Open access • Posted Content • DOI:10.1101/224774

Common risk variants identified in autism spectrum disorder — Source link

Jakob Grove, Stephan Ripke, Thomas Damm Als, Manuel Mattheisen ...+73 more authors

Institutions: Aarhus University, Harvard University, University of California, Los Angeles, University of Oslo ...+14 more institutions

Published on: 25 Nov 2017 - bioRxiv (Cold Spring Harbor Laboratory)

Topics: Autism spectrum disorder and Genome-wide association study

Common risk variants identified in autism spectrum disorder

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

* Co-last, co-corresponding authors.

Correspondence to: ADB (anders@biomed.au.dk) and MJD (mjdaly@atgu.mgh.harvard.edu)

1. The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, 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, Massachusetts, USA
6. Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA
7. Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA
8. Department of Psychiatry, Charite Universitatsmedizin Berlin Campus Benjamin Franklin, Berlin, Germany
9. Program in Neurogenetics, Department of Neurology, David Geffen School of Medicine, University of California, Los Angeles, USA
10. Center for Autism Research and Treatment and Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, USA
11. National Centre for Register-Based Research, Aarhus University, Aarhus, Denmark
12. Centre for Integrated Register-based Research, Aarhus University, Aarhus, Denmark
13. NORMENT - KG Jebsen Centre for Psychosis Research, University of Oslo, Oslo, Norway

14. Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway
15. MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, UK
16. Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, USA
17. Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
18. Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.
19. Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark
20. Department of Medical Epidemiology and Biostatistics, Karolinska, Sweden
21. Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA 15213, USA.
22. Department of Medical Genetics, Oslo University Hospital, Oslo, Norway
23. Institute of Biological Psychiatry, MHC Sct. Hans, Mental Health Services Copenhagen, Denmark
24. Department of Neurohabilitation, Oslo University Hospital, Oslo; Norway
25. Genomics plc, Oxford, UK
26. Department of Rare Disorders and Disabilities, Oslo University Hospital, Norway
27. Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland.
28. Language and Genetics Department, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
29. MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK;
30. Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands
31. Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK
32. Department of Epidemiology, Harvard Chan School of Public Health, Boston, Massachusetts, USA
33. Computational Biology Department, Carnegie Mellon University, Pittsburgh, PA 15213, USA
34. Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213, USA.
35. Institute for Genomics and Multiscale Biology, Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA
36. Mental Illness Research Education and Clinical Center (MIRECC), James J. Peters VA Medical Center, Bronx, New York, USA
37. The State Diagnostic and Counselling Centre, Digranesvegur 5, IS-200 Kópavogur, Iceland
38. Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK
39. deCODE genetics/Amgen, Sturlugata 8, IS-101 Reykjavík Iceland
40. Departments of Genetics and Psychiatry, University of North Carolina, Chapel Hill, NC, USA
41. Faculty of Medicine, University of Iceland, Reykjavik, Iceland
42. Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, California, USA
43. Mental Health Services in the Capital Region of Denmark, Mental Health Center Copenhagen, University of Copenhagen, Copenhagen, Denmark
44. Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark
45. Psychosis Research Unit, Aarhus University Hospital, Risskov, Denmark

Abstract

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 ASD cases and 27,969 controls that identifies five genome-wide significant loci. Leveraging GWAS results from three phenotypes with significantly overlapping genetic architectures (schizophrenia, major depression, and educational attainment), seven additional loci shared with other traits are identified at equally strict significance levels. Dissecting the polygenic architecture we find both quantitative and qualitative polygenic heterogeneity across ASD subtypes, in contrast to what is typically seen in other complex disorders. 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, just as it has been in a broad range of important psychiatric and diverse medical phenotypes.

