nature genetics

ARTICLES

https://doi.org/10.1038/s41588-019-0344-8

Identification of common genetic risk variants for autism spectrum disorder

Jakob Grove[1,2,3,4], Stephan Ripke[5,6,7], Thomas D. Als[1,2,3], Manuel Mattheisen[1,2,3,8,9], Raymond K. Walters[5,6], Hyejung Won[10,11], Jonatan Pallesen[1,2,3], Esben Agerbo[1,12,13], Ole A. Andreassen[14,15], Richard Anney[16], Swapnil Awashti[7], Rich Belliveau[6], Francesco Bettella[14,15], Joseph D. Buxbaum[17,18,19,20], Jonas Bybjerg-Grauholm[1,21], Marie Bækvad-Hansen[1,21], Felecia Cerrato[6], Kimberly Chambert[6], Jane H. Christensen[1,2,3], Claire Churchhouse[5,6,22], Karin Dellenvall[23], Ditte Demontis[1,2,3], Silvia De Rubeis[17,18], Bernie Devlin[24], Srdjan Djurovic[14,25], Ashley L. Dumont[6], Jacqueline I. Goldstein[5,6,22], Christine S. Hansen[1,21,26], Mads Engel Hauberg[1,2,3], Mads V. Hollegaard[1,21], Sigrun Hope[14,27], Daniel P. Howrigan[5,6], Hailiang Huang[5,6], Christina M. Hultman[23], Lambertus Klei[24], Julian Maller[6,28,29], Joanna Martin[6,16,23], Alicia R. Martin[5,6,22], Jennifer L. Moran[6], Mette Nyegaard[1,2,3], Terje Nærland[14,30], Duncan S. Palmer[5,6], Aarno Palotie[5,6,22,31], Carsten Bøcker Pedersen[1,12,13], Marianne Giørtz Pedersen[1,12,13], Timothy dPoterba[5,6,22], Jesper Buchhave Poulsen[1,21], Beate St Pourcain[32,33,34], Per Qvist[1,2,3], Karola Rehnström[35], Abraham Reichenberg[17,18,19], Jennifer Reichert[17,18], Elise B. Robinson[5,6,36], Kathryn Roeder[37,38], Panos Roussos[18,39,40,41], Evald Saemundsen[42], Sven Sandin[17,18,23], F. Kyle Satterstrom[5,6,22], George Davey Smith[33,43], Hreinn Stefansson[44], Stacy Steinberg[44], Christine R. Stevens[6], Patrick F. Sullivan[10,23,45], Patrick Turley[5,6], G. Bragi Walters[44,46], Xinyi Xu[17,18], Autism Spectrum Disorder Working Group of the Psychiatric Genomics Consortium[47], BUPGEN[47], Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium[48], 23andMe Research Team[48], Kari Stefansson[44,46], Daniel H. Geschwind[49,50,51], Merete Nordentoft[1,52], David M. Hougaard[1,21], Thomas Werge[1,26,53], Ole Mors[1,54], Preben Bo Mortensen[1,2,12,13], Benjamin M. Neale[5,6,22], Mark J. Daly[5,6,22,31*] and Anders D. Børglum[1,2,3*]

Autism spectrum disorder (ASD) is a highly heritable and heterogeneous group of neurodevelopmental phenotypes diagnosed in more than 1% of children. Common genetic variants contribute substantially to ASD susceptibility, but to date no individual variants have been robustly associated with ASD. With a marked sample-size increase from a unique Danish population resource, we report a genome-wide association meta-analysis of 18,381 individuals with ASD and 27,969 controls that identified five genome-wide-significant loci. Leveraging GWAS results from three phenotypes with significantly overlapping genetic architectures (schizophrenia, major depression, and educational attainment), we identified seven additional loci shared with other traits at equally strict significance levels. Dissecting the polygenic architecture, we found both quantitative and qualitative polygenic heterogeneity across ASD subtypes. These results highlight biological insights, particularly relating to neuronal function and corticogenesis, and establish that GWAS performed at scale will be much more productive in the near term in ASD.

ASD is the term for a group of pervasive neurodevelopmental disorders characterized by impaired social and communication skills along with repetitive and restrictive behavior. The clinical presentation is highly heterogeneous, including individuals with severe impairment and intellectual disability (ID) as well as individuals with above-average intelligence quotient (IQ) and high levels of academic and occupational functioning. ASD affects 1–1.5% of individuals and is highly heritable, and both common and rare variants contribute to its etiology[1–4]. Common variants have been estimated to account for a major part of ASD liability[2], as has been observed for other common neuropsychiatric disorders. In contrast, de novo mutations, mostly copy number variants (CNVs) and gene-disrupting point mutations, have larger individual effects but collectively explain <5% of the overall liability[1–3] and far less of the heritability. Although a number of genes have been convincingly implicated via excess statistical aggregation of de novo mutations, the largest genome-wide association study (GWAS) to date ($n = 7{,}387$ cases scanned)—although providing compelling evidence for the bulk contribution of common variants—did not conclusively identify single variants at genome-wide significance[5–7].

A full list of affiliations appears at the end of the paper.

These results underscore that common variants, as in other complex diseases such as schizophrenia, individually have low impact and that a substantial scale-up in sample numbers would be needed.

Here we report what are, to our knowledge, the first common risk variants robustly associated with ASD, by more than doubling the discovery sample size relative to that in previous GWAS[5–8]. We describe strong genetic correlations between ASD and other complex disorders and traits, confirming shared etiology, and we show results indicating differences in the polygenic architecture across clinical subtypes of ASD. Leveraging these relationships and recently introduced computational techniques[9], we identify additional previously undescribed ASD-associated variants that are shared with other phenotypes. Furthermore, by integrating with complementary data from Hi-C chromatin-interaction analysis of fetal brains and brain transcriptome data, we explore the functional implications of our top-ranking GWAS results.

Results

GWAS. As part of the iPSYCH project[10], we collected and genotyped a Danish nationwide population-based case–cohort sample including nearly all individuals born in Denmark between 1981 and 2005 and diagnosed with ASD (according to the International Statistical Classification of Diseases and Related Health Problems, 10th revision (ICD-10)) before 2014. We randomly selected controls from the same birth cohorts (Supplementary Table 1). We previously validated registry-based ASD diagnoses[11,12] and demonstrated the accuracy of genotyping DNA extracted and amplified from blood spots collected shortly after birth[13,14]. Genotypes were processed with Ricopili[15], performing stringent quality control of data, removal of related individuals, exclusion of ancestry outliers based on principal component analysis (PCA), and imputation by using the 1000 Genomes Project phase 3 reference panel. After this processing, genotypes from 13,076 cases and 22,664 controls from the iPSYCH sample were included in the analysis. As is now standard in human complex-trait genomics, our primary analysis was a meta-analysis of the iPSYCH ASD results with five family-based trio samples of European ancestry from the Psychiatric Genomics Consortium (PGC; 5,305 cases and 5,305 pseudocontrols)[16]. All PGC samples had been processed with the same Ricopili pipeline for quality control, imputation, and analysis as used here.

Supporting the consistency between the study designs, the iPSYCH population-based and PGC family-based analyses showed a high degree of genetic correlation with $r_G = 0.779$ (s.e.m. = 0.106; $P = 1.75 \times 10^{-13}$), findings similar to the genetic correlations observed between datasets in other mental disorders[17]. Likewise, polygenicity, as assessed by polygenic risk scores (PRSs), showed consistency across the samples, thus supporting homogeneity of the effects across samples and study designs (results below regarding PRSs on a five-way split of the sample). The SNP heritability (h^2_G) was estimated to be 0.118 (s.e.m. = 0.010), for a population prevalence of 0.012 (ref. [18]).

The main GWAS meta-analysis included a total of 18,381 ASD cases and 27,969 controls, and applied an inverse-variance-weighted fixed-effects model. To ensure that the analysis was well powered and robust, we examined markers with minor-allele frequency (MAF) ≥0.01 and imputation INFO score ≥0.7, which were supported by an effective sample size in >70% of the total. This final meta-analysis included results for 9,112,387 autosomal markers and yielded 93 genome-wide-significant markers in three separate loci (Fig. 1, Table 1a and Supplementary Figs. 1–44). Each locus was strongly supported by both the Danish case–control and the PGC family-based data. Although modest inflation was observed (lambda = 1.12, lambda1000 = 1.006), linkage disequilibrium (LD)-score regression analysis[19] indicated that this finding arose from polygenicity (>96%; Methods) rather than confounding. The strongest signal among 294,911 markers analyzed on the X chromosome was $P = 7.8 \times 10^{-5}$.

We next obtained replication data for the top 88 loci with P values $< 1 \times 10^{-5}$ in five cohorts of European ancestry, including a total of 2,119 additional cases and 142,379 controls (Supplementary Table 2 and 3). An overall replication of the direction of effects was observed (53 of 88 (60%) of $P < 1 \times 10^{-5}$; 16 of 23 (70%) at $P < 1 \times 10^{-6}$; sign tests, $P = 0.035$ and $P = 0.047$, respectively), and two additional loci achieved genome-wide significance in the combined analysis (Table 1a). More details on the identified loci can be found in Supplementary Table 4, and selected candidates are described in Box 1.

Correlation with other traits and multitrait GWAS. To investigate the extent of genetic overlap between ASD and other phenotypes, we estimated the genetic correlations with a broad set of psychiatric and other medical diseases, disorders, and traits available at LD Hub[20], by using bivariate LD-score regression (Fig. 2 and Supplementary Table 5). Significant correlations were found for several traits including schizophrenia[15] ($r_G = 0.211$, $P = 1.03 \times 10^{-5}$) and measures of cognitive ability, especially educational attainment[21] ($r_G = 0.199$, $P = 2.56 \times 10^{-9}$), thus indicating a substantial genetic overlap with these phenotypes and corroborating previous reports[5,22–24]. In contrast to findings in previous reports[16], we find a strong and highly significant correlation with major depression[25] ($r_G = 0.412$, $P = 1.40 \times 10^{-25}$), and we report a novel and prominent overlap with ADHD[26] ($r_G = 0.360$, $P = 1.24 \times 10^{-12}$). Moreover, we confirm the genetic correlation with social communication difficulties at age 8 in a non-ASD population sample previously reported and based on a subset of the ASD sample[27] ($r_G = 0.375$, $P = 0.0028$).

To leverage these observations for the discovery of loci that may be shared between ASD and these other traits, we selected three particularly well-powered and genetically correlated phenotypes. These were schizophrenia ($n = 79{,}641$)[15], major depression ($n = 424{,}015$)[25], and educational attainment ($n = 328{,}917$)[21]. We used the recently introduced MTAG method[9] which, in brief, generalizes the standard inverse-variance-weighted meta-analysis for multiple phenotypes. In this case, MTAG takes advantage of how, given an overall genetic correlation between ASD and a second trait, the effect-size estimate and evidence for association to ASD can be improved by appropriate use of the association information from the second trait. The results of these three ASD-anchored MTAG scans are correlated to the primary ASD scan (and to each other), but given the exploration of three scans, we used a more conservative threshold of 1.67×10^{-8} for declaring significance across these secondary scans giving an estimated maximum false discovery rate (maxFDR) of 0.021. In addition to stronger evidence for several of the ASD hits defined above, variants in seven additional regions achieved genome-wide significance, including three loci shared with educational attainment and four shared with major depression (Table 1b, Box 1, Supplementary Table 6 and Supplementary Figs. 49–55). We note that in these seven instances, the effect-size estimate is stronger in ASD than the secondary trait, and the result is not characteristic of the strongest signals in these other scans (Supplementary Table 7–9) (and in fact, three of these seven were not significant in the secondary trait and constitute potentially novel findings). Moreover, we benchmarked against MTAG running two very large and heritable traits (height[28], $n = 252{,}288$ and body mass index (BMI)[29], $n = 322{,}154$) with no expected links to ASD, and no significant loci were added to the list of ASD-only significant associations.

Gene and gene-set analysis. Next, we performed gene-based association analysis on our primary ASD meta-analysis by using MAGMA[30], testing for the joint association of all markers within a locus (across all protein-coding genes in the genome). This analysis identified 15 genes surpassing the significance threshold

