH, Krumm N, Bergmann S, *et al.* (2014): A  
797 Higher Mutational Burden in Females Supports a “Female Protective Model” in

- 798 Neurodevelopmental Disorders. *Am J Hum Genet* 94: 415–425.
- 799 92. Parikshak NN, Luo R, Zhang A, Won H, Lowe JK, Chandran V, *et al.* (2013): Integrative  
800 functional genomic analyses implicate specific molecular pathways and circuits in autism.  
801 *Cell* 155: 1008–1021.
- 802 93. Sato S-I, Inoue H, Kogure T, Tagaya M, Tani K (2010): Golgi-localized KIAA0725p regulates  
803 membrane trafficking from the Golgi apparatus to the plasma membrane in mammalian  
804 cells. *FEBS Letters*, vol. 584. pp 4389–4395.
- 805 94. Nagase T, Ishikawa K-I, Suyama M, Kikuno R, Hirose M, Miyajima N, *et al.* (1998):  
806 Prediction of the coding sequences of unidentified human genes. XII. The complete  
807 sequences of 100 new cDNA clones from brain which code for large proteins in vitro. *DNA*  
808 *Res* 5: 355–364.
- 809 95. Nakajima K-I, Sonoda H, Mizoguchi T, Aoki J, Arai H, Nagahama M, *et al.* (2002): A novel  
810 phospholipase A1 with sequence homology to a mammalian Sec23p-interacting protein,  
811 p125. *J Biol Chem* 277: 11329–11335.
- 812 96. Schuurs-Hoeijmakers JHM, Geraghty MT, Kamsteeg E-J, Ben-Salem S, de Bot ST, Nijhof  
813 B, *et al.* (2012): Mutations in DDHD2, encoding an intracellular phospholipase A(1), cause  
814 a recessive form of complex hereditary spastic paraplegia. *Am J Hum Genet* 91: 1073–  
815 1081.
- 816 97. Gonzalez M, Nampoothiri S, Kornblum C, Oteyza AC, Walter J, Konidari I, *et al.* (2013):  
817 Mutations in phospholipase DDHD2 cause autosomal recessive hereditary spastic  
818 paraplegia (SPG54). *Eur J Hum Genet* 21: 1214–1218.
- 819 98. Novarino G, Fenstermaker AG, Zaki MS, Hofree M, Silhavy JL, Heiberg AD, *et al.* (2014):  
820 Exome sequencing links corticospinal motor neuron disease to common neurodegenerative  
821 disorders. *Science* 343: 506–511.
- 822 99. Inloes JM, Hsu K-L, Dix MM, Viader A, Masuda K, Takei T, *et al.* (2014): The hereditary  
823 spastic paraplegia-related enzyme DDHD2 is a principal brain triglyceride lipase. *Proc Natl*

824 *Acad Sci U S A* 111: 14924–14929.

825 100. Sokhadze EM, Tasman A, Sokhadze GE, El-Baz AS, Casanova MF (2016): Behavioral,

826 Cognitive, and Motor Preparation Deficits in a Visual Cued Spatial Attention Task in Autism

827 Spectrum Disorder. *Appl Psychophysiol Biofeedback* 41: 81–92.

828 101. Peyrot WJ, Boomsma DI, Penninx BWJH, Wray NR (2016): Disease and Polygenic

829 Architecture: Avoid Trio Design and Appropriately Account for Unscreened Control Subjects

830 for Common Disease. *Am J Hum Genet* 98: 382–391.

831 102. Bodea CA, Neale BM, Ripke S, International IBD Genetics Consortium, Daly MJ, Devlin B,

832 Roeder K (2016): A Method to Exploit the Structure of Genetic Ancestry Space to Enhance

833 Case-Control Studies. *Am J Hum Genet* 98: 857–868.

834 103. Sirugo G, Williams SM, Tishkoff SA (2019): The Missing Diversity in Human Genetic

835 Studies. *Cell* 177: 1080.

836

## 837 Legends

### 838 **Figure 1. Genome-wide association of ASD in the SPARK dataset**

839 (A) GWAS result from SPARK full dataset ( $N_{\text{case+pseudocontrol}} = 12,444$ ). (B) Genetic correlations across ASD

840 GWAS. From left to right, iPSYCH versus PGC (50), SPARK EUR versus iPSYCH, SPARK EUR versus

841 PGC and SPARK EUR versus iPSYCH-PGC study (14). (C) GWAS results from the meta-analysis (SPARK

842 European population and iPSYCH-PGC,  $N_{\text{max\_case+control}} = 55,420$ ). For Manhattan plots (A, C), the x-axes

843 indicate the chromosomal position and y-axes indicate the significance of associations. The blue and red

844 lines denote the significance threshold at suggestive ( $P < 1 \times 10^{-6}$ ) and significant ( $P < 5 \times 10^{-8}$ ) levels. SNPs