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 very heterogeneous including individuals with severe impairment and intellectual disability as well as individuals with above average IQ and high levels of academic and occupational functioning. ASD affects 1-1.5% of individuals and is highly heritable, with both common and rare variants contributing 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. By 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. While a number of genes have been convincingly implicated via excess statistical aggregation of *de novo* mutations, the largest GWAS to date (n=7387 cases scanned) – while providing compelling evidence for the bulk contribution of common variants – did not conclusively identify single variants at genome-wide significance[5-7]. This underscored 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 the first common risk variants robustly associated with ASD by more than doubling the discovery sample size compared to previous GWAS[5-8]. We describe strong genetic correlations between ASDs and other complex disorders and traits, confirming shared etiology, and we show results indicating differences in the polygenic architecture across clinical sub-types of ASD. Leveraging these relationships and recently introduced computational techniques[9], we identify additional novel 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 (http://ipsych.au.dk)[10], we collected and genotyped a Danish nation-wide population-based case-cohort sample including nearly all individuals born in Denmark between 1981 and 2005 and diagnosed with ASD (according to ICD-10) before

2014. We randomly selected controls from the same birth cohorts (**Table S1.1.1**). We have previously validated registry-based ASD diagnoses[11,12] and demonstrated the accuracy of genotyping DNA extracted and amplified from bloodspots collected shortly after birth[13,14]. Genotypes were processed using Ricopili[15], performing stringent quality control of data, removal of related individuals, exclusion of ancestry outliers based on principal component analysis, and imputation[16,17] 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 analysis. As is now standard in human complex trait genomics, our primary analysis is 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 pseudo controls[18]. All PGC samples had been processed using the same Ricopili pipeline for QC, imputation and analysis as employed 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 (SE = 0.106; P = 1.75 x 10^{-13}), similar to the genetic correlations observed between datasets in other mental disorders[19]. Likewise, polygenicity as assessed by polygenic risk scores (PRS)[20] showed consistency across the samples supporting homogeneity of effects across samples and study designs (**Figure S4.4.4**). SNP heritability (h^2_G) was estimated to be 0.118 (SE = 0.010, for population prevalence of 0.012[21]).

The main GWAS meta-analysis totaled 18,381 ASD cases and 27,969 controls, and applied an inverse variance-weighted fixed effects model[22]. To ensure that the analysis was well-powered and robust, we examined markers with minor allele frequency (MAF) ≥ 0.01, imputation INFO score ≥ 0.7, and 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 (**Figure 1**; **Table 1a**; **Figures S4.2.1-4.2.44**). Each locus was strongly supported by both the Danish case-control and the PGC family-based data. While modest inflation was observed (lambda=1.12, lambda1000 = 1.006), LD score regression analysis[23] indicates this is arising from polygenicity (> 96%, see Methods) rather than confounding. The strongest signal among 265,846 markers analyzed on chromosome X was $P = 5.4 \times 10^{-6}$.

We next obtained replication data for the top 88 loci with p-values < 1×10^{-5} in five cohorts of European ancestry totaling 2,119 additional cases and 142,379 controls (**Table S1.3.2**). An overall replication of direction of effects was observed (53 of 88 (60%) of P < 1×10^{-5}; 16 of 23 (70%) at P < 1×10^{-6}; both sign tests P < 0.05) and two additional loci achieved genome-wide significance in the combined analysis (**Table 1a**). More details on the identified loci can be found in **Table S3.1.3** and selected candidates are described in **Box1**.

Correlation with other traits and multi-trait 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[24,25] using bivariate LD score regression (**Figure 2**, **Table S3.3.2**). 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[26] (r_G = 0.199, $p = 2.56 \times 10^{-9}$), indicating a substantial genetic overlap with these phenotypes and corroborating previous reports[5,24,27,28]. In contrast to previous reports[18], we found a strong and highly significant correlation with major depression[29] (r_G = 0.412, $p = 1.40 \times 10^{-25}$), and we are the first to report a prominent overlap

with ADHD[30] (r_G = 0.360, p = 1.24 x 10^{-12}). Moreover, we confirm the genetic correlation with social communication difficulties at age 8 in a non-ASD population sample reported previously based on a subset of the ASD sample[31] (r_G = 0.375, p = 0.0028).

