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

Nana Matoba¹,², Dan Liang¹,², Huaigu Sun¹,², Nil Aygün¹,², Jessica C. McAfee¹,², Jessica E. Davis³,⁸, Laura M. Raffle², Huijun Qian⁹, Joseph Piven¹⁰, Yun Li¹,¹¹,¹², Srijan Kosuri³,⁸, Hyejung Won¹,²,*, Jason L. Stein¹,²,*

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

* These authors equally supervised.

Corresponding authors: Hyejung Won or Jason L. Stein
Hyejung Won:
Address: 7202B Mary Ellen Jones Building, 116 Manning Drive, Chapel Hill, NC 27599
Phone: +1 (919) 966-4069
Email: hyejung_won@med.unc.edu;
Jason L. Stein:
Address: 7202A Mary Ellen Jones Building, 116 Manning Drive, Chapel Hill, NC 27599
Phone: +1 (919) 843-5541
Email: jason_stein@med.unc.edu

Short title: Common genetic risk variants for autism in SPARK
Keywords: Autism spectrum disorder, GWAS, common variant, MPRA, DDHD2
Abstract

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

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

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

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

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

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

Once we identify GWS loci, the critical next step is to understand their biological impact. This is especially challenging because most GWAS identified loci for neurodevelopmental disorders and other traits are located in poorly annotated non-coding regions with presumed gene regulatory function (19). In addition, most loci are comprised of multiple single nucleotide polymorphisms.
(SNPs) that are often inherited together, which makes it difficult to identify the true causal variant(s) and their regulatory effects (20,21). To overcome these problems, various experimental validation tools have been developed (22–24). One of these tools, a massively parallel reporter assay (MPRA), simultaneously evaluates allelic effects on enhancer activity for many variants. In this assay, exogenous DNA constructs, harboring risk and protective alleles at an associated variant, drive the expression of a barcoded transcript. Differences in barcode counts between the risk and protective alleles indicate the regulatory function of that variant (23,24). This assay thus demonstrates the regulatory potential of individual SNPs and provides evidence of the causal variants within an associated locus.

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

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

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

SPARK dataset

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

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

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

Genome-wide association analysis and Meta-analysis with iPSYCH-PGC study

We tested association within the SPARK all case-pseudocontrol pairs (full dataset; Supplementary Table S1) using PLINK2 generalized linear model (--glm) for SNPs with MAF ≥ 0.01 and imputation quality score from minimac4 (R2) > 0.5 (Supplementary Figure S4). In this model, we did not include any covariates since cases and pseudocontrols are matched on environmental variables and genetic ancestry. We performed secondary GWAS analyses by subsetting to only specific ancestry groups. We called ancestry using multidimensional scaling (MDS) analysis with 988 HapMap3 individuals and one random case from each trio (Supplementary Figure S3, Supplementary Table S2). Ancestry of individuals from SPARK was called as European, African or East Asian ancestries if they were within 5 standard deviations of
defined HapMap3 population (CEU/TSI; YRI/LWK; or CHB/CHD/JPT, respectively) centroids in MDS dimensions 1 and 2. Population-specific GWASs were carried out using the same association model as described above for the SPARK all ancestries dataset. Meta-analyses with iPSYCH-PGC study (14) were performed by METAL (release 2018-08-28) (41). Additional information for iPSYCH-PGC summary statistics is provided in Supplementary Methods.

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

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

Estimating polygenic risk score
Polygenic risk scores (PRSs) were calculated based on the iPSYCH-PGC study (14) using PRSice-2 (45) (https://www.prsice.info/). Details on generation of PRS, sex-stratified and family-type PRS, and parental origin PRS analyses are provided in Supplementary Methods.
SNP to Ensembl gene annotation was carried out by Hi-C coupled MAGMA (H-MAGMA) (https://github.com/thewonlab/H-MAGMA) by leveraging chromatin-interaction generated from fetal brain Hi-C (46) as previously described (31). Details on H-MAGMA and functional analyses of H-MAGMA genes are provided in Supplementary Methods.

Construction of a Massively Parallel Reporter Assay (MPRA) Library

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

Functional annotation of rs7001340 locus with multi-omic datasets

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

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

We obtained genotype and clinical diagnosis of ASD via self- or parent-report from 27,290 individuals who participated in the SPARK project (18). The majority of data comprised families, including those where both parents and multiple children were genotyped (quads; N=3,192 families), where both parents and one child were genotyped (trios; N=2,486 families), or where one parent and one child were genotyped (duos; N=2,448 families) (Supplementary Figure S1). Only 68 individuals were ascertained without family members (singletons). After genotyping quality control (Supplementary Methods), 375,918 variants from 26,883 individuals were retained. Because the SPARK dataset did not genotype unrelated controls, we created pseudocontrols from the alleles not transmitted from parents to probands (36). Case-pseudocontrol design requires genotyping of both parents, so singletons and duos were excluded from the analysis. Due to the diverse ancestry in the cohort (Supplementary Figure S2, Supplementary Table S1), genotypes of all individuals including pseudocontrols were imputed to a diverse reference panel (TOPMed Freeze 5b reference panel consisting of 125,568 haplotypes). After imputation quality control (Methods; Supplementary Figure S3), 90,051,896 autosomal SNPs were tested for association in 6,222 case-pseudocontrol pairs (SPARK full dataset) consisting of 4,956 males and 1,266 females from multiple ancestries including European (N = 4,535), African (N = 37), East Asian (N = 83) and other ancestries/admixed individuals (N = 1,567) (Supplementary Figure S1, Supplementary Table S2). We observed no inflation of test statistics (\(\lambda_{GC} = 1.00\)) (Supplementary Figure S3), indicating population stratification was well-controlled when using this case-pseudocontrol design. We identified two SNPs at one locus (index SNP: rs60527016-C; OR = 0.84, \(P = 4.70\times10^{-8}\)) at genome-wide significance (\(P < 5.0 \times 10^{-8}\)) (Figure 1A, Table 1), which were supported by the previous largest ASD GWAS (14) (OR = 0.95, \(P = 0.0047\)) derived from the PGC and iPSYCH cohorts (Supplementary Figure S4).
Figure 1. Genome-wide association of ASD in the SPARK dataset