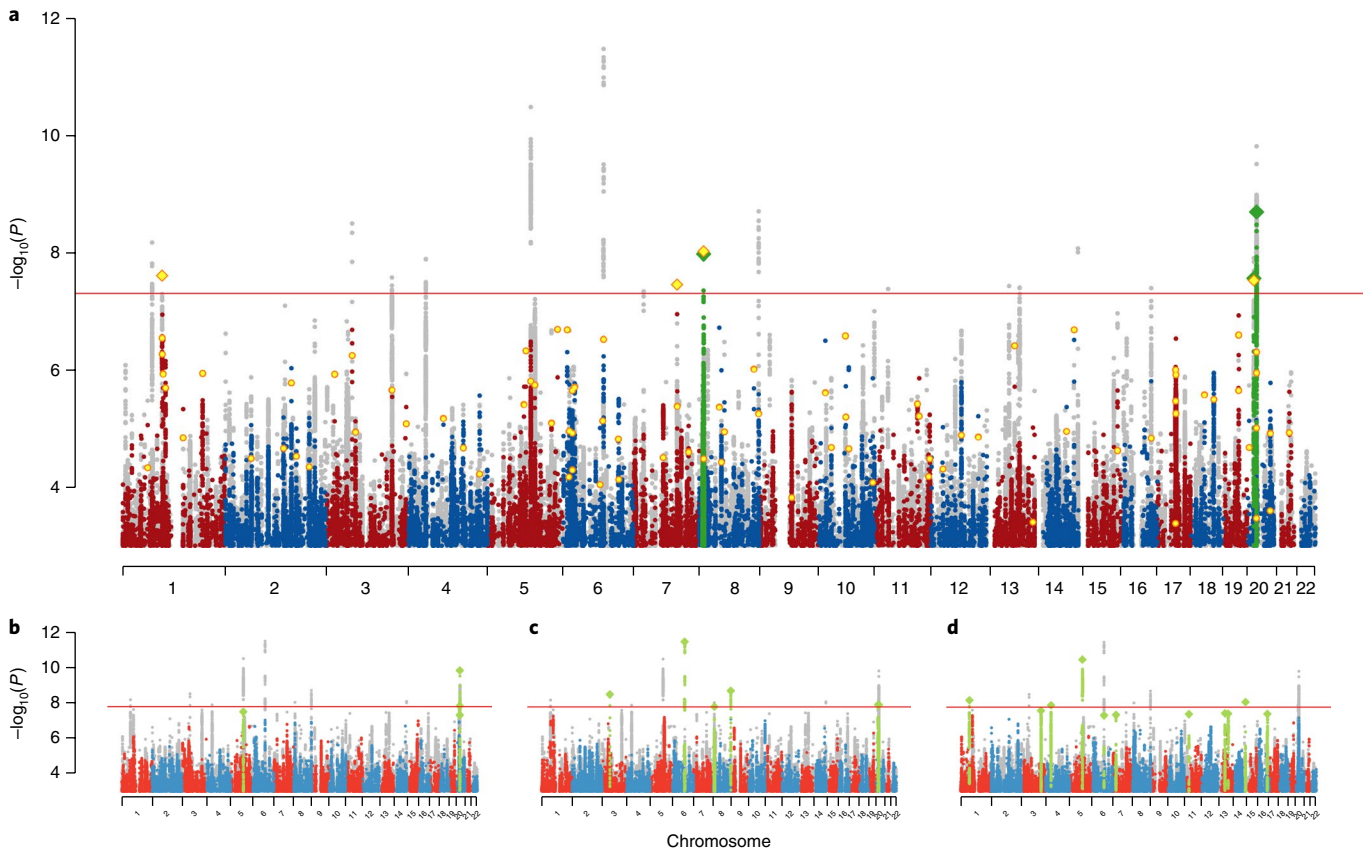

Fig. 1 | Manhattan plots. The *x* axis shows genomic position (chromosomes 1–22), and the *y* axis shows statistical significance as $-\log_{10}(P)$ of *z* statistics. **a**, The main ASD scan (18,381 cases and 27,969 controls), with the results of the combined analysis with the follow-up sample (2,119 cases and 142,379 controls) in yellow in the foreground. Genome-wide-significant clumps are green, and index SNPs are shown as diamonds. **b**–**d**: Manhattan plots for three MTAG scans of ASD together with schizophrenia[15] (34,129 cases and 45,512 controls; **b**), educational attainment[21] ($n=328{,}917$; **c**) and major depression[25] (111,902 cases and 312,113 controls; **d**). Full-size plots are shown in Supplementary Figs. 45–48. In all panels, the results of the composite of the five analyses (consisting of the minimal *P* value of the five for each marker) is shown in gray in the background.

(Supplementary Table 10). As expected, most of these genes were located within the genome-wide-significant loci identified in the GWAS, but seven genes were located in four additional loci: *KCNN2*, *MMP12*, *NTM*, and a cluster of genes on chromosome 17 (*KANSL1*, *WNT3*, *MAPT*, and *CRHR1*) (Supplementary Figs. 57–71). In particular, *KCNN2* was strongly associated ($P=1.02\times10^{-9}$), far beyond even single-variant statistical thresholds, and is included in the descriptions in Box 1.

Enrichment analyses using gene coexpression modules from human neocortex transcriptomic data (M13, M16, and M17 from Parikshak et al.[31]) and loss-of-function intolerant genes (probability of loss-of-function intolerance, pLI >0.9)[32,33], for which there is evidence of enrichment in neurodevelopmental disorders[26,31,34], yielded only nominal significance for the latter ($P=0.014$) and M16 ($P=0.050$) (Supplementary Table 11). Genes implicated in ASD by studies or rare variants in Sanders et al.[35] were just shy of showing nominally significant enrichment ($P=0.063$), whereas enrichment in the curated gene list from the SPARK consortium[36] was significant ($P=0.0034$). Likewise, analysis of Gene Ontology sets[37,38] for molecular function from the Molecular Signatures Database (MsigDB)[39] showed no significant sets after Bonferroni correction for multiple testing (Supplementary Table 12).

Dissection of the polygenic architecture. Because ASD is a highly heterogeneous disorder, we explored how h^2_G partitioned across phenotypic subcategories in the iPSYCH sample, and we estimated the genetic correlations among these groups by using GCTA[40]. We examined cases with ($n=1{,}873$) and those without ID and the ICD-10 diagnostic subcategories of childhood autism (F84.0, $n=3{,}310$), atypical autism (F84.1, $n=1{,}607$), Asperger's syndrome (F84.5, $n=4{,}622$), and other/unspecified pervasive developmental disorders (PDDs, F84.8-9, $n=5{,}795$), reducing to nonoverlapping groups when performing pairwise comparisons (Supplementary Table 13). Whereas the pairwise genetic correlations were consistently high among all subgroups (95% confidence intervals (CIs) including 1 in all comparisons), the h^2_G of Asperger's syndrome ($h^2_G=0.097$, s.e.m. $=0.001$) was found to be twice the h^2_G of both childhood autism ($h^2_G=0.049$, s.e.m. $=0.009$, $P=0.001$) and the group of other/unspecified PDDs ($h^2_G=0.045$, s.e.m. $=0.008$, $P=0.001$) (Supplementary Tables 14 and 15 and Supplementary Figs. 82 and 83). Similarly, the h^2_G of ASD without ID ($h^2_G=0.086$, s.e.m. $=0.005$) was three times higher than that for cases with ID ($h^2_G=0.029$, s.e.m. $=0.013$, $P=0.015$).

To further examine the apparent polygenic heterogeneity across subtypes, we investigated how PRSs trained on different phenotypes were distributed across distinct ASD subgroups. We focused on phenotypes showing strong genetic correlation with ASD (for example, educational attainment) but also included traits with little or no correlation to ASD (for example, BMI) as negative controls. In this analysis, we regressed the normalized scores on ASD subgroups while including covariates for batches and principal components (PCs) in a multivariate regression. Of the eight phenotypes evaluated, only the cognitive phenotypes showed strong heterogeneity (educational attainment[21], $P=1.8\times10^{-8}$; IQ[41], $P=3.7\times10^{-9}$)

Table 1 | Genome-wide-significant loci from ASD scans and MTAG analyses

	Index variant	Chr	BP	Analysis	*P*	*β*	s.e.	A1/A2	FRQ	Support from other scans			Nearest genes
										Scan	*P*	*β*	
a	rs910805	20	21248116	ASD	2.04×10^{-9}	−0.096	0.016	A/G	0.760	ASD-SCZ	1.5×10^{-10}	−0.069	*KIZ*, *XRN2*,
										ASD-Edu*	2.0×10^{-8}	−0.061	*NKX2-2*, *NKX2-4*
	rs10099100	8	10576775	ASD	1.07×10^{-8}	0.084	0.015	C/G	0.331	Comb ASD	9.6×10^{-9}	0.078	*C8orf74*, *SOX7*,
										ASD-Edu	1.6×10−8	0.056	*PINX1*
	rs201910565	1	96561801	Comb ASD	2.48×10^{-8}	−0.077	0.014	A/AT	0.689	ASD	3.4×10^{-7}	−0.033	*LOC102723661*, *PTBP2*
	rs71190156	20	14836243	ASD	2.75×10^{-8}	−0.078	0.014	GTTTT	0.481	Comb ASD	3.0×10−8	−0.072	*MACROD2*
								TTT/G		ASD-Edu	1.2×10^{-8}	0.053	
	rs111931861	7	104744219	Comb ASD	3.53×10^{-8}	−0.216	0.039	A/G	0.966	ASD	1.1×10^{-7}	−0.094	*KMT2E*, *SRPK2*
b	rs2388334	6	98591622	ASD-Edu	3.34×10^{-12}	−0.065	0.009	A/G	0.517	ASD	1.0×10^{-6}	−0.068	*MMS22L*, *POU3F2*
	rs325506	5	104012303	ASD-MD	3.26×10^{-11}	0.057	0.009	C/G	0.423	ASD	3.5×10^{-7}	0.071	*NUD12*
	rs11787216	8	142615222	ASD-Edu	1.99×10^{-9}	−0.058	0.010	T/C	0.364	ASD	2.6×10^{-6}	−0.030	*MROH5*
	rs1452075	3	62481063	ASD-Edu	3.17×10^{-9}	0.061	0.010	T/C	0.721	ASD	2.1×10^{-7}	0.035	*CADPS*
	rs1620977	1	72729142	ASD-MD	6.66×10^{-9}	0.056	0.010	A/G	0.260	ASD	1.2×10^{-4}	0.062	*NEGR1*
	rs10149470	14	104017953	ASD-MD	8.52×10^{-9}	−0.049	0.008	A/G	0.487	ASD	8.5×10^{-5}	−0.056	*MARK3*, *CKB*, *TRMT61A*, *BAG5*, *APOPT1*, *KLC1*, *XRCC3*
	rs16854048	4	42123728	ASD-MD	1.29×10^{-8}	0.069	0.012	A/C	0.858	ASD	5.9×10^{-5}	0.082	*SLC30A9*, *BEND4*, *TMEM33*, *DCAF4L1*

a, Loci reaching genome-wide significance in analysis of the ASD phenotype alone. The 'analysis' column indicates the minimum *P* value arising from the original scan (ASD) and the combined analysis with the follow-up sample (Comb ASD). The column 'support from other scans' lists the other analyses (including MTAG) that further support the locus at genome-wide significance. For the ASD scan results, genome-wide-significant results in the locus from the other scans are shown; for Comb ASD, the results from ASD are displayed. **b**, Additional genome-wide-significant loci identified in the three MTAG analyses. The three analyses are ASD with schizophrenia (SCZ)[15], educational attainment (Edu)[21], and major depression (MD)[25]. Here the 'analysis' column indicates which MTAG analysis gave the results (ASD-Edu or ASD-MD), and the columns 'support from other scans' provide the corresponding scan results in ASD alone. In both **a** and **b**, independent loci are defined to have r^2 <0.1 and distance >400 kb, and the index variant is displayed in the column 'index var'. Chr, chromosome; BP, chromosomal position; A1/A2, alleles; FRQ, allele frequency of A1; *β*, estimate of effect with respect to A1; s.e., standard error of *β*; *P*, association *P* value of the index variant (P). 'Nearest genes' lists nearest genes from within 50 kb of the region spanned by all SNPs with r^2 ≥0.6 to the index variant. Asterisks indicate a different lead SNP from the index variant.

(Supplementary Fig. 84). Interestingly, all case–control groups with or without ID showed significantly different loading for the two cognitive phenotypes: controls with ID had the lowest score, followed by ASD cases with ID, and ASD cases without ID again had significantly higher scores than those of any other group (educational attainment, $P=2.6\times10^{-12}$; IQ, $P=8.2\times10^{-12}$).

With respect to the diagnostic subcategories constructed hierarchically from ASD subtypes (Supplementary Table 13), the cognitive phenotypes again showed the strongest heterogeneity across the diagnostic classes (educational attainment, $P=2.6\times10^{-11}$; IQ, $P=3.4\times10^{-8}$), whereas neuroticism[23] ($P=0.0015$), chronotype[42] ($P=0.011$), and subjective well-being[23] ($P=0.029$) showed a weaker but nominally significant degree of heterogeneity, and schizophrenia, major depressive disorder, and BMI[29] were nonsignificant across the groups ($P>0.19$) (Fig. 3). This pattern weakened only slightly when we excluded subjects with ID (Supplementary Fig. 85). For neuroticism, there was a clear split, with atypical and other/unspecified PDD cases having significantly higher PRSs than childhood autism and Asperger's syndrome, $P=0.00013$. Given the genetic overlap of each subcategory with each phenotype, the hypothesis of homogeneity across subphenotypes was strongly rejected ($P=1.6\times10^{-11}$), thereby establishing that these subcategories indeed have differences in their genetic architectures.

Focusing on educational attainment, we found a significant enrichment of PRSs for Asperger's syndrome ($P=2.0\times10^{-17}$) in particular, and for childhood autism ($P=1.5\times10^{-5}$), but not for the group of other/unspecified PDD ($P=0.36$) or for atypical autism ($P=0.13$) (Fig. 3). Excluding individuals with ID only marginally changed this result: atypical autism became nominally significant ($P=0.020$) (Supplementary Fig. 85). These results show that the genetic architecture underlying educational attainment is indeed shared with ASD but to a variable degree across the disorder spectrum. We found that the observed excess in ASD subjects of alleles positively associated with education attainment[43,44] was confined to Asperger's syndrome and childhood autism, and it was not seen here in atypical autism nor in other/unspecified PDD.

Finally, we evaluated the predictive ability of ASD PRSs by using five different sets of target and training samples within the combined iPSYCH-PGC sample. The observed mean variance explained by PRSs (Nagelkerke's R^2) was 2.45% ($P=5.58\times10^{-140}$) with a pooled PRS-based case–control odds ratio (OR) = 1.33 (95% CI 1.30 –1.36) (Supplementary Figs. 89 and 91). Dividing the target samples into PRS decile groups revealed an increase in ORs with increasing PRSs. The ORs for subjects with the highest PRSs increased to OR = 2.80 (95% CI 2.53–3.10) relative to the lowest decile (Fig. 4a and Supplementary Fig. 92). By leveraging correlated phenotypes in an attempt to improve prediction of ASD, we generated a multiphenotype PRS as a weighted sum of phenotype-specific PRSs (Methods). As expected, Nagelkerkes's R^2 increased for each PRS included, attaining its maximum at the full model at 3.77% ($P=2.03\times10^{-215}$) for the pooled analysis with an OR = 3.57 (95% CI 3.22–3.96) for the highest decile (Fig. 4b and Supplementary Figs. 93 and 94). These results demonstrate that an individual's ASD risk depends on the level of polygenic burden of thousands of common variants in a dose-dependent manner, which can be reinforced by adding SNP weights from ASD-correlated traits.