In order to leverage these observations for the discovery of ASD loci shared with 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)[29] and educational attainment (N = 328,917)[26]. We utilize the recently introduced MTAG method[9] which, briefly, generalizes the standard inverse-variance weighted meta-analysis for multiple phenotypes. In this case, MTAG takes advantage of the fact that, 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 utilize a more conservative threshold of 1.67 x 10^{-8} for declaring significance across these secondary scans. In addition to stronger evidence for several of the ASD hits defined above, variants in seven additional regions achieve genome-wide significance, including three loci shared with educational attainment and four shared with major depression (**Table 1b, Box 1**).

Gene and gene-set analysis

Next, we performed gene-based association analysis on our primary ASD meta-analysis using MAGMA[32], 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 (**Table S3.1.2**). As expected, the majority of these genes were located within the genome-wide significant loci identified in the GWAS, but seven genes are located

in four additional loci including *KCNN2*, *MMP12, NTM* and a cluster of genes on chromosome 17 (*KANSLl*, *WNT3*, *MAPT* and *CRHRl)* (**Figures S4.2.46-60**). In particular, *KCNN2* was strongly associated ($P = 1.02x10^{-9}$), far beyond even single-variant statistical thresholds and is included in the descriptions in **Box 1**.

Enrichment analyses using gene co-expression modules from human neocortex transcriptomic data (M13, M16 and M17 from Parikshak et al. 2013[33] and loss-of-function intolerant genes (pLI > 0.9)[34,35], which previously have shown evidence of enrichment in neurodevelopmental disorders[30,33,36], yielded only nominal significance for the latter and M16 (**Table S3.1.4**). Likewise analysis of Gene Ontology sets[37,38] for molecular function from MsigDB[39] revealed no significant sets after Bonferroni correction for multiple testing (**Table S3.1.5**).

Dissection of the polygenic architecture

As ASD is a highly heterogeneous disorder, we explored how h_G^2 partitioned across phenotypic sub-categories in the iPSYCH sample and estimated the genetic correlations between these groups using GCTA[40-42]. We examined cases with and without intellectual disability (ID, N = 1,873) and the ICD-10 diagnostic sub-categories: 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 (PDD, F84.8-9, N = 5,795), reducing to non-overlapping groups when doing pairwise comparisons (see **Table S2.3.1**). While the pairwise genetic correlations were consistently high between all sub-groups (95% CIs including 1 in all comparisons), the h_G^2 of Asperger's syndrome (h_G^2 = 0.097, SE = 0.001) was found to be twice the h_G^2 of both childhood autism (h_G^2 = 0.049, SE = 0.009, P = 0.001) and the group of other/unspecified PDD (h_G^2 = 0.045, SE = 0.008, P = 0.001) (**Table**

S3.2.1 and **S3.3.1, Figure S4.3.1** and **S4.4.1**). Similarly, the h_G^2 of ASD without ID (h_G^2 = 0.086, SE = 0.005) was found to be three times higher than for cases with ID (h_G^2 = 0.029, SE = 0.013, P = 0.015).

To further examine the apparent polygenic heterogeneity across subtypes, we investigated how PRS trained on different phenotypes were distributed across distinct ASD subgroups. We focused on phenotypes showing strong genetic correlation with ASD (e.g., educational attainment), but included also traits with little or no correlation to ASD (e.g., BMI) as negative controls. In this analysis, we regressed the normalized scores on ASD subgroups while including covariates for batches and principal components in a multivariate regression. Of the eight phenotypes we evaluated, only the cognitive phenotypes showed strong heterogeneity (educational attainment[26] P = 1.8 x 10^{-8}, IQ[43] P = 3.7 x 10^{-9}) (**Figure S4.4.8**). Interestingly, all case-control groups with or without ID showed significantly different loading for the two cognitive phenotypes: controls with ID have the lowest score followed by ASD cases with ID, and ASD cases without ID have significantly higher scores again than any other group (P = 2.6 x 10^{-12} for educational attainment, P = 8.2 x 10^{-12} for IQ).