(A) GWAS result from SPARK full dataset (N\text{case+pseudocontrol} = 12,444). (B) Genetic correlations across ASD GWAS. From left to right, iPSYCH versus PGC (50), SPARK EUR versus iPSYCH, SPARK EUR versus PGC and SPARK EUR versus iPSYCH-PGC study (14). (C) GWAS results from the meta-analysis (SPARK European population and iPSYCH-PGC, N\text{max_case+control} = 55,420). For Manhattan plots (A, C), the x-axes indicate the chromosomal position and y-axes indicate the significance of associations. The blue and red lines denote the significance threshold at suggestive (P < 1\times10^{-6}) and significant (P<5\times10^{-8}) levels. SNPs at with P < 1\times10^{-6} are shown as a filled circle. Rs number indicates index SNPs from independent loci (1 MB apart from each other) at P < 1\times10^{-8}). Index SNPs at P < 5\times10^{-8} are shown as diamonds.
Table 1. Genome-wide significant loci associated with ASD risk

| SNP       | Position (hg38) | EA | OA | EAF | OR(95%CI) | P      | SPARK | iPSYCH+PGC | Meta(EUR)2 |
|-----------|-----------------|----|----|----|-----------|--------|-------|-----------|------------|
| rs16219   | 1:96104001      | T  | C  | 0.34 | 1.08 (1.03 - 1.14) | 0.003  | 1.08 (1.05 - 1.11) | 3.99×10⁻⁷ | 1.08 (1.05 - 1.11) | 6.42×10⁻⁹ |
| rs10099100| 8:10719265      | C  | G  | 0.31 | 1.08 (1.02 - 1.14) | 0.008  | 1.09 (1.06 - 1.12) | 1.07×10⁻⁸ | 1.08 (1.05 - 1.11) | 7.65×10⁻⁹ |
| rs60527016| 8:38442106      | C  | T  | 0.21 | 0.84 (0.79 - 0.90) | 0.003  | 0.95 (0.92 - 0.99) | 0.00466 | 0.93 (0.91 - 0.96) | 3.05×10⁻⁶ |
| rs112436750| 17:45887763     | A  | AT | 0.21 | 1.07 (1.01 - 1.14) | 0.027  | 1.09 (1.05 - 1.12) | 1.23×10⁻⁶ | 1.09 (1.06 - 1.12) | 2.62×10⁻⁸ |
| rs1000177 | 20:21252560     | T  | C  | 0.24 | 1.08 (1.02 - 1.15) | 0.014  | 1.10 (1.07 - 1.14) | 3.32×10⁻³ | 1.09 (1.06 - 1.13) | 1.34×10⁻⁹ |

Suggestive loci (5×10⁻⁸ ≤ P < 1×10⁻⁶)

| SNP       | Position (hg38) | EA | OA | EAF | OR(95%CI) | P      | SPARK | iPSYCH+PGC | Meta(EUR)2 |
|-----------|-----------------|----|----|----|-----------|--------|-------|-----------|------------|
| rs6701243 | 1:98627228      | A  | C  | 0.38 | 0.99 (0.94 - 1.00) | 0.610  | 0.93 (0.90 - 0.96) | 3.07×10⁻⁷ | 0.94 (0.91 - 0.96) | 5.90×10⁻⁷ |
| rs6743102 | 2:158521946     | G  | A  | 0.34 | 0.94 (0.89 - 0.99) | 0.021  | 0.94 (0.91 - 0.97) | 8.99×10⁻⁶ | 0.94 (0.91 - 0.96) | 4.07×10⁻⁷ |
| rs33966416| 4:170285452     | C  | A  | 0.50 | 0.95 (0.90 - 1.00) | 0.038  | 0.94 (0.91 - 0.96) | 2.73×10⁻⁶ | 0.94 (0.92 - 0.96) | 6.99×10⁻⁷ |
| rs4916723 | 5:88558577      | A  | C  | 0.40 | 1.10 (1.00 - 1.11) | 0.062  | 1.07 (1.04 - 1.10) | 1.92×10⁻⁶ | 1.07 (1.04 - 1.09) | 6.90×10⁻⁷ |
| rs416223  | 5:104655775     | C  | A  | 0.40 | 1.00 (0.96 - 1.10) | 0.730  | 1.07 (1.04 - 1.10) | 3.84×10⁻⁷ | 1.07 (1.04 - 1.09) | 3.56×10⁻⁷ |
| rs67248478| 6:134711094     | C  | T  | 0.34 | 0.94 (0.90 - 0.99) | 0.032  | 0.94 (0.91 - 0.96) | 3.22×10⁻⁶ | 0.94 (0.91 - 0.96) | 3.22×10⁻⁷ |
| rs76569799| 9:73565191      | C  | T  | 0.15 | 1.10 (0.99 - 1.11) | 0.076  | 1.09 (1.05 - 1.13) | 3.90×10⁻⁶ | 1.08 (1.05 - 1.12) | 9.99×10⁻⁷ |
| rs4750990 | 10:128689762     | T  | C  | 0.36 | 1.00 (0.98 - 1.10) | 0.250  | 1.07 (1.04 - 1.10) | 1.37×10⁻⁶ | 1.07 (1.04 - 1.09) | 4.89×10⁻⁷ |
| rs2224274 | 20:14780101      | C  | T  | 0.43 | 1.00 (0.97 - 1.10) | 0.310  | 1.07 (1.04 - 1.11) | 2.86×10⁻⁷ | 1.07 (1.05 - 1.10) | 5.56×10⁻⁸ |

1Index SNPs from loci that survived genome-wide significance in any of the GWASs including meta-analysis.
2Meta-analysis of SPARK European ancestries and iPSYCH+PGC

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

Replication of genetic risk factors for ASD

Given the phenotypic heterogeneity of ASD and potential technical differences such as genotyping platforms or data processing, we assessed the replication of genetic risk factors across cohorts by comparing previous major ASD studies including PGC and iPSYCH cohort.
(14,50) with the SPARK dataset subset to individuals of European descent (EUR) (Figure 1B).