Functional annotation. To obtain information on the possible biological underpinnings of our GWAS results, we conducted several

Box 1 | Selected loci and candidates (ordered by chromosome)

Gene	Locus[a] and supporting evidence	Gene function
NEGR1	Chr 1:72729142 Shared ASD-MDD locus This locus is also significant in depression[25,57], educational attainment[21], intelligence[41], obesity, and BMI[29,58–61], and in an ASD-schizophrenia meta-analysis[5]. *NEGR1* is the only protein-coding gene in the locus. *NEGR1* is supported by brain Hi-C and eQTL analyses[25].	Neuronal growth regulator 1 (NEGR1) is an adhesion molecule modulating synapse formation in hippocampal neurons[62,63] and neurite outgrowth[64,65]. It is a member of the IgLON protein family, which is implicated in synaptic plasticity and axon extension[66–68] NEGR1 is predominantly expressed (and developmentally upregulated) in the hippocampus and cortex[69], as well as the hypothalamus[70]
PTBP2	Chr 1:96561801 ASD locus This locus is also significant in BMI[29,58,60] weight[58], and educational attainment[21]. In schizophrenia, the locus shows a P value of 6.5×10^{-6} (ref. [15]). *PTBP2* is the nearest protein-coding gene, ~625 kb from the index SNP. De novo and rare variants in *PTBP2* have been reported in ASD cases[1,3,71]. *PTBP2* is supported by Hi-C results in this study (Fig. 5d).	PTBP2, also known as nPTB (neuronal PTB) or brPTB (brain PTB), is a splicing regulator. PTBP1 and its paralog PTBP2 bind intronic polypyrimidine tracts in precursor mRNAs and target large sets of exons, thereby coordinating alternative-splicing programs during development[72]. Several switches in the expression of PTBP1 and PTBP2 regulate alternative splicing during neurogenesis and neuronal differentiation[73–76].
CADPS	Chr 3:62481063 Shared ASD-Edu locus This locus is also significant in a study of cognitive-decline rate[77]. *CADPS* is supported by Hi-C results in this study (Fig. 5a).	*CADPS* encodes a calcium-binding protein involved in exocytosis of neurotransmitters and neuropeptides. In line with *CAPDS* mRNA being mainly expressed in the brain and pituitary (GTEx portal; see URLs), immunoreactive CAPS-1 is localized in neural and various endocrine tissues[78]. In hippocampal synapses, CADPS regulates the pool of readily releasable vesicles at presynaptic terminals[79,80]
KCNN2	Chr 5:113801423 ASD locus (gene-wise analysis) This locus is also significant in educational attainment[21,81]. KCNN2 synaptic levels are regulated by the E3 ubiquitin ligase UBE3A[82], whose overexpression has been linked to ASD risk[82,83].	KCNN2 is a voltage-independent Ca^{2+}-activated K^+ channel that responds to changes in intracellular calcium concentration and couples calcium metabolism to potassium flux and membrane excitability. In central-nervous-system neurons, activation of KCNN2 modulates neuronal excitability by causing membrane hyperpolarization[84]. Hippocampal KCNN2 has roles in the formation of new memory[85], encoding and consolidation of contextual fear[86], and in drug-induced plasticity[87].
KMT2E	Chr 7:104744219 ASD locus This locus is also significant in schizophrenia[15,88] and in ASD-schizophrenia meta-analysis[5]. *KMT2E* de novo mutations are associated with ASD at FDR <0.1 (ref. [35]). A *KMT2E* credible SNP is a loss-of-function variant (Supplementary Table 16).	*KMT2E* encodes histone-lysine N-methyltransferase 2E and forms a family together with SETD5 (refs. [89,90]). Evidence suggests that recognition of the histone H3K4me3 mark by the KMT2E PHD finger can facilitate recruitment of KMT2E to transcription-active chromatin regions[91,92]. KMT2E has been implicated in chromatin regulation, control of cell-cycle progression, and maintaining genomic stability[93].
MACROD2	Chr 20:14836243 ASD locus This locus is significant in previous ASD GWAS[94] but not supported in larger study[95]. *MACROD2* is the only protein-coding gene in the locus.	MACROD2 is a nuclear enzyme that binds mono-ADP-ribosylated (MARylated) proteins and functions as an eraser of mono-ADP-ribosylation[96]. Intracellular MARylated histones and GSK3β are substrates of MACROD2, and the removal of MAR from GSK3β is responsible for reactivation of its kinase activity[96]. This gene is expressed in the lung and multiple regions of the brain but has low or no expression across most other tissues (GTEx portal; see URLs).

[a]Position of index SNP is listed. Chr, chromosome.

analyses. First, we examined how the ASD h^2_G partitioned on functional genomic categories as well as on cell-type-specific regulatory elements, by using stratified LD-score regression[45]. This analysis identified significant enrichment of heritability in conserved DNA regions and monomethyl histone H3 Lys4 (H3K4me1) histone marks[46], as well as in genes expressed in central-nervous-system cell types as a group (Supplementary Figs. 95 and 96), in line with observations in schizophrenia[15], major depression[25], and bipolar disorder[22]. Analyzing the enhancer-associated mark H3K4me1 in individual cells/tissues[46], we found significant enrichment in brain and neuronal cell lines (Supplementary Fig. 97). The highest enrichment was observed in the developing brain, germinal matrix, cortex-derived neurospheres, and embryonic-stem-cell-derived neurons, results consistent with ASD as a neurodevelopmental disorder with largely prenatal origins, as supported by data from analysis of rare de novo variants[31].

Common variation in ASD is located in regions that are highly enriched with regulatory elements predicted to be active in human corticogenesis (Supplementary Figs. 95–97). Because most gene regulatory events occur at a distance via chromosome looping, we leveraged Hi-C data from the germinal zone (GZ) and postmitotic-zone cortical plate (CP) in the developing fetal brain to identify potential target genes for these variants[47]. We performed fine-mapping of 28 loci to identify the set of credible variants with likely

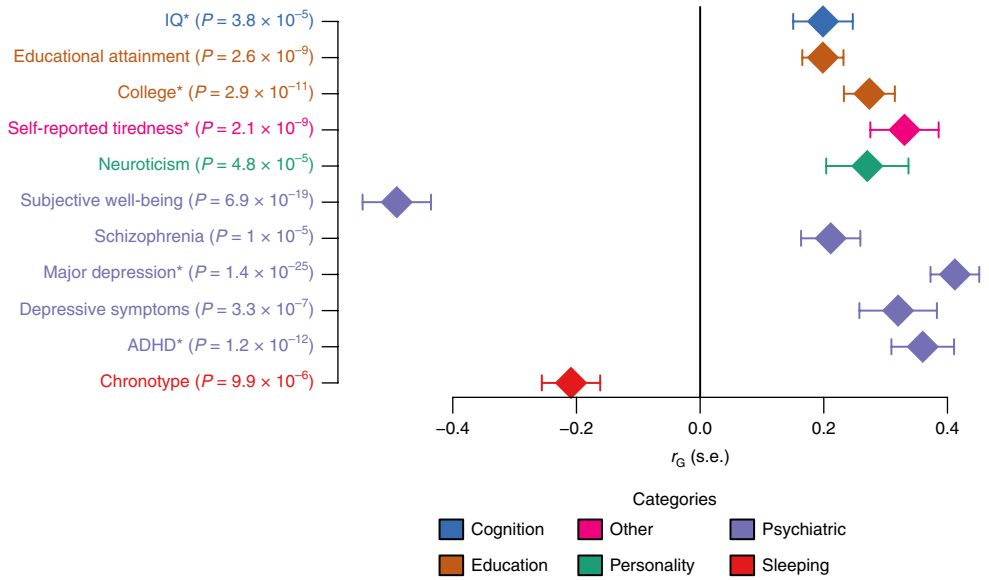

Fig. 2 | Genetic correlation with other traits. Significant genetic correlations between ASD ($n = 46,350$) and other traits after Bonferroni correction for testing a total of 234 traits available at LD Hub with the addition of several new phenotypes. Estimates and tests were performed with LDSC[19]. The results shown correspond to the following GWAS analyses: IQ[41] ($n = 78,308$), educational attainment[21] ($n = 328,917$), college[55] ($n = 111,114$), self-reported tiredness[56] ($n = 108,976$), neuroticism[23] ($n = 170,911$), subjective well-being[23] ($n = 298,420$), schizophrenia[15] ($n = 82,315$), major depression[25] ($n = 480,359$), depressive symptoms[23] ($n = 161,460$), attention deficit/hyperactivity disorder (ADHD)[26] ($n = 53,293$), and chronotype[42] ($n = 128,266$). Supplementary Table 5 shows the full output of this analysis. Asterisks indicate values are from in-house analyses of new summary statistics not yet included in LD Hub.

causal genetic risk[48] (Methods). Credible SNPs were significantly enriched in enhancer marks in the fetal brain (Supplementary Fig. 98), thus again confirming the likely regulatory role of these SNPs during brain development.

On the basis of location or evidence of physical contact from Hi-C, the 380 credible SNPs (28 loci) were assigned to 95 genes (40 protein coding), including 39 SNPs within promoters assigned to 9 genes, and 16 SNPs within the protein coding sequence of 8 genes (Supplementary Table 16 and Supplementary Fig. 98). Hi-C identified 86 genes, which interacted with credible SNPs in either the CP or GZ during brain development. Among these genes, 34 interacted with credible SNPs in both CP and GZ, thus representing a high-confidence gene list. Notable examples are illustrated in Fig. 5 and highlighted in Box 1. By analyzing their mean expression trajectory, we observed that the identified ASD-candidate genes (Supplementary Table 16) showed the highest expression during fetal corticogenesis, a finding in line with the enrichment of heritability in the regulatory elements in developing brain (Fig. 5e–g). Interestingly, both common and rare variation in ASD preferentially affects genes expressed during corticogenesis[31], thus highlighting a potential spatiotemporal convergence of genetic risk on this specific developmental epoch, despite the disorder's profound genetic heterogeneity.

Discussion

The high heritability of ASD has been recognized for decades and remains among the highest for any complex disease despite many clinical diagnostic changes over the past 30–40 years resulting in a broader phenotype that characterizes more than 1% of the population. Although early GWAS permitted estimates that common polygenic variation should explain a substantial fraction of the heritability of ASD, individually significant loci remained elusive. This lack of results was suspected to be due to limited sample size, because studies of schizophrenia—with similar prevalence and heritability, and lower fitness—and major depression achieved striking results only when sample sizes five to ten times larger than those available in ASD were used. This study has finally borne out that expectation with definitively demonstrated significant 'hits'.

Here we report what are, to our knowledge, the first reported common risk variants robustly associated with ASD, on the basis of unique Danish resources in conjunction with results of the earlier PGC data—more than tripling the previous largest discovery sample. Of these, five loci were defined in ASD alone, and seven additional suggested at a stricter threshold by using GWAS results from three correlated phenotypes (schizophrenia, depression, and educational attainment) and a recently introduced analytic approach, MTAG. Both genome-wide LD-score regression analysis and the finding that, even among the loci defined in ASD alone, additional evidence in these other trait scans indicated that the polygenic architecture of ASD is significantly shared with the risk of adult psychiatric illness and higher educational attainment and intelligence. Of note, the MTAG analyses were carried out as three pairwise analyses. Consequently, we avoided the complex interactions that might have arisen if we ran three or four correlated phenotypes at a time[9]. Indeed, despite the secondary summary statistics coming from large, high-powered studies, we obtained relatively modest weights of the contributions from these statistics, because the genetic correlations were modest. The largest weight was 0.27 for schizophrenia, followed by 0.24 for major depression, and 0.11 for educational attainment. Moreover, the estimated worst-case FDR was 0.021, just 0.001 higher than that of the ASD GWAS alone. Thus, all loci identified by MTAG were found with an acceptable degree of certainty and had substantial contributions from ASD alone (Table 1a,b and Supplementary Table 6). We expect that most or all such loci will probably be identified in future ASD-only GWAS as sample sizes are increased substantially; however, given how new these methods are, the precise phenotypic consequences of these particular variants await expansion of all these trait GWAS.

In most GWAS studies, there has been little evidence of heterogeneity of association across phenotypic subgroups. In this study, however, we observed strong heterogeneity of genetic overlap

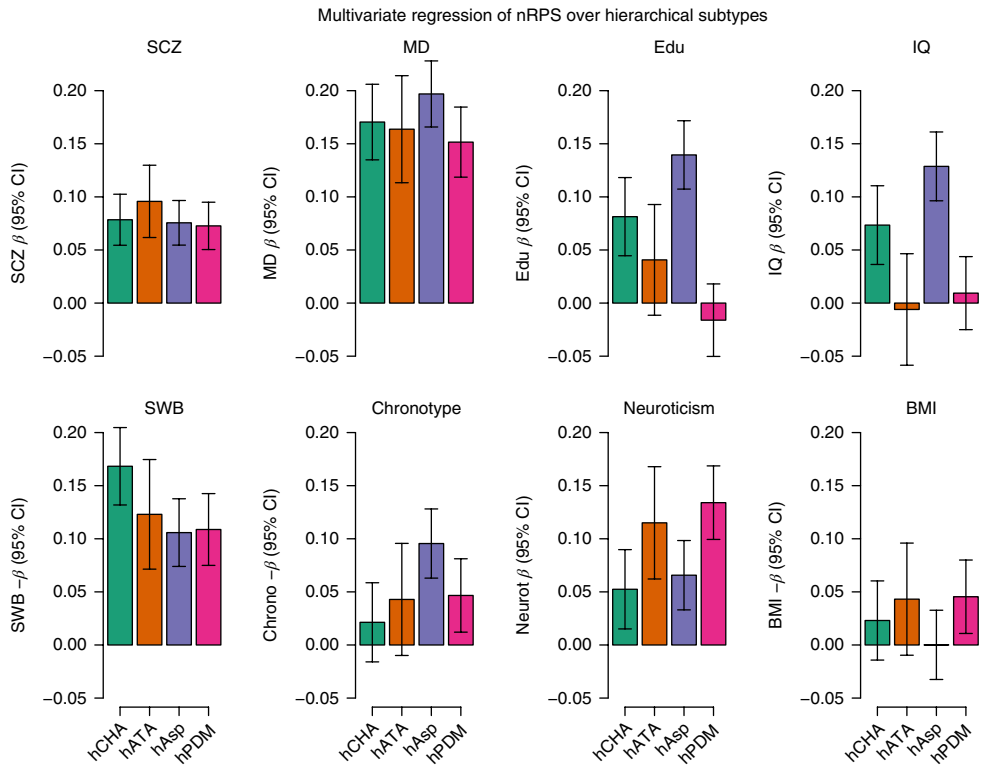

Fig. 3 | Profiling PRS load across distinct ASD subgroups. Results are shown for eight different phenotypes: schizophrenia (SCZ)[15], major depression (MD)[25], educational attainment (Edu)[21], human intelligence (IQ)[41], subjective well-being (SWB)[23], chronotype[42], neuroticism[23], and BMI[29]. The bars show coefficients from multivariate multivariable regression of the eight normalized scores on the distinct ASD subtypes of 13,076 cases and 22,664 controls, adjusting for batches and PCs. The subtypes are the hierarchically defined subtypes for childhood autism (hCHA, $n=3,310$), atypical autism (hATA, $n=1,494$), Asperger's syndrome (hAsp, $n=4,417$), and the lumped pervasive disorders developmental group (hPDM, $n=3,855$). The orientations of the scores for subjective well-being, chronotype and BMI have been switched to improve graphical presentation. The corresponding plot where subjects with ID have been excluded is shown in Supplementary Fig. 85, and with ID as a subtype in Supplementary Fig. 84. Applying the same procedure to the internally trained ASD score did not display systematic heterogeneity ($P=0.068$) except as expected for the ID groups ($P=0.00027$) (Supplementary Fig. 88). Linear hypotheses were tested with the Pillai test.