With respect to the diagnostic sub-categories constructed hierarchically from ASD subtypes (**Table S2.3.1**), it was again the cognitive phenotypes that showed the strongest heterogeneity across the diagnostic classes (educational attainment P = 2.6 x 10^{-11}, IQ P = 3.4 x 10^{-8}), while neuroticism[27] (P = 0.0015), chronotype[44] (P = 0.011) and subjective well-being[27] (P = 0.029) showed weaker but nominally significant degree of heterogeneity, and SCZ, MDD and BMI were non-significant across the groups (P > 0.19) (**Figure 3**). This pattern weakened only slightly when excluding subjects with ID (**Figure S4.4.9**). For neuroticism there was a clear split with atypical and other/unspecified PDD cases having

significantly higher PRS than childhood autism and Asperger's, P = 0.00013. Considering the genetic overlap of each subcategory with each phenotype, the hypothesis of homogeneity across sub-phenotypes was strongly rejected ($P = 1.6 \times 10^{-11}$), thereby establishing that these sub-categories indeed have differences in their genetic architectures.

Focusing on educational attainment, significant enrichment of PRS was found for Asperger's syndrome ($P = 2.0 \times 10^{-17}$) in particular, and for childhood autism ($P = 1.5 \times 10^{-5}$), but not for the group of other/unspecified PDD (P = 0.36) or for atypical autism (P = 0.13) (**Figure 3).** Excluding individuals with ID only changes this marginally, with atypical autism becoming nominally significant (P = 0.020) **(Figure S4.4.9**). These results reveal that the genetic architecture underlying educational attainment is indeed shared with ASD but to a variable degree across the disorder spectrum. We find that the observed excess in ASD subjects of alleles positively associated with education attainment[45,46] is confined to Asperger's and childhood autism, and it is not seen here in atypical autism nor in other/unspecified PDD.

Finally, we evaluated the predictive ability of ASD PRS using five different sets of target and training samples within the combined iPSYCH-PGC sample. The observed mean variance explained by PRS (Nagelkerke's R^2) was 2.45% ($P = 5.58 \times 10^{-140}$) with a pooled PRS-based case-control odds ratio OR = 1.33 (CI.95% 1.30 –1.36) (**Figures S4.4.2** and **S4.4.4)**. Dividing the target samples into PRS decile groups revealed an increase in OR with increasing PRS. The OR for subjects with the highest PRS increased to OR = 2.80 (CI.95% 2.53–3.10) relative to the lowest decile (**Figures 4a and S4.4.5**). Leveraging correlated phenotypes in an attempt to improve prediction of ASD, we generated a multi-phenotype PRS as a weighted sum of phenotype specific PRS (see Methods). As expected,

Nagelkerkes's R^2 increased for each PRS included attaining its maximum at the full model at 3.77% ($P = 2.03 \times 10^{-215}$) for the pooled analysis with an OR = 3.57 (CI.95% 3.22-3.96) for the highest decile (**Figures 4b and S4.4.6-7**). 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 way, which can be reinforced by adding SNP-weights from ASD correlated traits.

Functional annotation

In order to obtain information on possible biological underpinnings of our GWAS results we conducted several analyses. First, we examined how the ASD h_G^2 partitioned on functional genomic categories as well as on cell type-specific regulatory elements using stratified LD score regression[47]. This analysis revealed significant enrichment of heritability in conserved DNA regions and H3K4me1 histone marks[48], as well as in genes expressed in central nervous system (CNS) cell types as a group (**Figures S4.3.2** and **S4.3.3**), which is in line with observations in schizophrenia, major depression[29], and bipolar disorder[24]. Analyzing the enhancer associated mark H3K4me1 in individual cell/tissue[48], we found significant enrichment in brain and neuronal cell lines (**Figure S4.3.4**). The highest enrichment was observed in the developing brain, germinal matrix, cortex-derived neurospheres, and embryonic stem cell (ESC)-derived neurons, consistent with ASD as a neurodevelopmental disorder with largely prenatal origins, as supported by data from analysis of rare *de novo* variants[33].