Although each study included multiple ASD subtypes including ASD from DSM5, Asperger's, autism/autistic disorder, and Pervasive Developmental Disorder - Not otherwise specified (PDD-NOS) from DSM IV, and approaches differed across these samples from requiring community diagnosis to best-estimate diagnosis based on standardized assessment, we obtained high genetic correlations between the SPARK EUR dataset and the largest iPSYCH-PGC GWAS ($r_g = 0.82; P = 5.27 \times 10^{-14}$), suggesting the genetic risk factors for autism are largely shared among different ASD GWAS and are generalizable despite differences in diagnostic criteria and batch effects.

We next performed meta-analysis with SPARK EUR samples and iPSYCH-PGC samples ($N_{case} = 18,381$ and $N_{control} = 27,969$) to maximize power. The meta-analysis identified four additional loci associated with risk for ASD (Supplementary Figure S5-8). These included three previously reported loci (14) and one novel locus on chromosome 17, where a gene-based test from the iPSYCH-PGC study has previously shown association with risk for ASD (14) (Figure 1C, Table 1, Supplementary Figure S7). This novel locus was also reported to be associated with more than 60 phenotypes including neuroticism (51–55), educational attainment (56) and intracranial volume (57) (index SNPs $r^2 > 0.8$ in 1KG EUR) (Supplementary Table S3) indicating highly pleiotropic effects at this locus. The SNP based heritability in SPARK EUR samples was estimated ($h^2_G$) to be 0.117 (s.e. = 0.0082) for population prevalence of 0.012 (14,58) which was comparable with the previous report ($h^2_G = 0.118$; s.e. = 0.010) (14).

The generalization of effects across ancestries for the five index SNPs identified (Table 1) was examined (Supplementary Figure S2, S3, and Supplementary Table S4). The association results from the cross-ancestry dataset were mainly driven by the European population, as expected given the larger sampling from this population. We found that some regions showed population-differentiated allele frequencies. For example, rs10099100 was more common in European and African populations (MAF = 0.33, 0.39 from tested samples, respectively) than in East Asians.
(MAF = 0.02 from tested samples), necessitating a further investigation of genetic risk factors for ASD in populations of diverse ancestry (59,60).

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

![Figure 2. Comparison of polygenic risk scores between subgroups.](image)
Common variant risk burden is higher in cases compared to pseudocontrols. **(B)** Comparison of PRS across family types (from left to right, families with multiple affected children with affected parent(s), multiple affected children with unaffected parents, one affected child with affected parent(s), and one affected child with unaffected parents) shows no evidence for a higher common variant burden in multiplex families. **(C)** Comparison of PRS between male and female probands shows no evidence of enrichment of common variants impacting risk for ASD in females. **(D)** There is no evidence for a difference in the transmission of common variant risk burden from mother versus father.

Investigation of common variant burden impacting risk for ASD

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

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

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

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

Figure 3. Partitioned heritability enrichment of tissues implicates cortical development in ASD risk
(A) Heritability enrichment in active enhancer and promoter regions present in different tissues shows the critical role of developing the brain in ASD etiology. (B) Heritability enrichment in differential chromatin accessibility from the developing fetal cortical wall. (upper) The x-axis represents tissue types and the y-axis indicates heritability enrichment. The error bar shows a 95% confidence interval. (lower) The x-axis represents tissue types and the y-axis indicates statistical significance as -log10(FDR). BRN: Brain, ADRL: Adrenal, BLD: Blood, BRST: Breast, CRVX: Cervix, ESDR: ESC-derived, GI: GI_duodenum, GI_colon, GI_rectum, GI_stomach, GI_intestine, GI_colon, GI_rectum, GI_duodenum and GI_esophagus, HRT: Heart, KID: Kidney, LIV: Liver, LNG: Lung, MUS: Muscle, OVRY: Ovary, PANC: Pancreas, PLCNT: Placenta, SPLN: Spleen, STRM: Stromal connective, THYM: Thymus, VAS: Vascular, CP: Peaks more accessible in cortical plate, GZ: Peaks more accessible in germinal zone. * FDR < 0.05, ** FDR < 0.01.

H-MAGMA identified genes and pathways impacting risk for ASD

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

Rank-based gene ontology enrichment analysis (73) suggested that ASD risk genes were enriched in 188 terms including telencephalon development and regulation of synapse organization (Figure 4E, Supplementary Table S7).
Since heritability enrichment analyses suggested genetically mediated impacts on cortical development contribute to ASD risk (Figure 3), we explored whether the expression level of ASD risk genes from H-MAGMA is different between prenatal and postnatal cortex. As shown previously (14,31), we found ASD risk genes exhibited higher expression in the prenatal cortex as compared to the postnatal cortex ($P = 2.77 \times 10^{-62}$) (Figure 4F) (74). In particular, the expression level of ASD risk genes was highest between 20 and 30 post-conception weeks (Supplementary Figure S9). Taken together, our results demonstrate common risk variants for ASD play an important role in the developing cortex.
Figure 4. H-MAGMA identified 263 protein-coding genes linked to ASD