845 at with  $P < 1 \times 10^{-6}$  are shown as a filled circle.  $R_s$  number indicates index SNPs from independent loci (1

846 MB apart from each other) at  $P < 1 \times 10^{-8}$ . Index SNPs at  $P < 5 \times 10^{-8}$  are shown as diamonds.

### 847 **Figure 2. Comparison of polygenic risk scores between subgroups.**

848 (A) Common variant risk burden is higher in cases compared to pseudocontrols. (B) Comparison of PRS

849 across family types (from left to right, families with multiple affected children with affected parent(s), multiple

850 affected children with unaffected parents, one affected child with affected parent(s), and one affected child

851 with unaffected parents) shows no evidence for a higher common variant burden in multiplex families. (C)

852 Comparison of PRS between male and female probands shows no evidence of enrichment of common  
853 variants impacting risk for ASD in females. **(D)** There is no evidence for a difference in the transmission of  
854 common variant risk burden from mother versus father.

### 855 **Figure 3. Partitioned heritability enrichment of tissues implicates cortical development in ASD risk**

856 **(A)** Heritability enrichment in active enhancer and promoter regions present in different tissues shows the  
857 critical role of developing the brain in ASD etiology. **(B)** Heritability enrichment in differential chromatin  
858 accessibility from the developing fetal cortical wall. (upper) The x-axis represents tissue types and the y-  
859 axis indicates heritability enrichment. The error bar shows a 95% confidence interval. (lower) The x-axis  
860 represents tissue types and the y-axis indicates statistical significance as  $-\log_{10}(\text{FDR})$ . BRN: Brain, ADRL:  
861 Adrenal, BLD: Blood, BRST: Breast, CRVX: Cervix, ESDR: ESC\_derived, GI: GI\_duodenum, GI\_colon,  
862 GI\_rectum, GI\_stomach, GI\_intestine, GI\_colon, GI\_rectum, GI\_duodenum and GI\_esophagus, HRT:  
863 Heart, KID: Kidney, LIV: Liver, LNG: Lung, MUS: Muscle, OVRY: Ovary, PANC: Pancreas, PLCNT:  
864 Placenta, SPLN: Spleen, STRM: Stromal connective, THYM: Thymus, VAS: Vascular, CP: Peaks more  
865 accessible in cortical plate, GZ: Peaks more accessible in germinal zone. \* FDR < 0.05, \*\* FDR < 0.01.

### 866 **Figure 4. H-MAGMA identified 263 protein-coding genes linked to ASD**

867 **(A)** Schematic diagram of H-MAGMA. SNP based association P-values were aggregated to gene-based P-  
868 values using positional information as well as chromatin interaction in the fetal brain. **(B)** Gene-based  
869 association result from H-MAGMA. The x-axis indicates the start position of genes (hg19). **(C)** Overlap of  
870 ASD risk genes harboring common and rare variants (11). **(D)** Overlapped genes with differential  
871 expression from post-mortem brains in individuals with ASD patients and neurotypical controls (72). **(E)**  
872 Gene ontologies enriched for ASD linked genes (top 20). **(F)** Developmental expression pattern of ASD risk  
873 genes (74).

### 874 **Figure 5. Genetic correlation of ASD against twelve brain and behavioral phenotypes**

875 The x-axis represents an estimate of the genetic correlation ( $r_g$ ). Error bars represent the 95% confidence  
876 interval. P-values at FDR < 0.05 are shown in bold. MDD: Major depressive disorder, ADHD: Attention-  
877 Deficit/Hyperactivity Disorder.

878

### 879 **Figure 6. Identification of putative causal variant and gene impacting risk for ASD**

880 **(A)** Annotated locus plot near rs60527016 ASD risk index variant, from top panel to bottom, ASD  
881 associations within SPARK full dataset (n = 6,222 case-pseudocontrol pairs), eQTL for DDHD2 in fetal  
882 brains (n = 235) and adult brain (n = 1,387), MPRA expression (n=6), ATAC-seq averaged depth in neuron  
883 (n = 61) and progenitor (n = 73). Differential open chromatin accessibility peaks from ATAC-seq, and gene  
884 model (NCBI Refseq). LD was calculated to rs7001340 within SPARK parents of cases, fetal brain donors,  
885 or 1KG EUR and colored accordingly. **(B)** The barcoded expression level of mRNA based from each allele  
886 at rs7001340 from the MPRA experiment. **(C)** The expression level of DDHD2 by rs7001340 genotypes in



887 the fetal brain. **(D)** The expression level of DDHD2 by rs7001340 genotypes in adult brain. Individuals with  
888 allele dosage (0-0.1 as C/C, 0.9-1.1 as C/T, 1.9-2.0 as T/T) are shown. For **(B)** to **(D)**, ASD risk allele for  
889 rs7001340 is T and protective allele is C.