Common variation in ASD is located in regions that are highly enriched with regulatory elements predicted to be active in human corticogenesis (**Figures S4.3.2-4**). Since most gene regulatory events occur at a distance via chromosome looping, we leveraged Hi-C data from

germinal zone (GZ) and post-mitotic zones cortical plate (CP) in the developing fetal brain to identify potential target genes for these variants[49]. We performed fine mapping of 29 loci to identify the set of credible variants containing likely causal genetic risk[50] (see Methods). Credible SNPs were significantly enriched with enhancer marks in the fetal brain (**Figure S4.5.1**), again confirming the likely regulatory role of these SNPs during brain development.

Based on location or evidence of physical contact from Hi-C, the 380 credible SNPs (29 loci) could be assigned to 95 genes (40 protein-coding), including 39 SNPs within promoters that were assigned to 9 genes, and 16 SNPs within the protein coding sequence of 8 genes (**Table S3.4.1, Figure S4.5.1**). Hi-C identified 86 genes, which interacted with credible SNPs in either the CP or GZ during brain development. Among these genes, 34 are interacting with credible SNPs in both CP and GZ, which represent a high-confidence gene list. Notable examples are illustrated in **Figure 5** and highlighted in **Box 1**. By analyzing their mean expression trajectory, we observed that the identified ASD candidate genes (**Table S3.4.1**) show highest expression during fetal corticogenesis, which is in line with the enrichment of heritability in the regulatory elements in developing brain (**Figure 5e-g**). Interestingly, both common and rare variation in ASD preferentially affects genes expressed during corticogenesis[33], 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. While early GWAS permitted estimates that common polygenic variation should explain a

substantial fraction of the heritability of ASD, individually significant loci remained elusive. This was suspected to be due to limited sample size since studies of schizophrenia – with similar prevalence, heritability and reduced fitness – and major depression achieved striking results only when sample sizes five to ten times larger than available in ASD were employed. This study has finally borne out that expectation with definitively demonstrated significant "hits".

Here we report the first common risk variants robustly associated with ASD by using 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 utilizing 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 fact that even among the loci defined in ASD alone there was additional evidence in these other trait scans indicate that the polygenic architecture of ASD is significantly shared with risk to adult psychiatric illness and higher educational attainment and intelligence. It should be noted that the MTAG analyses were carried out as three pairwise analyses. This way we avoid the complex interactions that could arise if we ran three or four correlated phenotypes at a time[9]. Indeed, what we get, despite the secondary summary statistics coming from large, high-powered studies, are relatively modest weights to the contributions from these statistics, because the genetic correlations are modest. The largest weight was 0.27 for schizophrenia, followed by 0.24 for major depression, and 0.11 for educational attainment. Thus all loci identified by MTAG have substantial contributions from ASD alone (Table 1a, b) and will most likely also be identified in future ASD-only GWAS with increasing sample sizes.

In most GWAS studies there has been little evidence of heterogeneity of association across phenotypic subgroups. In this study, however, we see strong heterogeneity of genetic overlap with other traits when our ASD samples are broken into distinct subsets. In particular, the excess of alleles associated with higher intelligence and educational attainment was only observed in the higher functioning categories (particularly Asperger's syndrome and individuals without comorbid ID) – and not in the other/unspecified PDD and ID categories. This is reminiscent, and logically inverted, from the much greater role of spontaneous mutations in these latter categories, particularly in genes known to have an even larger impact in cohorts ascertained for ID/DD[51]. Interestingly, other/unspecified PDD and atypical autism also have a significantly higher PRS for neuroticism than childhood autism and Asperger's. These different enrichment profiles observed provide evidence for a heterogeneous and qualitatively different genetic architecture between sub-types 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 and without ID, and highest in Asperger's provide genetic evidence of longstanding observations. In particular, this aligns well with the observation that *de novo* variants are more frequently observed in ASD cases with ID compared to cases without comorbid ID, that IQ correlates positively with family history of psychiatric disorders[52] and that severe ID (encompassing many syndromes that confer high risk to ASD) show far less heritability than is observed for mild ID[53], intelligence in general[54] and ASDs. Thus it is perhaps unsurprising that our data suggests that the contribution of common variants may be more prominent in high-functioning ASD cases 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 (**Figures S4.3.2-4)** and brain transcriptome enrichment (**Figure 5e-g**)) as well as at single loci (**Figure 5a-d, Box 1**) highlighted the involvement of processes relating to brain development and neuronal function. Notably, a number of genes located in the identified loci have previously been linked to ASD risk in studies of *de novo* and rare variants (**Box 1**, **Table S3.1.3**), including *PTBP2, CADPS,* and *KMT2E*, which were found to interact with credible SNPs in the Hi-C analysis (*PTBP2, CADPS*) or contain a loss-of-function credible SNP (*KMT2E*). Interestingly, aberrant splicing of CADPS' sister gene CADPS2, which has almost identical function, has been found in autism cases and *Cadps2* knockout mice display behavioral anomalies with translational relevance to autism[55]. *PTBP2* encodes a neuronal splicing factor and alterations in alternative splicing have been identified in brains from individuals diagnosed with ASD[56].