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

Genetic correlation between ASD and 12 brain and behavioral phenotypes

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

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

Functional validation to fine-map causal variants and prioritize genes

Interestingly, the novel locus identified by the SPARK full dataset (rs60527016 at chr8:38.19M - chr8:38.45M, Figure 1A, 6A) was also identified as a pleiotropic locus in a recent cross-disorder meta-analysis on eight psychiatric disorders (79). This locus was not only associated with ASD but also with schizophrenia, bipolar disorder and obsessive-compulsive disorder (OCD), suggesting that understanding the regulatory mechanism at this locus may reveal the basis for
pleiotropic effects across psychiatric disorders. The index SNP (rs60527016) was located within a 300 kb LD block ($r^2 > 0.5$ in SPARK full dataset) that contains seven genes (Figure 6A). To prioritize causal variants within this locus, we performed a massively parallel reporter assay (MPRA) (23,24) on 98 credible SNPs in this region in HEK 293 cells. MPRA measures barcoded transcriptional activity driven by each allele in a high-throughput fashion (Supplementary Figure S13). Surprisingly, SNP rs7001340 exhibited the strongest allelic difference in barcoded expression ($P = 1.51 \times 10^{-24}$) even though it is 37 kb away from the GWAS index SNP ($r^2 = 0.85$ with the index SNP in SPARK full dataset) (Figure 6A, B, Supplementary Table S9), demonstrating the regulatory potential of this SNP and suggesting its causal role in psychiatric disorders, including ASD. While MPRA was performed in HEK cells, the SNP was located in a regulatory element with higher chromatin accessibility in human neural progenitors compared with postmitotic neurons (Figure 6A) (unpublished data from Stein lab), indicating its regulatory potential in the developing brain. The risk allele (T) at this SNP was associated with downregulation of barcoded expression in MPRA (Figure 6B), and was predicted to disrupt two transcription factor binding motifs (89) (TBX1 and SMARCC1) (Supplementary Figure S14), providing a possible mechanism of action of this variant. We next investigated potential target genes impacted by regulatory changes at this SNP by using eQTL data from fetal (33) and adult brain tissues (32). Expression levels of three eGenes were significantly associated with this SNP (DDHD2 from the fetal brain and DDHD2, LSM1, LETM2 from the adult brain) (Figure 6A). Of these three genes, two genes (DDHD2, LETM2) showed the direction of the effect expected from the MPRA result (risk allele downregulates the eGene) (Supplementary Figure S15). It is of note that DDHD2 was identified in both fetal and adult brain eQTL datasets (beta = -0.080, $P = 2.212 \times 10^{-13}$; beta = -0.177, $P = 1.38 \times 10^{-20}$, respectively; Figure 6C, D). Notably, DDHD2 was also significantly downregulated in the post-mortem cortex of individuals with autism (logFC = -0.28, FDR = 0.013), providing an added layer of evidence supporting its role in ASD risk (72). DDHD2 also was identified by H-MAGMA (Figure 4F), and a copy number variation (CNV) containing
DDHD2 (deletions) was found in proband-sibling pairs with discordant social-behavior phenotypes (90). Collectively, by integrating existing multi-level functional genomic resources and an experimental fine-mapping approach using MPRA, we suggest DDHD2 as a strong candidate gene impacting risk for ASD.
Figure 6. Identification of putative causal variant and gene impacting risk for ASD

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

Discussion

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

Using a PRS derived from a previous study (14), we found enrichment of risk variants in SPARK cases, indicating the contribution of common genetic risk factors to ASD is consistent across cohorts. However, despite several hypotheses that rare variants associated with risk for ASD are enriched in certain subgroups of individuals with ASD, such as in females compared to males (female protective effect) (9,64–66,91), multiplex families compared to simplex families (12,61,62), or maternal alleles compared to paternal alleles (10,67), we do not find evidence to support the increased burden of common risk variants within those subgroups. This result indicates potential rare and common variant differences in contribution to subgroup risk for ASD. However, given the small sample size of specific subgroups ($N = 835$ in female whereas $N = 3,262$ in male, $N = 14$ for families with multiple affected children vs $N = 3,618$ with one affected children), our study may have limited power to identify the differences among subgroups. Thus, a larger sample size would be warranted to compare the difference in the role of common variants within these categories.
Identifying locations in the genome associated with risk for ASD does not in itself lead to insights into what tissues or developmental time points are crucial for the etiology of ASD. Here, by integrating SNP association statistics with existing annotations of regulatory elements active during specific developmental time periods or within specific brain regions, we found an excess of common genetic risk for ASD in the fetal brain regulatory elements (brain germinal matrix and primary cultured neurospheres from the fetal cortex), and progenitor enriched germinal zone of the developing cortex, confirming previous findings that alterations of gene regulation in the prenatal cortex play a key role in ASD etiology (14).

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

Since identification of a GWS locus does not elucidate the causal variant(s), we performed MPRA and identified a potential causal SNP (rs7001340) at a novel ASD locus discovered in the SPARK sample. Interestingly, the individual variant with the strongest regulatory effect (rs7001340; $r^2$ = 0.85 with the index SNP in SPARK full dataset) was different from the SNP with the strongest association with ASD (rs60527016), highlighting the importance of experimental validation in identifying causal variants. It should be noted that the regulatory effects of these variants were assessed in non-neural (HEK) cells, so further validation of these effects in ASD-relevant cell types would provide increased confidence in declaring this SNP as causal. The experimentally validated regulatory SNP (rs7001340) is in the intron of LETM2, and is also an eQTL for LETM2.
LSM1 (247 kb away) and DDHD2 (173 kb away), indicating that the SNP functions as a distal regulatory element. The risk allele (T) was associated with decreased expression of barcoded transcripts in the MPRA and downregulation of DDHD2 from eQTL in both fetal and adult brains, implying a consistent direction of the allelic effects on gene regulation. The risk allele showed the same direction of effect for LETM2 in adult brain tissue, but was not significantly associated in fetal brain tissue (P-values = 0.33). Intriguingly, DDHD2 was also downregulated in the cortex from individuals with ASD compared to neurotypical controls (72), providing an additional level of support for this gene as a risk factor for ASD. DDHD2 (DDHD domain-containing protein 2), also known as KIAA0725p, encodes a phospholipase and is localized in the Golgi (93). DDHD2 plays a role in the efficient regulation of membrane trafficking between the Golgi and cytosol (93) and is highly expressed in the brain (94–96). Mutations in DDHD2 have been previously found in individuals with spastic paraplegia type 54 (SPG) (96–98). Dhd2 null mice exhibited motor and cognitive impairments (99), which are frequent comorbidities of ASD (100). We, therefore, conclude DDHD2 is a strong candidate risk gene for ASD through multiple lines of evidence.