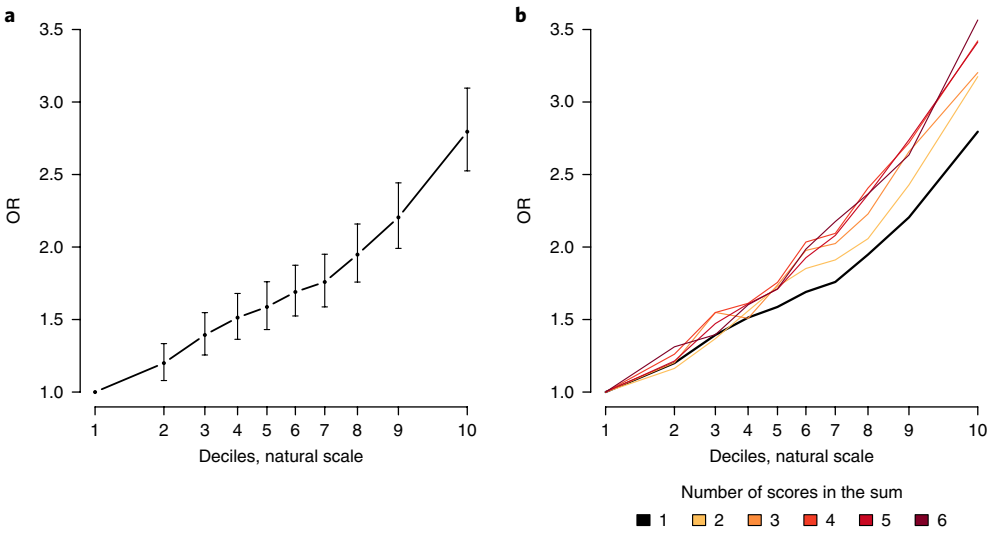

Fig. 4 | Decile plots (OR) by PRS within each decile for 13,076 cases and 22,664 controls. a, Decile plot with 95% CI for the internally trained ASD score (*P*-value threshold of 0.1). **b**, Decile plots on weighted sums of PRSs, starting with the ASD score of **a** and successively adding the scores for major depression[25], subjective well-being[23], schizophrenia[15], educational attainment[21], and chronotype[42]. In all instances, the *P*-value threshold for the score used was the one with the highest Nagelkerke's R^2. Supplementary Figs. 92 and 94 show the stability across leave-one-out groups that was used to create these combined results.

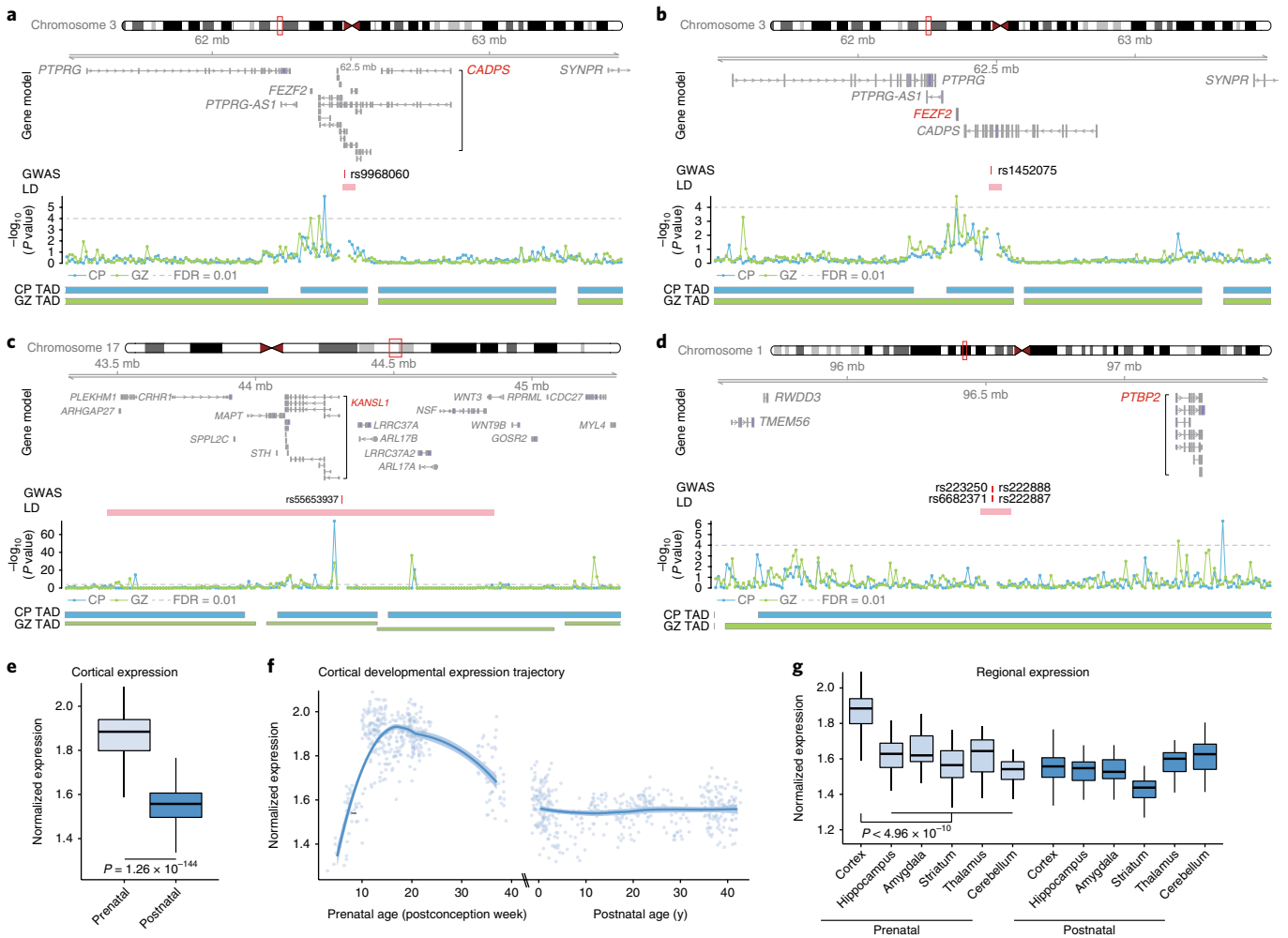

Fig. 5 | Chromatin interactions identify putative target genes of ASD loci. a–d. Chromatin-interaction maps of credible SNPs to the 1-Mb flanking region, providing putative candidate genes that physically interact with credible SNPs. The gene model is based on Gencode v19, and putative target genes are in red; the genomic coordinate for a credible SNP is labeled as GWAS; $-\log_{10}$(*P* value), the significance of the interaction between a SNP and each 10-kb bin, is shown with a gray dotted line for FDR = 0.01 (one-sided significance test, calculated as the probability of observing a higher contact frequency under the fitted Weibull distribution matched by chromosome and distance); topologically associated domain (TAD) borders in the CP and GZ are shown. **e–g**, Developmental expression trajectories of ASD candidate genes show the highest expression in prenatal periods. Significance was determined by *t* test ($n = 410$ and 453 for prenatal and postnatal samples, respectively). Box plots showing median, interquartile range (IQR) with whiskers adding IQR to the first and third quartiles (**e**,**g**) and LOESS smooth curves plotted with actual data points (**f**). In **g**, ASD candidate genes are highly expressed in the developing cortex compared with other brain regions. Data were analyzed with one-way analysis of variance and post hoc Tukey test, FDR corrected ($n = 410/453$, 39/36, 33/37, 48/34, 37/36, and 32/39 for prenatal/postnatal cortex, hippocampus, amygdala, striatum, thalamus, and cerebellum, respectively).

with other traits when our ASD samples were divided into distinct subsets. In particular, the excess of alleles associated with higher intelligence and educational attainment was observed only in the higher-functioning categories (particularly in individuals with Asperger's syndrome and individuals without comorbid ID) and not in the other/unspecified PDD and ID categories. These results are reminiscent of, and logically inverted relative to, the much greater role of spontaneous mutations in these latter categories, particularly in genes known to have an even larger effect in cohorts ascertained for ID/developmental delay[49]. Interestingly, other/unspecified PDDs and atypical autism also have significantly higher PRSs for neuroticism than childhood autism and Asperger's syndrome. The different enrichment profiles observed provide evidence of a heterogeneous and qualitatively different genetic architecture among subtypes of ASD, which should inform future studies aiming at identifying etiologies and disease mechanisms in ASD.

The strong differences in estimated SNP heritability between ASD cases with versus without ID, and the highest values observed in Asperger's syndrome, provide genetic evidence of longstanding observations. In particular, the results align well with the observation that de novo variants are more frequently observed in ASD cases with ID than in cases without comorbid ID, that IQ correlates positively with family history of psychiatric disorders[50]; and that severe ID (encompassing many syndromes that confer high risk of ASD) show far less heritability than that observed for mild ID[51], intelligence in general[52], and ASDs. Thus, it is perhaps unsurprising that our data suggest that the contribution of common variants may be more prominent in high-functioning ASD, such as Asperger's syndrome.

We further explored the functional implications of these results with complementary functional genomics data including Hi-C analyses of fetal brains and brain transcriptome data. Analyses at

genome-wide scale (partitioned h^2_G (Supplementary Figs. 95–97) and brain transcriptome enrichment (Fig. 5e–g)) as well as at single loci (Fig. 5a–d and Box 1) highlighted the involvement of processes relating to brain development and neuronal function. Notably, several genes located in the identified loci have previously been linked to ASD risk in studies of de novo and rare variants (Box 1 and Supplementary Table 4), including *PTBP2*, *CADPS*, and *KMT2E*, which were found to interact with credible SNPs in the Hi-C analysis (*PTBP2* and *CADPS*) or to contain a loss-of-function credible SNP (*KMT2E*). Interestingly, aberrant splicing of the sister gene of *CADPS*, *CADPS2*, which has almost identical function, has been found in autism cases, and *Cadps2*-knockout mice display behavioral anomalies with translational relevance to autism[53]. *PTBP2* encodes a neuronal splicing factor, and alterations in alternative splicing have been identified in brains from individuals diagnosed with ASD[54].

In summary, we established an initial robust set of common variant associations in ASD and have begun laying the groundwork through which the biology of ASD and related phenotypes will inevitably be better articulated.

URLs. GenomeDK high-performance-computing cluster in Denmark, https://genome.au.dk/; iPSYCH project, http://ipsych.au.dk/, iPSYCH download site, http://ipsych.au.dk/downloads/; NIMH Repository, https://www.nimhgenetics.org/available_data/autism/; PGC download site, https://www.med.unc.edu/pgc/results-and-downloads/; LISA cluster at SURFsara, https://userinfo.surfsara.nl/systems/lisa/; plink 1.9, http://www.cog-genomics.org/plink/1.9/; LDSC and associated files, https://github.com/bulik/ldsc/; LD Hub, http://ldsc.broadinstitute.org/ldhub/; GTEx portal, https://gtexportal.org/home/

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41588-019-0344-8.