In summary, we have established a first compelling 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.

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 May 1st 1981 and December 31st 2005. Eligible were singletons born to a known mother and resident in Denmark on their one-year birthday. 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 onwards) as well as at outpatient psychiatric clinics (from 1995 onwards). Cases 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). As controls we selected a random sample from the set of eligible children excluding those with an ASD diagnosis by 2013.

The samples were linked using the unique 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 amplified in triplicates as described previously[13,57]. Genotyping was performed at the Broad Institute of Harvard and MIT (Cambridge, MA, USA) using the PsychChip array from Illumina (CA, San Diego, USA) according to the instructions of the manufacturer. Genotype calling of markers with minor allele frequency (maf) > 0.01 was performed by merging callsets from GenCall[58] and Birdseed[59] while less frequent variants were called with zCall[60]. Genotyping and data processing was 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 (https://genome.au.dk).

The study was approved by the Regional Scientific Ethics Committee in Denmark and the Danish Data Protection Agency.

Psychiatric Genomic Consortium (PGC) samples

In brief, five cohorts provided genotypes to the sample (n denoting the number of trios for which genotypes were available): The Geschwind Autism Center of Excellence (ACE; N = 391), the Autism Genome Project[61] (AGP; N = 2272), the Autism Genetic Resource Exchange[62,63] (AGRE; N = 974), the NIMH Repository (https://www.nimhgenetics.org/available_data/autism/), the Montreal[64]/Boston Collection (MONBOS; N = 1396, and the Simons Simplex Collection [65,66](SSC; N = 2231). The trios were analyzed as cases and pseudo controls. A detailed description of the sample is available on the PGC web site: https://www.med.unc.edu/pgc/files/resultfiles/PGCASDEuro_Mar2015.readme.pdf and even more details are provided in Anney et al[5]. Analyses of the PGC genotypes were conducted on LISA on the Dutch HPC center SURFsara (https://userinfo.surfsara.nl/systems/lisa).

Follow-up samples

As follow-up for the loci with p-values less than 10^{-6} we asked for look up in 5 samples of Nordic and Eastern European origin with altogether 2,119 cases and 142,379 controls: 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), deCODE (Iceland 574 cases and 136,968 controls; Eastern Europe: 296 cases and 388 controls). See supplementary for details.

GWAS analysis

Ricopili[15], the pipeline developed by the Psychiatric Genomics Consortium (PGC) Statistical Analysis Group was used for quality control, imputation, principle component analysis (PCA) and primary association analysis. For details see supplementary information. The data was processed separately in the 23 genotyping batches in the case of iPSYCH and separately for each study in the PGC sample. Phasing was achieved using SHAPEIT[16] and imputation done by IMPUTE2[67,68] with haplotypes from the 1000 Genomes Project, phase 3[69,70] (1kGP3) as reference. Chromosome X was imputed using 1000 Genomes Project, phase 1[71] haplotypes.