There is still a large amount of common variant heritability not explained in this study indicating that further increases in sample size will be necessary to explain the common inherited component of ASD risk. While the combination of TOPMed imputation and the case-pseudocontrol study model enabled us to include individuals from multiple ancestries, the case-pseudocontrol model is lower powered compared to case-unscreened control models because a pseudocontrol might have greater liability for ASD than the average individual in the population (101). In addition, the case-pseudocontrol model cannot incorporate duos or singletons due to the lack of parental genotype information, which resulted in over 2,000 individuals with ASD with genotyping information in the SPARK project not being included in our analysis. Future studies could potentially increase power by including all cases available in SPARK and using unscreened population matched controls (102). Secondly, subsequent analyses including PRS, LDSC
regression, and H-MAGMA were limited to individuals from European ancestries only, because most resources and software are designed to be used only within one population, generally European ancestry (103). Including other ancestries for these analyses will be able to uncover risk factors shared or specific to existing human populations.

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

Acknowledgments

This work was supported by a grant from the Simon Foundation (SFARI[Award #: 605259], H.W and J.L.S), NIMH (R00MH113823 and DP2MH122403 to H.W. and R01MH118349, R01MH120125, R01MH121433 to J.L.S), NIGMS (DP2GM114829 to S.K.) and the NARSAD Young Investigator Award (H.W.). LMR was supported by T32 HL129982. We are grateful to all of the families in SPARK, the SPARK clinical sites and SPARK staff. We appreciate obtaining access to genetic and phenotypic data on SFARI Base. Approved researchers can obtain the SPARK population dataset described in this study (more details available at https://www.sfari.org/resource/spark/) by applying at https://base.sfari.org. Trans-Omics in Precision Medicine (TOPMed) program imputation panel (version Freeze5) supported by the National Heart, Lung and Blood Institute (NHLBI); see www.nhlbiwgs.org.
TOPMed study investigators contributed data to the reference panel, which can be accessed through the Michigan Imputation Server; see [https://imputationserver.sph.umich.edu](https://imputationserver.sph.umich.edu). The panel was constructed and implemented by the TOPMed Informatics Research Center at the University of Michigan (3R01HL-117626-02S1; contract HHSN268201800002I). The TOPMed Data Coordinating Center (3R01HL-120393-02S1; contract HHSN268201800001I) provided additional data management, sample identity checks, and overall program coordination and support. We gratefully acknowledge the studies and participants who provided biological samples and data for TOPMed. Summary statistics from the GWAS conducted here will be available upon publication. Code is available at [https://github.com/thewonlab/GWAS_ASD_SPARK](https://github.com/thewonlab/GWAS_ASD_SPARK).

**Disclosures**

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

**References**

1. American Psychiatric Association (2013): *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*. American Psychiatric Pub.
2. Lee SH, Ripke S, Neale BM, Faraone SV, Purcell SM, Perlis RH, et al. (2013): Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet* 45: 984–994.
3. Gaugler T, Klei L, Sanders SJ, Bodea CA, Goldberg AP, Lee AB, et al. (2014): Most genetic risk for autism resides with common variation. *Nat Genet* 46: 881–885.
4. de la Torre-Ubieta L, Won H, Stein JL, Geschwind DH (2016): Advancing the understanding of autism disease mechanisms through genetics. *Nat Med* 22: 345–361.
5. Sandin S, Lichtenstein P, Kuja-Halkola R, Hultman C, Larsson H, Reichenberg A (2017): The Heritability of Autism Spectrum Disorder. *JAMA* 318: 1182–1184.
6. Bai D, Yip BHK, Windham GC, Sourander A, Francis R, Yoffe R, et al. (2019): Association of Genetic and Environmental Factors With Autism in a 5-Country Cohort. *JAMA Psychiatry*. https://doi.org/10.1001/jamapsychiatry.2019.1411

7. Michaelson JJ, Shi Y, Gujral M, Zheng H, Malhotra D, Jin X, et al. (2012): Whole-genome sequencing in autism identifies hot spots for de novo germline mutation. *Cell* 151: 1431–1442.

8. O’Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, et al. (2012): Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* 485: 246–250.

9. Sanders SJ, He X, Willsey AJ, Ercan-Sencicek AG, Samocha KE, Cicek AE, et al. (2015): Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron* 87: 1215–1233.

10. Leppa VM, Kravitz SN, Martin CL, Andrieux J, Le Caignec C, Martin-Coignard D, et al. (2016): Rare Inherited and De Novo CNVs Reveal Complex Contributions to ASD Risk in Multiplex Families. *Am J Hum Genet* 99: 540–554.

11. Kyle Satterstrom F, Kosmicki JA, Wang J, Breen MS, De Rubeis S, An J-Y, et al. (2019, April 24): Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. *bioRxiv*. p 484113.

12. Ruzzo EK, Pérez-Cano L, Jung J-Y, Wang L-K, Kashef-Haghhighi D, Hartl C, et al. (2019): Inherited and De Novo Genetic Risk for Autism Impacts Shared Networks. *Cell*, vol. 178. pp 850–866.e26.

13. Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J, et al. (2012): De novo gene disruptions in children on the autistic spectrum. *Neuron* 74: 285–299.

14. Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, et al. (2019): Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet* 51: 431–444.

15. Park J-H, Wacholder S, Gail MH, Peters U, Jacobs KB, Chanock SJ, Chatterjee N (2010):
Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nat Genet* 42: 570–575.

16. Gibson G (2012): Rare and common variants: twenty arguments. *Nat Rev Genet* 13: 135–145.

17. Robinson MR, Wray NR, Visscher PM (2014): Explaining additional genetic variation in complex traits. *Trends Genet* 30: 124–132.

18. SPARK Consortium (2018): SPARK: A US Cohort of 50,000 Families to Accelerate Autism Research. *Neuron* 97: 488–493.

19. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. (2012): Systematic localization of common disease-associated variation in regulatory DNA. *Science* 337: 1190–1195.