Received: 24 November 2017; Accepted: 12 December 2018;
Published online: 25 February 2019

References

1. De Rubeis, S. et al. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209–215 (2014).
2. Gaugler, T. et al. Most genetic risk for autism resides with common variation. *Nat. Genet.* **46**, 881–885 (2014).
3. Iossifov, I. et al. The contribution of de novo coding mutations to autism spectrum disorder. *Nature* **515**, 216–221 (2014).
4. Krumm, N. et al. Excess of rare, inherited truncating mutations in autism. *Nat. Genet.* **47**, 582–588 (2015).
5. Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium. 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 (2017).
6. Ma, D. et al. A genome-wide association study of autism reveals a common novel risk locus at 5p14.1. *Ann. Hum. Genet.* **73**, 263–273 (2009).
7. Devlin, B., Melhem, N. & Roeder, K. Do common variants play a role in risk for autism? Evidence and theoretical musings. *Brain Res.* **1380**, 78–84 (2011).
8. Anney, R. et al. Individual common variants exert weak effects on the risk for autism spectrum disorders. *Hum. Mol. Genet.* **21**, 4781–4792 (2012).
9. Turley, P. et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat. Genet.* **50**, 229–237 (2018).
10. Pedersen, C. B. et al. The iPSYCH2012 case-cohort sample: new directions for unravelling genetic and environmental architectures of severe mental disorders. *Mol. Psychiatry* **23**, 6–14 (2018).
11. Lauritsen, M. B. et al. Validity of childhood autism in the Danish Psychiatric Central Register: findings from a cohort sample born 1990–1999. *J. Autism Dev. Disord.* **40**, 139–148 (2010).
12. Mors, O., Perto, G. P. & Mortensen, P. B. The Danish Psychiatric Central Research Register. *Scand. J. Public Health* **39** (Suppl.), 54–57 (2011).
13. Hollegaard, M. V. et al. Robustness of genome-wide scanning using archived dried blood spot samples as a DNA source. *BMC Genet.* **12**, 58 (2011).
14. Hollegaard, M. V. et al. Genome-wide scans using archived neonatal dried blood spot samples. *BMC Genomics* **10**, 297 (2009).
15. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
16. Cross-Disorder Group of the Psychiatric Genomics Consortium. et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* **45**, 984–994 (2013).
17. Gratten, J., Wray, N. R., Keller, M. C. & Visscher, P. M. Large-scale genomics unveils the genetic architecture of psychiatric disorders. *Nat. Neurosci.* **17**, 782–790 (2014).
18. Hansen, S. N., Overgaard, M., Andersen, P. K. & Parner, E. T. Estimating a population cumulative incidence under calendar time trends. *BMC Med. Res. Methodol.* **17**, 7 (2017).
19. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
20. Zheng, J. et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272–279 (2017).
21. Okbay, A. et al. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539–542 (2016).
22. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
23. Okbay, A. et al. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat. Genet.* **48**, 624–633 (2016).
24. Clarke, T.-K. et al. Common polygenic risk for autism spectrum disorder (ASD) is associated with cognitive ability in the general population. *Mol. Psychiatry* **21**, 419–425 (2016).
25. Wray, N. R. et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).
26. Demontis, D. et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet.* **51**, 63–75 (2019).
27. St Pourcain, B. et al. ASD and schizophrenia show distinct developmental profiles in common genetic overlap with population-based social communication difficulties. *Mol. Psychiatry* **23**, 263–270 (2018).
28. Wood, A. R. et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46**, 1173–1186 (2014).
29. Locke, A. E. et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
30. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
31. Parikshak, N. N. et al. Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* **155**, 1008–1021 (2013).
32. Samocha, K. E. et al. A framework for the interpretation of de novo mutation in human disease. *Nat. Genet.* **46**, 944–950 (2014).
33. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
34. Pardiñas, A. F. et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat. Genet.* **50**, 381–389 (2018).
35. Sanders, S. J. et al. Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. *Neuron* **87**, 1215–1233 (2015).
36. SPARK Consortium. SPARK: a US cohort of 50,000 families to accelerate autism research. *Neuron* **97**, 488–493 (2018).
37. Ashburner, M. et al. Gene ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
38. Gene Ontology Consortium. Gene Ontology Consortium: going forward. *Nucleic Acids Res.* **43**, D1049–D1056 (2015).
39. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
40. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
41. Sniekers, S. et al. Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nat. Genet.* **49**, 1107–1112 (2017).
42. Jones, S. E. et al. Genome-wide association analyses in 128,266 individuals identifies new morningness and sleep duration loci. *PLoS Genet.* **12**, e1006125 (2016).

43. Robinson, E. B. et al. Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. *Nat. Genet.* **48**, 552–555 (2016).
44. Weiner, D. J. et al. Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. *Nat. Genet.* **49**, 978–985 (2017).
45. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
46. Shlyueva, D., Stampfel, G. & Stark, A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet.* **15**, 272–286 (2014).
47. Won, H. et al. Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* **538**, 523–527 (2016).
48. Hormozdiari, F., Kostem, E., Kang, E. Y., Pasaniuc, B. & Eskin, E. Identifying causal variants at loci with multiple signals of association. *Genetics* **198**, 497–508 (2014).
49. Kosmicki, J. A. et al. Refining the role of de novo protein-truncating variants in neurodevelopmental disorders by using population reference samples. *Nat. Genet.* **49**, 504–510 (2017).
50. Robinson, E. B. et al. Autism spectrum disorder severity reflects the average contribution of de novo and familial influences. *Proc. Natl Acad. Sci. USA* **111**, 15161–15165 (2014).
51. Reichenberg, A. et al. Discontinuity in the genetic and environmental causes of the intellectual disability spectrum. *Proc. Natl Acad. Sci. USA* **113**, 1098–1103 (2016).
52. Polderman, T. J. C. et al. Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat. Genet.* **47**, 702–709 (2015).
53. Sadakata, T. et al. Autistic-like phenotypes in Cadps2-knockout mice and aberrant CADPS2 splicing in autistic patients. *J. Clin. Invest.* **117**, 931–943 (2007).
54. Parikshak, N. N. et al. Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. *Nature* **540**, 423–427 (2016).
55. Davies, G. et al. Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N=112 151). *Mol. Psychiatry* **21**, 758–767 (2016).
56. Deary, V. et al. Genetic contributions to self-reported tiredness. *Mol. Psychiatry* **23**, 609–620 (2017).
57. Hyde, C. L. et al. Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nat. Genet.* **48**, 1031–1036 (2016).
58. Thorleifsson, G. et al. Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat. Genet.* **41**, 18–24 (2009).
59. Willer, C. J. et al. Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat. Genet.* **41**, 25–34 (2009).
60. Speliotes, E. K. et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* **42**, 937–948 (2010).
61. Berndt, S. I. et al. Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. *Nat. Genet.* **45**, 501–512 (2013).
62. Hashimoto, T., Yamada, M., Maekawa, S., Nakashima, T. & Miyata, S. IgLON cell adhesion molecule Kilon is a crucial modulator for synapse number in hippocampal neurons. *Brain Res.* **1224**, 1–11 (2008).
63. Hashimoto, T., Maekawa, S. & Miyata, S. IgLON cell adhesion molecules regulate synaptogenesis in hippocampal neurons. *Cell Biochem. Funct.* **27**, 496–498 (2009).
64. Pischedda, F. et al. A cell surface biotinylation assay to reveal membrane-associated neuronal cues: Negr1 regulates dendritic arborization. *Mol. Cell. Proteomics* **13**, 733–748 (2014).
65. Pischedda, F. & Piccoli, G. The IgLON family member Negr1 promotes neuronal arborization acting as soluble factor via FGFR2. *Front. Mol. Neurosci.* **8**, 89 (2016).
66. Marg, A. et al. Neurotractin, a novel neurite outgrowth-promoting Ig-like protein that interacts with CEPU-1 and LAMP. *J. Cell Biol.* **145**, 865–876 (1999).
67. Funatsu, N. et al. Characterization of a novel rat brain glycosylphosphatidylinositol-anchored protein (Kilon), a member of the IgLON cell adhesion molecule family. *J. Biol. Chem.* **274**, 8224–8230 (1999).
68. Sanz, R., Ferraro, G. B. & Fournier, A. E. IgLON cell adhesion molecules are shed from the cell surface of cortical neurons to promote neuronal growth. *J. Biol. Chem.* **290**, 4330–4342 (2015).
69. Schäfer, M., Bräuer, A. U., Savaskan, N. E., Rathjen, F. G. & Brümmendorf, T. Neurotractin/kilon promotes neurite outgrowth and is expressed on reactive astrocytes after entorhinal cortex lesion. *Mol. Cell. Neurosci.* **29**, 580–590 (2005).
70. Lee, A. W. S. et al. Functional inactivation of the genome-wide association study obesity gene neuronal growth regulator 1 in mice causes a body mass phenotype. *PLoS One* **7**, e41537 (2012).
71. Doan, R. N. et al. Mutations in human accelerated regions disrupt cognition and social behavior. *Cell* **167**, 341–354.e12 (2016).
72. Vuong, J. K. et al. PTBP1 and PTBP2 serve both specific and redundant functions in neuronal pre-mRNA splicing. *Cell Rep.* **17**, 2766–2775 (2016).
73. Boutz, P. L. et al. A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons. *Genes Dev.* **21**, 1636–1652 (2007).
74. Makeyev, E. V., Zhang, J., Carrasco, M. A. & Maniatis, T. The microRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing. *Mol. Cell* **27**, 435–448 (2007).
75. Spellman, R., Llorian, M. & Smith, C. W. J. Crossregulation and functional redundancy between the splicing regulator PTB and its paralogs nPTB and ROD1. *Mol. Cell* **27**, 420–434 (2007).
76. Zheng, S. et al. Psd-95 is post-transcriptionally repressed during early neural development by PTBP1 and PTBP2. *Nat. Neurosci.* **15**, 381–388 (2012).
77. Li, Q. S., Parrado, A. R., Samtani, M. N. & Narayan, V. A. & Alzheimer's Disease Neuroimaging Initiative. Variations in the fra10ac1 fragile site and 15q21 are associated with cerebrospinal fluid aβ1–42 level. *PLoS One* **10**, e0134000 (2015).
78. Wassenberg, J. J. & Martin, T. F. J. Role of CAPS in dense-core vesicle exocytosis. *Ann. NY Acad. Sci.* **971**, 201–209 (2002).
79. Shinoda, Y. et al. CAPS1 stabilizes the state of readily releasable synaptic vesicles to fusion competence at CA3-CA1 synapses in adult hippocampus. *Sci. Rep.* **6**, 31540 (2016).
80. Farina, M. et al. Caps-1 promotes fusion competence of stationary dense-core vesicles in presynaptic terminals of mammalian neurons. *eLife* **4**, e05438 (2015).
81. Rietveld, C. A. et al. Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *Proc. Natl Acad. Sci. USA* **111**, 13790–13794 (2014).
82. Sun, J. et al. Ube3a regulates synaptic plasticity and learning and memory by controlling sk2 channel endocytosis. *Cell Rep.* **12**, 449–461 (2015).
83. Cook, E. H. Jr. et al. Autism or atypical autism in maternally but not paternally derived proximal 15q duplication. *Am. J. Hum. Genet.* **60**, 928–934 (1997).
84. Lin, M. T., Luján, R., Watanabe, M., Adelman, J. P. & Maylie, J. SK2 channel plasticity contributes to LTP at Schaffer collateral-CA1 synapses. *Nat. Neurosci.* **11**, 170–177 (2008).
85. Hammond, R. S. et al. Small-conductance Ca^{2+}-activated K^+ channel type 2 (SK2) modulates hippocampal learning, memory, and synaptic plasticity. *J. Neurosci.* **26**, 1844–1853 (2006).
86. Murthy, S. R. K. et al. Small-conductance Ca^{2+}-activated potassium type 2 channels regulate the formation of contextual fear memory. *PLoS One* **10**, e0127264 (2015).
87. Fakira, A. K., Portugal, G. S., Carusillo, B., Melyan, Z. & Morón, J. A. Increased small conductance calcium-activated potassium type 2 channel-mediated negative feedback on N-methyl-d-aspartate receptors impairs synaptic plasticity following context-dependent sensitization to morphine. *Biol. Psychiatry* **75**, 105–114 (2014).
88. Goes, F. S. et al. Genome-wide association study of schizophrenia in Ashkenazi Jews. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* **168**, 649–659 (2015).
89. Mas-Y-Mas, S. et al. The human mixed lineage leukemia 5 (mll5), a sequentially and structurally divergent set domain-containing protein with no intrinsic catalytic activity. *PLoS One* **11**, e0165139 (2016).
90. Sun, X.-J. et al. Genome-wide survey and developmental expression mapping of zebrafish SET domain-containing genes. *PLoS One* **3**, e1499 (2008).
91. Ali, M. et al. Molecular basis for chromatin binding and regulation of MLL5. *Proc. Natl Acad. Sci. USA* **110**, 11296–11301 (2013).
92. Lemak, A. et al. Solution NMR structure and histone binding of the PHD domain of human MLL5. *PLoS One* **8**, e77020 (2013).
93. Zhang, X., Novera, W., Zhang, Y. & Deng, L.-W. MLL5 (KMT2E): structure, function, and clinical relevance. *Cell. Mol. Life Sci.* **74**, 2333–2344 (2017).
94. Anney, R. et al. A genome-wide scan for common alleles affecting risk for autism. *Hum. Mol. Genet.* **19**, 4072–4082 (2010).
95. Torrico, B. et al. Lack of replication of previous autism spectrum disorder GWAS hits in European populations. *Autism Res.* **10**, 202–211 (2017).
96. Feijs, K. L. H., Forst, A. H., Verheugd, P. & Lüscher, B. Macrodomain-containing proteins: regulating new intracellular functions of mono(ADP-ribosyl)ation. *Nat. Rev. Mol. Cell Biol.* **14**, 443–451 (2013).

Acknowledgements

The iPSYCH project is funded by the Lundbeck Foundation (R102-A9118 and R155-2014-1724) and the universities and university hospitals of Aarhus and Copenhagen. Genotyping of iPSYCH and PGC samples was supported by grants from the Lundbeck Foundation, the Stanley Foundation, the Simons Foundation (SFARI 311789 to M.J.D.), and NIMH (5U01MH094432-02 to M.J.D.). The Danish National Biobank resource

was supported by the Novo Nordisk Foundation. Data handling and analysis on the GenomeDK HPC facility was supported by NIMH (1U01MH109514-01 to M.C.O.D and A.D.B.). High-performance computer capacity for handling and statistical analysis of iPSYCH data on the GenomeDK HPC facility was provided by the Centre for Integrative Sequencing, iSEQ, Aarhus University, Denmark (grant to A.D.B.). S.D.R. and J.D.B. were supported by NIH grants MH097849 (to J.D.B.) and MH111661 (to J.D.B.), and by the Seaver Foundation (to S.D.R. and J.D.B.). J. Martine was supported by the Wellcome Trust (grant 106047). O.A.A. received funding from the Research Council of Norway (213694, 223273, 248980, and 248778), Stiftelsen KG Jebsen, and South-East Norway Health Authority. We thank the research participants and employees of 23andMe for making this work possible.

Author contributions

Analysis: J.G., S.R., T.D.A., M.M., R.K.W., H.W., J.P., S.A., F.B., J.H.C., C.C., K.D., S.D.R., B.D., S.D., M.E.H., S.H., D.P.H., H.H., L.K., J. Maller, J. Martin, A.R.M., M. Nyegaard, T.N., D.S.P., T.P., B.S.P., P.Q., J.R., E.B.R., K. Roeder, P.R., S. Sandin, F.K.S., S. Steinberg, P.F.S., P.T., G.B.W., X.X., D.H.G., B.M.N., M.J.D., A.D.B. J.G., B.M.N., M.J.D., and A.D.B. supervised and coordinated the analyses. Sample and/or data provider and processing: J.G., S.R., M.M., R.K.W., E.A., O.A.A., R.A., R.B., J.D.B., J.B.-G., M.B.-H., F.C., K.C., D.D., A.L.D., J.I.G., C.S.H., M.V.H., C.M.H., J.L.M., A.P., C.B.P., M.G.P., J.B.P., K. Rehnström, A.R., E.S., G.D.S., H.S., C.R.S., Autism Spectrum Disorder Working Group of the Psychiatric Genomics Consortium, BUPGEN, Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, 23andMe Research Team, K.S., D.M.H., O.M., P.B.M., B.M.N., M.J.D., and A.D.B. Core PI group: K.S., D.H.G., M. Nordentoft, D.M.H., T.W., O.M., P.B.M., B.M.N., M.J.D., and A.D.B. Core writing group: J.G., M.J.D., and A.D.B. Direction of study: M.J.D. and A.D.B.