After excluding regions of high LD[72], the genotypes were pruned down to a set of roughly 30k markers. See supplementary information for details. Using PLINK's[73,74] 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 using smartPCA[75,76]. In iPSYCH a subsample of European ancestry was selected as an ellipsoid in the space of PC1-3 centred and scaled using the mean and 8 standard deviation of the subsample whose parents and grandparents were all known to have been born in Denmark (n=31500). In the PGC sample the CEU subset was chosen using a Euclidian distance measure weighted by the variance explain for each of the first 3 PCs. Individuals more distant than 10 standard deviations from the combined CEU and TSI HapMap reference populations were excluded. We conducted a secondary PCA to provide covariates for the

association analyses. Numbers of subjects in the data generation flow for the iPSYCH sample are found in the descriptive table **Table S1.1.1.**

Association analyses were done by applying PLINK 1.9 to the imputed dosage data (the sum of imputation probabilities P(A1A2) + 2P(A1A1)). In iPSYCH we included the first four principal components (PCs) as covariates as well as any PC beyond that, which were significantly associated with ASD in the sample, while the case-pseudo-controls from the PGC trios required no PC covariates. Combined results for iPSYCH and for iPSYCH with the PGC was achieved by meta-analysing batch-wise and study-wise results using METAL[77] (July 2010 version) employing an inverse variance weighted fixed effect model[22]. 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 ≥ 0.01 and an effective sample size (see supplementary info) of at least 70% of the study maximum. The degree to which the deviation in the test statistics can be ascribed to cryptic relatedness and population stratification rather than to polygenicity was measured from the intercept in LD score regression[23] (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 was paired up with the summary statistics for each of major depression[29] (excluding the Danish sampled but including summary statistics from 23andMe[78], 111,902 cases, 312,113 controls, and mean χ^2=1.477), schizophrenia[15] (also excluding the Danish samples, 34,129 cases, 45,512 controls, and mean χ^2=1.804) and educational attainment[26] (328,917 samples and mean χ^2=1.648). These are studies that have considerably more statistical power than the ASD scan, but the genetic correlations are modest in the context of MTAG, so the

weights ascribed to the secondary phenotypes in the MTAG analyses remain relatively low (no higher than 0.27). See supplementary methods for details.

The results were clumped and we highlighted loci of interest by selecting those that were significant at 5 x 10^{-8} in the iPSYCH-PGC meta-analysis or the meta-analysis with the follow-up sample or were significant at 1.67 x $l0^{-8}$ in any of the three MTAG analyses. The composite GWAS consisting of the minimal p-values at each marker over these five analyses wsa used as a background when creating Manhattan plots for the different analyses showing both what is maximally achieved and what the individual analysis contributes to that.

Gene-based association and gene-set analyses.

MAGMA 1.06[32] was applied to the summary statistics to test for gene-based association. Using NCBI 37.3 gene definitions and restricting the analysis to SNPs located within the transcribed region, mean SNP association was tested with the sum of -log(SNP p-value) as 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. 2013[33] as well as constrained, loss-of-function intolerant genes (pLI>0.9)[34,35] derived from data of the Exome Aggregation Consortium (see supplementary methods for details). Another was an agnostic analysis of the Gene Ontology sets[37,38] for molecular function from MsigDB 6.0[39]. We analyzed only genes outside the broad MHC region (hg19:chr6:25-35M) and included only gene sets with 10-1000 genes.

SNP heritability

SNP heritability, h_G^2, was estimated using LDSC[23] for the full sample and GCTA[40-42] for subsamples too small for LDSC. For LDSC we used precomputed LD scores based on the European ancestry samples of the 1000 Genomes Project[71] restricted to HapMap3[79] SNPs

(downloaded from the https://github.com/bulik/ldsc). The summary stats with standard LDSC filtering were regressed onto these scores. For liability scale estimates, we used a population prevalence for Denmark of 1.22%[21]. Lacking proper prevalence estimates for subtypes, we scaled the full spectrum prevalence based on the composition of the case sample.