20. Tak YG, Farnham PJ (2015): Making sense of GWAS: using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome. *Epigenetics Chromatin* 8: 57.

21. Schaid DJ, Chen W, Larson NB (2018): From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat Rev Genet* 19: 491–504.

22. Muerdter F, Boryń ŁM, Arnold CD (2015): STARR-seq—principles and applications. *Genomics* 106: 145–150.

23. Inoue F, Ahituv N (2015): Decoding enhancers using massively parallel reporter assays. *Genomics* 106: 159–164.

24. Davis JE, Insigne KD, Jones EM, Hastings QB, Kosuri S (2019, May 13): Multiplexed dissection of a model human transcription factor binding site architecture. *bioRxiv*. p 625434.

25. Rockman MV, Kruglyak L (2006): Genetics of global gene expression. *Nat Rev Genet* 7: 862–872.

26. Nica AC, Dermitzakis ET (2013): Expression quantitative trait loci: present and future. *Philos*
27. GTEx Consortium, Laboratory, Data Analysis & Coordinating Center (LDACC)—Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) groups, NIH Common Fund, NIH/NCI, et al. (2017): Genetic effects on gene expression across human tissues. *Nature* 550: 204–213.

28. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. (2009): Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326: 289–293.

29. Sanyal A, Lajoie BR, Jain G, Dekker J (2012): The long-range interaction landscape of gene promoters. *Nature* 489: 109–113.

30. Jin F, Li Y, Dixon JR, Selvaraj S, Ye Z, Lee AY, et al. (2013): A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* 503: 290–294.

31. Nancy Y, Fauni H, Ma W, Won H (2019): Connecting gene regulatory relationships to neurobiological mechanisms of brain disorders. *bioRxiv*. Retrieved from https://www.biorxiv.org/content/10.1101/681353v1.abstract

32. Wang D, Liu S, Warrell J, Won H, Shi X, Navarro FCP, et al. (2018): Comprehensive functional genomic resource and integrative model for the human brain. *Science* 362. https://doi.org/10.1126/science.aat8464

33. Walker RL, Ramaswami G, Hartl C, Mancuso N, Gandal MJ, de la Torre-Ubieta L, et al. (2019): Genetic Control of Expression and Splicing in Developing Human Brain Informs Disease Mechanisms. *Cell* 179: 750–771.e22.

34. Loh P-R, Danecek P, Palamara PF, Fuchsberger C, A Reshef Y, K Finucane H, et al. (2016): Reference-based phasing using the Haplotype Reference Consortium panel. *Nat Genet* 48: 1443–1448.

35. Chang CC, Chow CC, Tellier LC, Vattikutti S, Purcell SM, Lee JJ (2015): Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4: 7.
36. Cordell HJ, Barratt BJ, Clayton DG (2004): Case/pseudocontrol analysis in genetic association studies: A unified framework for detection of genotype and haplotype associations, gene-gene and gene-environment interactions, and parent-of-origin effects. *Genet Epidemiol* 26: 167–185.

37. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. (2016): Next-generation genotype imputation service and methods. *Nat Genet* 48: 1284–1287.

38. Taliun D, Harris DN, Kessler MD, Carlson J, Szpiech ZA, Torres R, et al. (2019, March 6): Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *bioRxiv*. p 563866.

39. Kowalski MH, Qian H, Hou Z, Rosen JD, Tapia AL, Shan Y, et al. (2019, July 2): Use of >100,000 NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium whole genome sequences improves imputation quality and detection of rare variant associations in admixed African and Hispanic/Latino populations. *bioRxiv*. p 683201.

40. Sariya S, Lee JH, Mayeux R, Vardarajan BN, Reyes-Dumeyer D, Manly JJ, et al. (2019): Rare Variants Imputation in Admixed Populations: Comparison Across Reference Panels and Bioinformatics Tools. *Front Genet* 10: 239.

41. Willer CJ, Li Y, Abecasis GR (2010): METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26: 2190–2191.

42. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. (2019): The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* 47: D1005–D1012.

43. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics Consortium, et al. (2015): LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* 47: 291–295.

44. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh P-R, et al. (2015): Partitioning heritability by functional annotation using genome-wide association summary
statistics. Nat Genet 47: 1228–1235.

45. Choi SW, O'Reilly PF (2019): PRSice-2: Polygenic Risk Score software for biobank-scale data. Gigascience 8. https://doi.org/10.1093/gigascience/giz082

46. Won H, de la Torre-Ubieta L, Stein JL, Parikshak NN, Huang J, Opland CK, et al. (2016): Chromosome conformation elucidates regulatory relationships in developing human brain. Nature 538: 523–527.

47. Pardiñas AF, Holmans P, Pocklington AJ, Escott-Price V, Ripke S, Carrera N, et al. (2018): Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. Nat Genet 50: 381–389.

48. Law CW, Chen Y, Shi W, Smyth GK (2014): voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol 15: R29.

49. Myint L, Avramopoulos DG, Goff LA, Hansen KD (2019): Linear models enable powerful differential activity analysis in massively parallel reporter assays. BMC Genomics 20: 209.

50. Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium (2017): Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. Mol Autism 8: 21.

51. Smith DJ, Escott-Price V, Davies G, Bailey MES, Colodro-Conde L, Ward J, et al. (2016): Genome-wide analysis of over 106,000 individuals identifies 9 neuroticism-associated loci. Mol Psychiatry 21: 749–757.

52. Luciano M, Hagenaars SP, Davies G, Hill WD, Clarke T-K, Shirali M, et al. (2018): Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism. Nat Genet 50: 6–11.

53. Nagel M, Watanabe K, Stringer S, Posthuma D, van der Sluis S (2018): Item-level analyses reveal genetic heterogeneity in neuroticism. Nat Commun 9: 905.

54. Nagel M, Jansen PR, Stringer S, Watanabe K, de Leeuw CA, Bryois J, et al. (2018): Meta-
analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. *Nat Genet* 50: 920–927.