Competing interests

H.S., K.S., S. Steinberg, and G.B.W. are employees of deCODE genetics/Amgen. The 23andMe Research Team members are employed by 23andMe. D.H.G. is a scientific advisor for Ovid Therapeutic, Falcon Computing, and Axial Biotherapeutics. T.W. has acted as scientific advisor and lecturer for H. Lundbeck A/S.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-019-0344-8.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to M.J.D. or A.D.B.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

[1]The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Aarhus, Denmark. [2]Centre for Integrative Sequencing, iSEQ, Aarhus University, Aarhus, Denmark. [3]Department of Biomedicine–Human Genetics, Aarhus University, Aarhus, Denmark. [4]Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark. [5]Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. [6]Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA, USA. [7]Department of Psychiatry and Psychotherapy, Charité-Universitätsmedizin, Berlin, Germany. [8]Department of Psychiatry, Psychosomatics and Psychotherapy, University of Würzburg, Würzburg, Germany. [9]Centre for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden. [10]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [11]UNC Neuroscience Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [12]National Centre for Register-Based Research, Aarhus University, Aarhus, Denmark. [13]Centre for Integrated Register-based Research, Aarhus University, Aarhus, Denmark. [14]NORMENT-KG Jebsen Centre for Psychosis Research, University of Oslo, Oslo, Norway. [15]Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway. [16]MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, UK. [17]Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [18]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [19]Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [20]Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [21]Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark. [22]Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA. [23]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [24]Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA. [25]Department of Medical Genetics, Oslo University Hospital, Oslo, Norway. [26]Institute of Biological Psychiatry, MHC SctHans, Mental Health Services, Copenhagen, Denmark. [27]Department of Neurohabilitation, Oslo University Hospital, Oslo, Norway. [28]Genomics plc, Oxford, UK. [29]Vertex Pharmaceuticals, Abingdon, UK. [30]NevSom, Department of Rare Disorders and Disabilities, , Oslo University Hospital, Oslo, Norway. [31]Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland. [32]Language and Genetics Department, Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands. [33]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK. [34]Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, the Netherlands. [35]Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK. [36]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA. [37]Computational Biology Department, Carnegie Mellon University, Pittsburgh, PA, USA. [38]Department of Statistics and Data Science, Carnegie Mellon University, Pittsburgh, PA, USA. [39]Institute for Genomics and Multiscale Biology, Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [40]Friedman Brain Institute, Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [41]Mental Illness Research Education and Clinical Center (MIRECC), James J. Peters VA Medical Center, Bronx, NY, USA. [42]The State Diagnostic and Counselling Centre, Kópavogur, Iceland. [43]Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK. [44]deCODE genetics/Amgen, Reykjavík, Iceland. [45]Department of Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [46]Faculty of Medicine, University of Iceland, Reykjavik, Iceland. [47]A list of members and affiliations appears in the Supplementary Note. [48]A list of members and affiliations appears at the end of the paper. [49]Program in Neurogenetics, Department of Neurology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. [50]Center for Autism Research and Treatment and Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, Los Angeles, CA, USA. [51]Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. [52]Mental Health Services in the Capital Region of Denmark, Mental Health Center Copenhagen, University of Copenhagen, Copenhagen, Denmark. [53]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark. [54]Psychosis Research Unit, Aarhus University Hospital, Risskov, Denmark. *e-mail: mjdaly@atgu.mgh.harvard.edu; anders@biomed.au.dk

Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium

Naomi R. Wray[55,56], Maciej Trzaskowski[55], Enda M. Byrne[55], Abdel Abdellaoui[57], Mark J. Adams[58], Tracy M. Air[59], Till F. M. Andlauer[60,61], Silviu-Alin Bacanu[62], Aartjan T. F. Beekman[63], Tim B. Bigdeli[62,64], Elisabeth B. Binder[60,65], Douglas H. R. Blackwood[58], Julien Bryois[23], Henriette N. Buttenschøn[1,2,66], Na Cai[67,68], Enrique Castelao[69], Toni-Kim Clarke[58], Jonathan R. I. Coleman[70], Lucía Colodro-Conde[71], Baptiste Couvy-Duchesne[72,73], Nick Craddock[74], Gregory E. Crawford[75,76], Gail Davies[77], Ian J. Deary[77], Franziska Degenhardt[78,79], Eske M. Derks[71], Nese Direk[80,81], Conor V. Dolan[57], Erin C. Dunn[6,82,83], Thalia C. Eley[70], Valentina Escott-Price[84], Farnush Farhadi Hassan Kiadeh[85], Hilary K. Finucane[36,86], Andreas J. Forstner[78,79,87,88], Josef Frank[89], Héléna A. Gaspar[70], Michael Gill[90], Fernando S. Goes[91], Scott D. Gordon[71], Lynsey S. Hall[58,92], Thomas F. Hansen[93,94,95], Stefan Herms[78,79,88], Ian B. Hickie[96], Per Hoffmann[78,79,88], Georg Homuth[97], Carsten Horn[98], Jouke-Jan Hottenga[57], Marcus Ising[99], Rick Jansen[63,63], Eric Jorgenson[100], James A. Knowles[101], Isaac S. Kohane[102,103,104], Julia Kraft[105], Warren W. Kretzschmar[106], Jesper Krogh[107], Zoltán Kutalik[108,109], Yihan Li[106], Penelope A. Lind[71], Donald J. MacIntyre[110,111], Dean F. MacKinnon[91], Robert M. Maier[56], Wolfgang Maier[112], Jonathan Marchini[113], Hamdi Mbarek[57], Patrick McGrath[114], Peter McGuffin[70], Sarah E. Medland[71], Divya Mehta[56,115], Christel M. Middeldorp[57,116,117], Evelin Mihailov[118], Yuri Milaneschi[63,63], Lili Milani[118], Francis M. Mondimore[91], Grant W. Montgomery[55], Sara Mostafavi[119,120], Niamh Mullins[70], Matthias Nauck[121,122], Bernard Ng[120], Michel G. Nivard[57], Dale R. Nyholt[123], Paul F. O'Reilly[70], Hogni Oskarsson[124], Michael J. Owen[16], Jodie N. Painter[71], Roseann E. Peterson[62,125], Erik Pettersson[23], Wouter J. Peyrot[63], Giorgio Pistis[69], Danielle Posthuma[126,127], Jorge A. Quiroz[128], John P. Rice[129], Brien P. Riley[62], Margarita Rivera[70,130], Saira Saeed Mirza[80], Robert Schoevers[131], Eva C. Schulte[132,133], Ling Shen[100], Jianxin Shi[134], Stanley I. Shyn[135], Engilbert Sigurdsson[136], Grant C. B. Sinnamon[137], Johannes H. Smit[63], Daniel J. Smith[138], Fabian Streit[89], Jana Strohmaier[89], Katherine E. Tansey[139], Henning Teismann[140], Alexander Teumer[141], Wesley Thompson[1,14,15,94,142], Pippa A. Thomson[143], Thorgeir E. Thorgeirsson[144], Matthew Traylor[145], Jens Treutlein[89], Vassily Trubetskoy[105], André G. Uitterlinden[146], Daniel Umbricht[147], Sandra Van der Auwera[148], Albert M. van Hemert[149], Alexander Viktorin[23], Peter M. Visscher[55,56], Yunpeng Wang[1,14,15,94], Bradley T. Webb[125], Shantel Marie Weinsheimer[1,94], Jürgen Wellmann[140], Gonneke Willemsen[57], Stephanie H. Witt[89], Yang Wu[55], Hualin S. Xi[150], Jian Yang[56,151], Futao Zhang[55], Volker Arolt[152], Bernhard T. Baune[59], Klaus Berger[140], Dorret I. Boomsma[57], Sven Cichon[78,88,153,154], Udo Dannlowski[152], E. J. C. de Geus[57,155], J. Raymond DePaulo[91], Enrico Domenici[156], Katharina Domschke[157], Tõnu Esko[22,118], Hans J. Grabe[148], Steven P. Hamilton[158], Caroline Hayward[159], Andrew C. Heath[129], Kenneth S. Kendler[62], Stefan Kloiber[99,160,161], Glyn Lewis[162], Qingqin S. Li[163], Susanne Lucae[99], Pamela A. F. Madden[129], Patrik K. Magnusson[23], Nicholas G. Martin[71], Andrew M. McIntosh[58,77], Andres Metspalu[118,164], Bertram Müller-Myhsok[60,61,165], Markus M. Nöthen[78,79], Michael C. O'Donovan[16], Sara A. Paciga[166], Nancy L. Pedersen[23], Brenda W. J. H. Penninx[63], Roy H. Perlis[82,167], David J. Porteous[143], James B. Potash[168], Martin Preisig[69], Marcella Rietschel[89], Catherine Schaefer[100], Thomas G. Schulze[89,91,133,169,170], Jordan W. Smoller[6,82,83], Henning Tiemeier[80,171,172], Rudolf Uher[173], Henry Völzke[141], Myrna M. Weissman[114,174], Cathryn M. Lewis[70,175], Douglas F. Levinson[176] and Gerome Breen[70,177]

23andMe Research Team

Michelle Agee[178], Babak Alipanahi[178], Adam Auton[178], Robert K. Bell[178], Katarzyna Bryc[178], Sarah L. Elson[178], Pierre Fontanillas[178], Nicholas A. Furlotte[178], Bethann S. Hromatka[178], Karen E. Huber[178], Aaron Kleinman[178], Nadia K. Litterman[178], Matthew H. McIntyre[178], Joanna L. Mountain[178], Elizabeth S. Noblin[178], Carrie A. M. Northover[178], Steven J. Pitts[178], J. Fah Sathirapongsasuti[178], Olga V. Sazonova[178], Janie F. Shelton[178], Suyash Shringarpure[178], Joyce Y. Tung[178], Vladimir Vacic[178] and Catherine H. Wilson[178]

[55]Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland, Australia. [56]Queensland Brain Institute, University of Queensland, Brisbane, Queensland, Australia. [57]Department of Biological Psychology & EMGO+ Institute for Health and Care Research, Vrije Universiteit, Amsterdam, Amsterdam, the Netherlands. [58]Division of Psychiatry, University of Edinburgh, Edinburgh, UK. [59]Discipline of Psychiatry, University of Adelaide, Adelaide, South Australia, Australia. [60]Department of Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Munich, Germany. [61]Munich Cluster for Systems Neurology (SyNergy), Munich, Germany. [62]Department of Psychiatry, Virginia Commonwealth University, Richmond, VA, USA. [63]Department of Psychiatry, Vrije Universiteit Medical Center and GGZ inGeest, Amsterdam, the Netherlands. [64]Virginia Institute for Psychiatric and Behavior Genetics, Richmond, VA, USA. [65]Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, GA, USA. [66]Department of Clinical Medicine, Translational Neuropsychiatry Unit, Aarhus University, Aarhus, Denmark. [67]Human Genetics, Wellcome Trust Sanger Institute, Cambridge, UK. [68]Statistical genomics and systems genetics, European Bioinformatics Institute (EMBL-EBI), Cambridge, UK. [69]Department of Psychiatry, University Hospital of Lausanne, Prilly, Switzerland. [70]MRC Social Genetic and Developmental Psychiatry Centre, King's College London, London, UK. [71]Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. [72]Centre for Advanced Imaging, University of Queensland, Saint Lucia, Queensland, Australia. [73]Queensland Brain Institute, University of Queensland, Saint Lucia, Queensland, Australia. [74]Psychological Medicine, Cardiff University, Cardiff, UK. [75]Center for Genomic and Computational Biology, Duke University, Durham, NC, USA. [76]Department of Pediatrics, Division of Medical Genetics, Duke University, Durham, NC, USA. [77]Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK. [78]Institute of Human Genetics, University of Bonn, Bonn, Germany. [79]Life&Brain Center, Department of Genomics, University of Bonn, Bonn, Germany. [80]Epidemiology, Erasmus MC, Rotterdam, the Netherlands. [81]Psychiatry, Dokuz Eylul University School Of Medicine, Izmir, Turkey. [82]Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA. [83]Psychiatric and Neurodevelopmental Genetics Unit (PNGU), Massachusetts General Hospital, Boston, MA, USA. [84]Neuroscience and Mental Health, Cardiff University, Cardiff, UK. [85]Bioinformatics, University of British Columbia, Vancouver, British Columbia, Canada. [86]Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, USA. [87]Department of Psychiatry (UPK), University of Basel, Basel, Switzerland. [88]Human Genomics Research Group, Department of Biomedicine, University of Basel, Basel, Switzerland. [89]Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty, Mannheim, Heidelberg University, Mannheim, Baden-Württemberg, Germany. [90]Department of Psychiatry, Trinity College Dublin, Dublin, Ireland. [91]Department of Psychiatry and Behavioral Sciences, Johns Hopkins University, Baltimore, MD, USA. [92]Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, UK. [93]Danish Headache Centre, Department of Neurology, Rigshospitalet, Glostrup, Denmark. [94]Institute of Biological Psychiatry, Mental Health Center SctHans, Mental Health Services Capital Region of Denmark, Copenhagen, Denmark. [95]iPSYCH, Lundbeck Foundation Initiative for Psychiatric Research, Copenhagen, Denmark. [96]Brain and Mind Centre, University of Sydney, Sydney, New South Wales, Australia. [97]Interfaculty Institute for Genetics and Functional Genomics, Department of Functional Genomics, University Medicine and Ernst Moritz Arndt University Greifswald, Greifswald, Mecklenburg-Vorpommern, Germany. [98]Roche Pharmaceutical Research and Early Development, Pharmaceutical Sciences, Roche Innovation Center Basel, FHoffmann-La Roche Ltd, Basel, Switzerland. [99]Max Planck Institute of Psychiatry, Munich, Germany. [100]Division of Research, Kaiser Permanente Northern California, Oakland, CA, USA. [101]Psychiatry & The Behavioral Sciences, University of Southern California, Los Angeles, CA, USA. [102]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA. [103]Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA. [104]Informatics Program, Boston Children's Hospital, Boston, MA, USA. [105]Department of Psychiatry and Psychotherapy, Universitätsmedizin Berlin Campus Charité Mitte, Berlin, Germany. [106]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. [107]Department of Endocrinology at Herlev University Hospital, University of Copenhagen, Copenhagen, Denmark. [108]Institute of Social and Preventive Medicine (IUMSP), University Hospital of Lausanne, Lausanne, Switzerland. [109]Swiss Institute of Bioinformatics, Lausanne, Switzerland. [110]Division of Psychiatry, Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK. [111]Mental Health, NHS 24, Glasgow, UK. [112]Department of Psychiatry and Psychotherapy, University of Bonn, Bonn, Germany. [113]Statistics, University of Oxford, Oxford, UK. [114]Psychiatry, Columbia University College of Physicians and Surgeons, New York, NY, USA. [115]School of Psychology and Counseling, Queensland University of Technology, Brisbane, Queensland, Australia. [116]Child and Youth Mental Health Service, Children's Health Queensland Hospital and Health Service, South Brisbane, Queensland, Australia. [117]Child Health Research Centre, University of Queensland, Brisbane, Queensland, Australia. [118]Estonian Genome Center, University of Tartu, Tartu, Estonia. [119]Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada. [120]Statistics, University of British Columbia, Vancouver, British Columbia, Canada. [121]DZHK (German Centre for Cardiovascular Research), Partner Site Greifswald, University Medicine, University Medicine Greifswald, Greifswald, Germany. [122]Institute of Clinical Chemistry and Laboratory Medicine, University Medicine, Greifswald, Germany. [123]Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Queensland, Australia. [124]Humus, Reykjavik, Iceland. [125]Virginia Institute for Psychiatric & Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA. [126]Clinical Genetics, Vrije Universiteit Medical Center, Amsterdam, the Netherlands. [127]Complex Trait Genetics, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. [128]Solid Biosciences, Boston, MA, USA. [129]Department of Psychiatry, Washington University in Saint Louis School of Medicine, Saint Louis, MO, USA. [130]Department of Biochemistry and Molecular Biology II, Institute of Neurosciences, Center for Biomedical Research, University of Granada, Granada, Spain. [131]Department of Psychiatry, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. [132]Department of Psychiatry and Psychotherapy, Medical Center of the University of Munich, Campus Innenstadt, Munich, Germany. [133]Institute of Psychiatric Phenomics and Genomics (IPPG), Medical Center of the University of Munich, Campus Innenstadt, Munich, Germany. [134]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA. [135]Behavioral Health Services, Kaiser Permanente Washington, Seattle, WA, USA. [136]Faculty of Medicine, Department of Psychiatry, University of Iceland, Reykjavik, Iceland. [137]School of Medicine and Dentistry, James Cook University, Townsville, Queensland, Australia. [138]Institute of Health and Wellbeing, University of Glasgow, Glasgow, UK. [139]College of Biomedical and Life Sciences, Cardiff University, Cardiff, UK. [140]Institute of Epidemiology and Social Medicine, University of Münster, Münster, Germany. [141]Institute for Community Medicine, University Medicine, Greifswald, Germany. [142]Department of Psychiatry, University of

California, San Diego, San Diego, CA, USA. [143]Medical Genetics Section, CGEM, IGMM, University of Edinburgh, Edinburgh, UK. [144]deCODE Genetics/Amgen, Reykjavik, Iceland. [145]Clinical Neurosciences, University of Cambridge, Cambridge, UK. [146]Internal Medicine, Erasmus MC, Rotterdam, the Netherlands. [147]Roche Pharmaceutical Research and Early Development, Neuroscience, Ophthalmology and Rare Diseases Discovery & Translational Medicine Area, Roche Innovation Center Basel, FHoffmann-La Roche Ltd, Basel, Switzerland. [148]Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, Germany. [149]Department of Psychiatry, Leiden University Medical Center, Leiden, the Netherlands. [150]Computational Sciences Center of Emphasis, Pfizer Global Research and Development, Cambridge, MA, USA. [151]Institute for Molecular Bioscience, Queensland Brain Institute, University of Queensland, Brisbane, Queensland, Australia. [152]Department of Psychiatry, University of Münster, Münster, Germany. [153]Institute of Medical Genetics and Pathology, University Hospital Basel, University of Basel, Basel, Switzerland. [154]Institute of Neuroscience and Medicine (INM-1), Research Center Juelich, Juelich, Germany. [155]Amsterdam Public Health Institute, Vrije Universiteit Medical Center, Amsterdam, the Netherlands. [156]Centre for Integrative Biology, Università degli Studi di Trento, Trento, Trentino-Alto Adige, Italy. [157]Department of Psychiatry and Psychotherapy, Medical Center, Faculty of Medicine, University of Freiburg, Freiburg, Germany. [158]Psychiatry, Kaiser Permanente Northern California, San Francisco, CA, USA. [159]Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. [160]Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada. [161]Centre for Addiction and Mental Health, Toronto, Ontario, Canada. [162]Division of Psychiatry, University College London, London, UK. [163]Neuroscience Therapeutic Area, Janssen Research and Development, LLC, Titusville, NJ, USA. [164]Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia. [165]University of Liverpool, Liverpool, UK. [166]Human Genetics and Computational Biomedicine, Pfizer Global Research and Development, Groton, CT, USA. [167]Psychiatry, Harvard Medical School, Boston, MA, USA. [168]Psychiatry, University of Iowa, Iowa City, IA, USA. [169]Department of Psychiatry and Psychotherapy, University Medical Center Göttingen, Göttingen, Germany. [170]Human Genetics Branch, NIMH Division of Intramural Research Programs, Bethesda, MD, USA. [171]Child and Adolescent Psychiatry, Erasmus MC, Rotterdam, the Netherlands. [172]Psychiatry, Erasmus MC, Rotterdam, the Netherlands. [173]Psychiatry, Dalhousie University, Halifax, Nova Scotia, Canada. [174]Division of Epidemiology, New York State Psychiatric Institute, New York, NY, USA. [175]Department of Medical & Molecular Genetics, King's College London, London, UK. [176]Psychiatry & Behavioral Sciences, Stanford University, Stanford, CA, USA. [177]NIHR BRC for Mental Health, King's College London, London, UK. [178]23andMe, Inc., Mountain View, CA, USA.

Methods

Subjects. *iPSYCH sample.* The iPSYCH ASD sample is a part of a population based case–cohort sample extracted from a baseline cohort[10] consisting of all children born in Denmark between 1 May 1981 and 31 December 2005. Singletons who were born to a known mother and were resident in Denmark on their first birthday were included. Cases were identified from the Danish Psychiatric Central Research Register (DPCRR)[12], which includes data on all individuals treated in Denmark at psychiatric hospitals (from 1969 onward) as well as at outpatient psychiatric clinics (from 1995 onward). Subjects were diagnosed with ASD in 2013 or earlier by a psychiatrist according to ICD10, including diagnoses of childhood autism (ICD10 code F84.0), atypical autism (F84.1), Asperger's syndrome (F84.5), other pervasive developmental disorders (F84.8), and pervasive developmental disorder, unspecified (F84.9). For controls, we selected a random sample from the set of eligible children excluding those with an ASD diagnosis by 2013.

The samples were linked by using the unique national personal identification number to the Danish Newborn Screening Biobank (DNSB) at Statens Serum Institute (SSI), where DNA was extracted from Guthrie cards, and whole-genome amplification was performed in triplicate, as described previously[13,97]. Genotyping was performed at the Broad Institute of Harvard and MIT (Cambridge, MA, USA) with PsychChip arrays from Illumina according to the manufacturer's instructions. Genotype calling of markers with MAF >0.01 was performed by merging call sets from GenCall[98] and Birdseed[99], and less frequent variants were called with zCall[100]. Genotyping and data processing were carried out in 23 waves.

All analyses of the iPSYCH sample and joint analyses with the PGC samples were performed at the secured national GenomeDK high-performance computing cluster in Denmark. The study was approved by the Regional Scientific Ethics Committee in Denmark and the Danish Data Protection Agency.

PGC samples. In brief, five cohorts provided genotypes to the sample (n denotes the number of trios for which genotypes were available): the Geschwind Autism Center of Excellence (ACE; $n=391$), the Autism Genome Project[94] (AGP; $n=2{,}272$), the Autism Genetic Resource Exchange[101,102] (AGRE; $n=974$), the NIMH Repository, the Montreal[103]/Boston Collection (MONBOS; $n=1{,}396$, and the Simons Simplex Collection[104,105](SSC; $n=2{,}231$). The trios were analyzed as cases and pseudocontrols. A detailed description of the sample is available on the PGC website, and additional details are provided in Anney et al.[5]. Analyses of the PGC genotypes were conducted on the computer cluster LISA at the Dutch HPC center SURFsara.

Follow-up samples. As follow-up for the loci with P values $<10^{-6}$, we asked for look-up in five samples of Nordic and Eastern European origin, including 2,119 cases and 142,379 controls in total: BUPGEN (Norway: 164 cases and 656 controls), PAGES (Sweden: 926 cases and 3,841 controls not part of the PGC sample above), the Finnish autism case–control study (Finland: 159 cases and 526 controls), and deCODE (Iceland: 574 cases and 136,968 controls; Eastern Europe: 296 cases and 388 controls) (details in Supplementary Note).

Statistical analyses. All statistical tests were two sided unless otherwise stated. Software versions and additional information can be found in the Nature Research Reporting Summary.

GWAS analysis. Ricopili[15], the pipeline developed by the PGC Statistical Analysis Group was used for quality control, imputation, PCA, and primary association analysis (details in the Supplementary Note). The data were processed separately in the 23 genotyping batches in the case of iPSYCH and separately for each study in the PGC sample. Phasing was achieved with SHAPEIT[106], and imputation was done with IMPUTE2 (refs. [107,108]) with haplotypes from the 1000 Genomes Project, phase 3 (ref. [109]) as a reference.

After exclusion of regions of high LD[110], the genotypes were pruned down to a set of approximately 30,000 markers (details in Supplementary Note). With PLINK's[111] identity by state analysis, pairs of subjects were identified with $\hat{\pi} > 0.2$, and one subject of each such pair was excluded at random (with a preference for keeping cases). PCA was carried out with smartPCA[112,113]. In iPSYCH, a subsample of European ancestry was selected as an ellipsoid in the space of PC1–3 and centered and scaled by using the mean and eight s.d. of the subsample whose parents and grandparents were all known to have been born in Denmark ($n=31{,}500$). In the PGC sample, the European (CEU) subset was chosen by using a Euclidian-distance measure weighted by the variance explained by each of the first three PCs. Individuals more distant than ten s.d. from the combined CEU and Toscani in Italy (TSI) HapMap reference populations were excluded. We conducted a secondary PCA on the remaining 13,076 cases and 22,664 controls to provide covariates for the association analyses. Numbers of subjects in the data-generation flow for the iPSYCH sample can be found in Supplementary Table 1.

We performed association analyses by applying PLINK 1.9 to the imputed dosage data (the sum of imputation probabilities $P(\text{A1A2}) + 2P(\text{A1A1})$). In iPSYCH, we included the first four PCs as covariates as well as any PC beyond that, which were significantly associated with ASD in the sample, whereas the case–pseudocontrols from the PGC trios required no PC covariates. Combined results for iPSYCH and for iPSYCH with the PGC were achieved by meta-analysis of batchwise and studywise results by using METAL[114] (July 2010 version) with an inverse-variance-weighted fixed-effect model[115]. On chromosome X, males and females were analyzed separately and then meta-analyzed together. Subsequently, we applied a quality filter allowing only markers with an imputation info score 0.7, MAF of 0.01 and an effective sample size (Supplementary Note) of at least 70% of the study maximum. The degree to which the deviation in the test statistics could be ascribed to cryptic relatedness and population stratification rather than to polygenicity was measured from the intercept in LD-score regression[19] (LDSC) as the ratio of (intercept – 1) and (mean χ^2 – 1).

MTAG[9] was applied with standard settings. The iPSYCH-PGC meta-analysis summary statistics were paired with the summary statistics for each of major depression[25] (excluding the Danish samples but including summary statistics from 23andMe[57]; 111,902 cases, 312,113 controls, and mean $\chi^2=1.477$), schizophrenia[15] (also excluding the Danish samples; 34,129 cases, 45,512 controls, and mean $\chi^2=1.804$) and educational attainment[21] (328,917 samples and mean $\chi^2=1.648$). These are studies that have considerably more statistical power than the ASD scan, but because the genetic correlations are modest in the context of MTAG, the weights ascribed to the secondary phenotypes in the MTAG analyses remain relatively low (no higher than 0.27). The maximum FDR was estimated as recommended in the MTAG paper[9] (details in the Supplementary Note).

The results were clumped, and we highlighted loci of interest by selecting those that were significant at 5×10^{-8} in the iPSYCH-PGC meta-analysis or the meta-analysis with the follow-up sample or were significant at 1.67×10^{-8} in any of the three MTAG analyses. The composite GWAS consisting of the minimal P values at each marker over these five analyses was used as a background when creating Manhattan plots for the different analyses showing both what was maximally achieved and what the individual analysis contributed to that.

Gene-based association and gene-set analyses. MAGMA 1.06 (ref. [30]) was applied to the ASD GWAS summary statistics to test for gene-based association. By using NCBI 37.3 gene definitions and restricting the analysis to SNPs located within the transcribed region, we tested mean SNP association with the sum of –log(SNP P value) as the test statistic. The resulting gene-based P values were further used in competitive gene-set enrichment analyses in MAGMA. One analysis explored the candidate sets M13, M16, and M17 from Parikshak et al.[31], constrained, loss-of-function intolerant genes (pLI >0.9; refs. [32,33]) derived from data from the Exome Aggregation Consortium (details in Supplementary Note), as well as gene sets found in studies of rare variants in autism by Sanders et al.[35] and the curated gene list from the SPARK consortium[36]. Another was an agnostic analysis of the Gene Ontology sets[37,38] for molecular function from MsigDB 6.0 (ref. [39]). We analyzed only genes outside the broad MHC region (hg19: Chr 6: 25–35 Mb) and included only gene sets with 10–1,000 genes. The gene sets from Sanders et al. and SPARK included only one gene in MHC and were exempt from the MHC exclusion to be as true to the set as possible. All gene sets with significant enrichment were inspected to ensure that the signal was not driven by one or a few associated loci with multiple genes in close LD.