55. Kichaev G, Bhatia G, Loh P-R, Gazal S, Burch K, Freund MK, *et al.* (2019): Leveraging Polygenic Functional Enrichment to Improve GWAS Power. *Am J Hum Genet* 104: 65–75.

56. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, *et al.* (2018): Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet* 50: 1112–1121.

57. Adams HHH, Hibar DP, Chouraki V, Stein JL, Nyquist PA, Renteria ME, *et al.* (2016): Novel genetic loci underlying human intracranial volume identified through genome-wide association. *Nat Neurosci* 19: 1569–1582.

58. Hansen SN, Overgaard M, Andersen PK, Parner ET (2017): Estimating a population cumulative incidence under calendar time trends. *BMC Med Res Methodol* 17: 7.

59. Peterson RE, Kuchenbaecker K, Walters RK, Chen C-Y, Popejoy AB, Periyasamy S, *et al.* (2019): Genome-wide Association Studies in Ancestrally Diverse Populations: Opportunities, Methods, Pitfalls, and Recommendations. *Cell* 179: 589–603.

60. Duncan L, Shen H, Gelaye B, Meijsen J, Ressler K, Feldman M, *et al.* (2019): Analysis of polygenic risk score usage and performance in diverse human populations. *Nat Commun* 10: 3328.

61. Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, *et al.* (2007): Strong association of de novo copy number mutations with autism. *Science* 316: 445–449.

62. Ronemus M, Iossifov I, Levy D, Wigler M (2014): The role of de novo mutations in the genetics of autism spectrum disorders. *Nat Rev Genet* 15: 133–141.

63. Loomes R, Hull L, Mandy WPL (2017): What Is the Male-to-Female Ratio in Autism Spectrum Disorder? A Systematic Review and Meta-Analysis. *J Am Acad Child Adolesc Psychiatry* 56: 466–474.

64. Levy D, Ronemus M, Yamrom B, Lee Y-H, Leotta A, Kendall J, *et al.* (2011): Rare de novo
and transmitted copy-number variation in autistic spectrum disorders. *Neuron* 70: 886–897.

65. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, et al. (2012): De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* 485: 237–241.

66. Dong S, Walker MF, Carriero NJ, DiCola M, Willsey AJ, Ye AY, et al. (2014): De Novo Insertions and Deletions of Predominantly Paternal Origin Are Associated with Autism Spectrum Disorder. *Cell Rep* 9: 16–23.

67. Zhao X, Leotta A, Kustanovich V, Lajonchere C, Geschwind DH, Law K, et al. (2007): A unified genetic theory for sporadic and inherited autism. *Proc Natl Acad Sci U S A* 104: 12831–12836.

68. Hazlett HC, Gu H, Munsell BC, Kim SH, Styner M, Wolff JJ, et al. (2017): Early brain development in infants at high risk for autism spectrum disorder. *Nature* 542: 348–351.

69. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. (2015): Integrative analysis of 111 reference human epigenomes. *Nature* 518: 317–330.

70. de la Torre-Ubieta L, Stein JL, Won H, Opland CK, Liang D, Lu D, Geschwind DH (2018): The Dynamic Landscape of Open Chromatin during Human Cortical Neurogenesis. *Cell* 172: 289–304.e18.

71. Zhou J, Park CY, Theesfeld CL, Wong AK, Yuan Y, Scheckel C, et al. (2019): Whole-genome deep-learning analysis identifies contribution of noncoding mutations to autism risk. *Nat Genet* 51: 973–980.

72. Parikshak NN, Swarup V, Belgard TG, Irimia M, Ramaswami G, Gandal MJ, et al. (2016): Genome-wide changes in IncRNA, splicing, and regional gene expression patterns in autism. *Nature* 540: 423–427.

73. Reimand J, Kull M, Peterson H, Hansen J, Vilo J (2007): g:Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res* 35: W193–W200.
746 74. Kang HJ, Kawasawa YI, Cheng F, Zhu Y, Xu X, Li M, et al. (2011): Spatio-temporal transcriptome of the human brain. *Nature* 478: 483–489.

747 75. Simonoff E, Pickles A, Charman T, Chandler S, Loucas T, Baird G (2008): Psychiatric disorders in children with autism spectrum disorders: prevalence, comorbidity, and associated factors in a population-derived sample. *J Am Acad Child Adolesc Psychiatry* 47: 921–929.

748 76. van Steensel FJA, Bögels SM, Perrin S (2011): Anxiety disorders in children and adolescents with autistic spectrum disorders: a meta-analysis. *Clin Child Fam Psychol Rev* 14: 302–317.

749 77. Antshel KM, Zhang-James Y, Wagner KE, Ledesma A, Faraone SV (2016): An update on the comorbidity of ADHD and ASD: a focus on clinical management. *Expert Rev Neurother* 16: 279–293.

750 78. Romero M, Aguilar JM, Del-Rey-Mejías Á, Mayoral F, Rapado M, Peciña M, et al. (2016): Psychiatric comorbidities in autism spectrum disorder: A comparative study between DSM-IV-TR and DSM-5 diagnosis. *Int J Clin Health Psychol* 16: 266–275.

751 79. Cross-Disorder Group of the Psychiatric Genomics Consortium (2019): Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. *Cell* 179: 1469–1482.e11.

752 80. Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E, et al. (2019): Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat Genet* 51: 63–75.

753 81. Savage JE, Jansen PR, Stringer S, Watanabe K, Bryois J, de Leeuw CA, et al. (2018): Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat Genet* 50: 912–919.

754 82. Stahl EA, Breen G, Forstner AJ, McQuillen A, Ripke S, Trubetskoy V, et al. (2019): Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat Genet* 51:
83. Howard DM, Adams MJ, Clarke T-K, Hafferty JD, Gibson J, Shirali M, et al. (2019): Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat Neurosci* 22: 343–352.

84. Pasman JA, Verweij KJH, Gerring Z, Stringer S, Sanchez-Roige S, Treur JL, et al. (2018): GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal influence of schizophrenia. *Nat Neurosci* 21: 1161–1170.

85. Liu M, Jiang Y, Wedow R, Li Y, Brazel DM, Chen F, et al. (2019): Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat Genet* 51: 237–244.

86. Jansen IE, Savage JE, Watanabe K, Bryois J, Williams DM, Steinberg S, et al. (2019): Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer’s disease risk. *Nat Genet* 51: 404–413.

87. Nalls MA, Blauwendraat C, Vallerga CL, Heilbron K, Bandres-Ciga S, Chang D, et al. (2018, August 9): Expanding Parkinson’s disease genetics: novel risk loci, genomic context, causal insights and heritable risk [no. biorxiv;388165v3]. *Genetics*. bioRxiv, p 223.

88. Brainstorm Consortium, Anttila V, Bulik-Sullivan B, Finucane HK, Walters RK, Bras J, et al. (2018): Analysis of shared heritability in common disorders of the brain. *Science* 360. https://doi.org/10.1126/science.aap8757

89. Coetze SG, Coetzee GA, Hazelett DJ (2015): motifbreakR: an R/Bioconductor package for predicting variant effects at transcription factor binding sites. *Bioinformatics* 31: 3847–3849.

90. Krumm N, O’Roak BJ, Karakoc E, Mohajeri K, Nelson B, Vives L, et al. (2013): Transmission disequilibrium of small CNVs in simplex autism. *Am J Hum Genet* 93: 595–606.

91. Jacquemont S, Coe BP, Hersch M, Duyzend MH, Krumm N, Bergmann S, et al. (2014): Higher Mutational Burden in Females Supports a “Female Protective Model” in
Neurodevelopmental Disorders. *Am J Hum Genet* 94: 415–425.

92. Parikshak NN, Luo R, Zhang A, Won H, Lowe JK, Chandran V, *et al.* (2013): Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* 155: 1008–1021.

93. Sato S-I, Inoue H, Kogure T, Tagaya M, Tani K (2010): Golgi-localized KIAA0725p regulates membrane trafficking from the Golgi apparatus to the plasma membrane in mammalian cells. *FEBS Letters*, vol. 584. pp 4389–4395.

94. Nagase T, Ishikawa K-I, Suyama M, Kikuno R, Hirosawa M, Miyajima N, *et al.* (1998): Prediction of the coding sequences of unidentified human genes. XII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro. *DNA Res* 5: 355–364.

95. Nakajima K-I, Sonoda H, Mizoguchi T, Aoki J, Arai H, Nagahama M, *et al.* (2002): A novel phospholipase A1 with sequence homology to a mammalian Sec23p-interacting protein, p125. *J Biol Chem* 277: 11329–11335.

96. Schuurs-Hoeijmakers JHM, Geraghty MT, Kamsteeg E-J, Ben-Salem S, de Bot ST, Nijhof B, *et al.* (2012): Mutations in DDHD2, encoding an intracellular phospholipase A(1), cause a recessive form of complex hereditary spastic paraplegia. *Am J Hum Genet* 91: 1073–1081.

97. Gonzalez M, Nampoothiri S, Kornblum C, Oteyza AC, Walter J, Konidari I, *et al.* (2013): Mutations in phospholipase DDHD2 cause autosomal recessive hereditary spastic paraplegia (SPG54). *Eur J Hum Genet* 21: 1214–1218.

98. Novarino G, Fenstemaker AG, Zaki MS, Hofree M, Silhavy JL, Heiberg AD, *et al.* (2014): Exome sequencing links corticospinal motor neuron disease to common neurodegenerative disorders. *Science* 343: 506–511.

99. Inloes JM, Hsu K-L, Dix MM, Viader A, Masuda K, Takei T, *et al.* (2014): The hereditary spastic paraplegia-related enzyme DDHD2 is a principal brain triglyceride lipase. *Proc Natl
100. Sokhadze EM, Tasman A, Sokhadze GE, El-Baz AS, Casanova MF (2016): Behavioral, Cognitive, and Motor Preparation Deficits in a Visual Cued Spatial Attention Task in Autism Spectrum Disorder. *Appl Psychophysiol Biofeedback* 41: 81–92.

101. Peyrot WJ, Boomsma DI, Penninx BWJH, Wray NR (2016): Disease and Polygenic Architecture: Avoid Trio Design and Appropriately Account for Unscreened Control Subjects for Common Disease. *Am J Hum Genet* 98: 382–391.

102. Bodea CA, Neale BM, Ripke S, International IBD Genetics Consortium, Daly MJ, Devlin B, Roeder K (2016): A Method to Exploit the Structure of Genetic Ancestry Space to Enhance Case-Control Studies. *Am J Hum Genet* 98: 857–868.

103. Sirugo G, Williams SM, Tishkoff SA (2019): The Missing Diversity in Human Genetic Studies. *Cell* 177: 1080.

**Legends**

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

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

Figure 2. Comparison of polygenic risk scores between subgroups.

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

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

(A) Heritability enrichment in active enhancer and promoter regions present in different tissues shows the critical role of developing the brain in ASD etiology. (B) Heritability enrichment in differential chromatin accessibility from the developing fetal cortical wall. (upper) The x-axis represents tissue types and the y-axis indicates heritability enrichment. The error bar shows a 95% confidence interval. (lower) The x-axis represents tissue types and the y-axis indicates statistical significance as -log10(FDR). BRN: Brain, ADRL: Adrenal, BLD: Blood, BRST: Breast, CRVX: Cervix, ESDR: ESC Derived, GI: GI_duodenum, GI_colon, GI_rectum, GI_stomach, GI_intestine, GI_colon, GI_rectum and GI_esophagus, HRT: Heart, KID: Kidney, LIV: Liver, LNG: Lung, MUS: Muscle, OVRY: Ovary, PANC: Pancreas, PLCNT: Placenta, SPLN: Spleen, STRM: Stromal connective, THYM: Thymus, VAS: Vascular, CP: Peaks more accessible in cortical plate, GZ: Peaks more accessible in germinal zone. * FDR < 0.05, ** FDR < 0.01.

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

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

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

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

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

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