SNP heritability. SNP heritability, h^2_G, was estimated by using LDSC[19] for the full ASD GWAS sample and GCTA[40,116,117] for subsamples too small for LDSC. For LDSC, we used precomputed LD scores based on the European-ancestry samples of the 1000 Genomes Project[118] restricted to HapMap3 (ref. [119]) SNPs. The summary statistics with standard LDSC filtering were regressed onto these scores. For liability-scale estimates, we used a population prevalence for Denmark of 1.22% (ref. [18]). Lacking proper prevalence estimates for subtypes, we scaled the full spectrum prevalence on the basis of the composition of the case sample.

For subsamples too small for LDSC, the GREML approach of GCTA[40,116,117] was used. On best-guess genotypes (genotype probability >0.8, missing rate <0.01, and MAF >0.05) with indels removed, a genetic relatedness matrix was fitted for the association sample (i.e., the subjects of European ancestry with $\hat{\pi} \leq 0.2$), thus providing a relatedness estimate for all pairwise combinations of individuals. Estimation of the phenotypic variance explained by the SNPs (REML) was performed by including PC1–4 as continuous covariates together with any other PC that was nominally significantly associated with the phenotype as well as batches as categorical indicator covariates. Testing equal heritability for nonoverlapping groups was performed with permutation tests (with 1,000 permutations), keeping the controls and randomly assigning the different case labels.

Following Finucane et al.[45], we conducted an enrichment analysis of the heritability for SNPs for functional annotation and for SNPs located in cell-type-specific regulatory elements. Using first the same 24 overlapping functional annotations (stripped down from 53), as in Finucane et al., we regressed the χ^2 from the ASD GWAS summary statistics on the cell-type-specific LD scores downloaded from the site mentioned above with baseline scores, regression weights, and allele frequencies based on European-ancestry 1000 Genome Project data. The enrichment of a category was defined as the proportion of SNP heritability in the category divided by the proportion of SNPs in that category. Still following Finucane et al., we performed a similar analysis using 220 cell-type-specific annotations divided into ten overlapping groups. In addition, we conducted an analysis based on annotations derived from data on

H3K4me1 imputed gapped peak data from the Roadmap Epigenomics Mapping Consortium[120], more specifically information excluding the broad MHC region (Chr 6: 25–35 Mb).

Genetic correlation. For the main ASD samples, SNP correlations, r_G, were estimated by using LDSC[19], and for the analysis of ASD subtypes and subgroups in which the sample sizes were generally small, we used GCTA[40]. In both cases, we followed the same procedures as those explained above. For all but a few phenotypes, LDSC estimates of correlation were achieved by upload to LD Hub[20] for comparison to 234 phenotypes in total.

Polygenic risk scores. For the PRSs, we clumped the summary statistics, applying standard Ricopili parameters (details in the Supplementary Note). To avoid potential strand conflicts, we excluded all ambiguous markers for summary statistics not generated by Ricopili by using the same imputation reference. PRSs were generated at the default *P*-value thresholds (5×10^{-8}, 1×10^{-6}, 1×10^{-4}, 0.001, 0.01, 0.05, 0.1, 0.2, 0.5, and 1) as a weighted sum of the allele dosages in the ASD GWAS sample, summing over the markers abiding by the *P*-value threshold in the training set and weighing by the additive scale effect measure of the marker (log(OR) or β) as estimated in the training set. Scores were normalized before analysis.

We evaluated the predictive power by using Nagelkerke's R^2 and plots of ORs and CIs over score deciles. Both R^2 and ORs were estimated in regression analyses including the relevant PCs and indicator variables for genotyping waves.

Lacking a large ASD sample outside of iPSYCH and PGC, we trained a set of PRSs for ASD internally as follows. We divided the sample into five subsamples of approximately equal size, respecting the division, into batches. We then ran five GWAS, leaving out each group in turn from the training set, then performed meta-analysis of these with the PGC results. This procedure produced a set of PRSs for each of the five subsamples trained on their complement. Before analyses, each score was normalized to the group in which it was defined. We evaluated the predictive power in each group and on the whole sample combined.

To exploit the genetic overlap with other phenotypes to improve prediction, we created a series of new PRSs by adding to the internally trained ASD score the PRSs of other highly correlated phenotypes in a weighted sum (details in the Supplementary Note).

To analyze ASD subtypes in relation to PRSs, we defined a hierarchical set of phenotypes in the following way: The first hierarchical subtypes were childhood autism; hierarchical atypical autism was defined as all individuals with atypical autism and no childhood autism diagnosis, and hierarchical Asperger's syndrome was defined as all individuals with an Asperger's syndrome diagnosis and neither childhood autism nor atypical autism. Finally, we lumped other pervasive developmental disorders and pervasive developmental disorder, unspecified into pervasive disorders developmental mixed, and the hierarchical version consisted of all subjects with such a diagnosis and none of the preceding diagnoses (Supplementary Table 13). We examined the distribution over the distinct ASD subtypes of PRSs for a number of phenotypes showing high r_G with ASD (as well as a few with low r_G as negative controls), by performing multivariate regression of the scores on the subtypes while adjusting for relevant PCs and wave-indicator variables in a linear regression (details in the Supplementary Note).

Hi-C analysis. The Hi-C data were generated from two major cortical laminae: the GZ, containing primarily mitotically active neural progenitors, and the cortical and subcortical plate, consisting primarily of postmitotic neurons[47]. We first derived a set of credible SNPs (putative causal SNPs) from the identified top-ranking loci in the ASD GWAS by using CAVIAR[48]. The 30 loci showing the strongest association were intersected with the Hi-C reference data, thus resulting in 28 loci for analysis. To test whether credible SNPs were enriched in active marks in the fetal brain[120], we used GREAT, as previously described[47,121]. Credible SNPs were subgrouped into SNPs without known function (unannotated) and functionally annotated SNPs (SNPs in the gene promoters and SNPs causing nonsynonymous variants) (Supplementary Fig. 98). Then we integrated unannotated credible SNPs with chromatin-contact profiles during fetal corticogenesis[47], defining genes physically interacting with intergenic or intronic SNPs (Supplementary Fig. 98).

The spatiotemporal transcriptomic atlas of the human brain was obtained from Kang et al.[122]. We used transcriptomic profiles of multiple brain regions with developmental epochs spanning prenatal (6–37 weeks postconception) and postnatal (4 months to 42 years) periods. Expression values were log-transformed and centered to the mean expression level for each sample by using a scale(center = *T*, scale = *F*)+1 function in R. ASD candidate genes identified by Hi-C analyses (Supplementary Fig. 98) were selected for each sample, and their average centered expression values were calculated and plotted.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary.

Data availability

The summary statistics are available for download the iPSYCH and at the PGC download sites (see URLs). For access to genotype data from the PGC samples and the iPSYCH sample, researchers should contact the lead principal investigators M.J.D. and A.D.B. for PGC-ASD and iPSYCH-ASD, respectively.

References

97. Børglum, A. D. et al. Genome-wide study of association and interaction with maternal cytomegalovirus infection suggests new schizophrenia loci. *Mol. Psychiatry* **19**, 325–333 (2014).
98. Illumina, Inc. *Illumina Gencall Data Analysis Software*. (Illumina, Inc., San Diego, 2005).
99. Korn, J. M. et al. Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat. Genet.* **40**, 1253–1260 (2008).
100. Goldstein, J. I. et al. zCall: a rare variant caller for array-based genotyping: genetics and population analysis. *Bioinformatics* **28**, 2543–2545 (2012).
101. Lajonchere, C. M., AGRE Consortium. Changing the landscape of autism research: the autism genetic resource exchange. *Neuron* **68**, 187–191 (2010).
102. Geschwind, D. H. et al. The autism genetic resource exchange: a resource for the study of autism and related neuropsychiatric conditions. *Am. J. Hum. Genet.* **69**, 463–466 (2001).
103. Gauthier, J. et al. Autism spectrum disorders associated with X chromosome markers in French-Canadian males. *Mol. Psychiatry* **11**, 206–213 (2006).
104. Fischbach, G. D. & Lord, C. The Simons Simplex Collection: a resource for identification of autism genetic risk factors. *Neuron* **68**, 192–195 (2010).
105. Chaste, P. et al. A genome-wide association study of autism using the Simons Simplex Collection: does reducing phenotypic heterogeneity in autism increase genetic homogeneity? *Biol. Psychiatry* **77**, 775–784 (2015).
106. Delaneau, O., Marchini, J. & Zagury, J.-F. A linear complexity phasing method for thousands of genomes. *Nat. Methods* **9**, 179–181 (2011).
107. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* **44**, 955–959 (2012).
108. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5**, e1000529 (2009).
109. 1000 Genomes Project Consortium. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
110. Price, A. L. et al. Long-range LD can confound genome scans in admixed populations. *Am. J. Hum. Genet.* **83**, 132–135 (2008). author reply 135–139.
111. Chang, C. C. et al. Second-generation plink: rising to the challenge of larger and richer datasets. *GigaScience* **4**, 7 (2015).
112. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
113. Price, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
114. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
115. Begum, F., Ghosh, D., Tseng, G. C. & Feingold, E. Comprehensive literature review and statistical considerations for GWAS meta-analysis. *Nucleic Acids Res.* **40**, 3777–3784 (2012).
116. Lee, S. H., Wray, N. R., Goddard, M. E. & Visscher, P. M. Estimating missing heritability for disease from genome-wide association studies. *Am. J. Hum. Genet.* **88**, 294–305 (2011).
117. Yang, J. et al. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42**, 565–569 (2010).
118. 1000 Genomes Project Consortium. et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
119. Altshuler, D. M. et al. Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010).
120. Kundaje, A. et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
121. McLean, C. Y. et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).
122. Kang, H. J. et al. Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483–489 (2011).

nature research

Corresponding author(s): Anders D. Børglum

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	This is fully described in the Online Methods and associated Supplemental Note. In brief: GenCall, Birdseed and zCall were used for genotype calling and the genotypes merged.
Data analysis	This is fully described in the Online Methods and associated Supplemental Note. In brief: Genotype calling was done using GenCall (1.6.2.2), Birdseed (1.6) and zCall version 1 (Autocall, https://github.com/jigold/zCall). Quality control, imputation, association analyses, and polygenic risk scoring was done using the Ricopili pipeline: https://github.com/Nealelab/ricopili, which relies on the following software: SHAPEIT v2, IMPUTE2, Eigensoft 6.0.1 (incl. smartPCA), Plink 1.9, METAL 2011-03-25. For gene-based and gene-set analyses we used MAGMA 1.06. Estimation of credible SNPs were done using CAVIAR v1, and to test whether credible SNPs are enriched in active marks in the fetal brain and in-house implementation of the GREAT method was used (see Won et al. Nature. 2016;538(7626):523-527). Functional annotation of credible SNPs was done using Ensemble Variant Effect Predictor (VEP). SNP heritability, partitioning of the heritability and genetic correlations were estimated using LD score regression (https://github.com/bulik/ldsc) and LD hub (http://ldsc.broadinstitute.org/) for the large samples. Genetic correlation between ASD subtypes was estimated using GCTA v1.26.0. Multitrait association analyzes were conducted using MTAG (https://github.com/omeed-maghzian/mtag). R v3.4 was used in general for statistical analyzes and plotting (https://www.Rproject.org).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

As stated in the manuscript, summary statistics has already been made available at http://ipsych.au.dk/downloads/ and https://www.med.unc.edu/pgc/results-and-downloads. For access to genotypes from the PGC samples and the iPSYCH sample, researchers should contact the lead PIs Mark J. Daly and Anders D. Børglum for PGC-ASD and iPSYCH-ASD respectively.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculation was made. Previous studies of psychiatric disorders that are very polygenic (e.g. schizophrenia) have demonstrated that high numbers of cases and controls (in line with the sample size analyzed in this study) yield enough power to detect common risk variants with low effect sizes.
Data exclusions	Within each analyzed cohort we aimed at analyzing genetically homogeneous samples. Genetic outliers were excluded based on principal component analyses and related individuals were removed.
Replication	Consistency was checked internally between the 23 batches in iPSYCH and between iPSYCH and PGC. The 88 strongest signals were followed up in independent samples from other Nordic countries and Eastern Europe. We evaluated our results in three ways: sign test, genetic correlation, and metaanalysis of the combined samples, and these generally support the results from our primary GWAS. In addition support for the reported loci were obtained by MTAG analysis of correlated traits.
Randomization	It is an observational study comparing everybody in the selected birth cohorts with and ASD diagnosis as cases, and a random sample from the complement in said cohort as controls.
Blinding	In iPSYCH, diagnoses are drawn from registries. These are administrative data bases populated by data from the clinicians long before the current study. The blood samples are pulled from a biobank. Hence, the study participants and diagnosing clinicians are blinded with respect to this study. Genotyping is done on a massive scale on 85.000 individuals on 500.000 variables (which by imputation is expanded to ~10 million variables), and the data is generated without a specif goal or effect in mind except for an overall goal of investigating the genetic and environmental effects on psychiatric disorders. So although it is in principle possible for analysts in the lab to look up crude diagnostic data for a sample, it will not change the genotyping. - In the meta analysis we include data from the Psychiatric Genetics Consortium (PGC) which has been reported in an earlier publication. There design was different, but analyses analogous.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	In the meta-analysis we included samples from the Psychiatric Genetics Consortium (PGC) and the iPSYCH sample. The iPSYCH sample was processed in 23 batches (genotyping, qc and imputation was done separately for theses batches) of approximately 3,500 individuals each. All analyzes were adjusted for batch, and principal components included to control for population stratification. The PGC samples are trio samples, hence there was no need to adjust for populations stratification there.
Recruitment	In iPSYCH, diagnoses are drawn from national registries and the blood samples are pulled from a the Danish Neonatal Screening Biobank. Hence, it is a population sample and bias from self-selection is impossible.