For subsamples too small for LDSC, the GREML approach of GCTA[40-42] 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 (GRM) was fitted for the association sample (i.e. the subjects of European ancestry with $\hat{\pi} \leq 0.2$) providing a relatedness estimate for all pairwise combinations of individuals. Estimation of the phenotypic variance explained by the SNPs (REML) was performed including PC 1-4 as continuous covariates together with any other PC that was nominally significantly associated to the phenotype as well as batches as categorical indicator covariates. Testing equal heritability for non-overlapping groups was done by permutation test (with 1000 permutations) keeping the controls and randomly the assigning the different case labels.

Following Finucane et al[47], 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 summary statistics on to the cell-type specific LD scores download 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 did a similar analysis using 220 cell type–specific annotations divided into 10 overlapping groups.

In addition to this, we conducted an analysis based on annotation derived from data on H3K4Me1 imputed gapped peaks data from the Roadmap Epigenomics Mapping Consortium[80,81]; more specifically information excluding the broad MHC-region (chr6:25-35MB).

Genetic correlation

For the main samples, SNP correlations, r_G, were estimated using LDSC[23] and for the analysis of ASD subtypes and subgroups where the sample size were generally small, we used GCTA[42]. In both cases, we followed the same procedures as explained above. For all but a few phenotypes, LDSC estimates of correlation were achieved by uploaded to LD hub[25] for comparison to all together 234 phenotypes.

Polygenic risk scores

For the polygenic risk scores (PRS) we clumped the summary stats applying standard Ricopili parameters (see supplementary methods for details). To avoid potential strand conflicts we excluded all ambiguous markers for summary statistics not generated by Ricopili using the same imputation reference. PRS were generated at the default p-value thresholds (5e-8, 1e-6, 1e-4, 0.001, 0.01, 0.05, 0.1, 0.2, 0.5 and 1) as a weighted sum of the allele dosages. 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 prior to analysis.

We evaluated the predictive power using Nagelkerke's R^2 and plots of odds ratios and confidence intervals over score deciles. Both R^2 and odds ratios were estimated in regression analyses including the relevant PCs and indicator variable for genotyping waves.

Lacking a large ASD sample outside of iPSYCH and PGC, we trained a set of PRS for ASD internally in the following way. We divided the sample in five subsamples of roughly equal size respecting the division into batches. We then ran five GWAS leaving out each group in turn from the training set and meta-analyzed these with the PGC results. This produced a set of PRS for each of the five subsamples trained on their complement. Prior to analyses, each score was normalized on the group where 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 PRS by adding to the internally trained ASD score the PRS of other highly correlated phenotypes in a weighted sum. See supplementary info for details.

To analyze ASD subtypes in relation PRS we defined a hierarchical set of phenotypes in the following way: First hierarchical subtypes was childhood autism, hierarchical atypical autism was defined as everybody with atypical autism and no childhood autism diagnosis, hierarchical Asperger's as everybody with an Asperger's 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 of that consists of everybody with such a diagnosis and none of neither preceding ones (**Table S2.3.1**). We examined the distribution over the distinct ASD subtypes of PRS for a number of phenotypes showing high r_G with ASD (as well as a few with low r_G as negative controls), by doing multivariate regression of the scores on the subtypes while adjusting for relevant PCs and wave indicator variables in a linear regression. See supplementary methods for details.

Hi-C analysis

The Hi-C data was generated from two major cortical laminae: the germinal zone (GZ), containing primarily mitotically active neural progenitors, and the cortical and subcortical plate (CP), consisting primarily of post-mitotic neurons[49]. We first derived a set of credible SNPs (putative causal SNPs) from the identified top ranking 29 loci using CAVIAR[50]. To test whether credible SNPs are enriched in active marks in the fetal brain[81], we employed GREAT as previously described[49,82]. Credible SNPs were, sub-grouped into those without known function (unannotated) and functionally annotated SNPs (SNPs in the gene promoters and SNPs that cause nonsynonymous variants) (**Figure S4.5.1**). Then we integrated unannotated credible SNPs with chromatin contact profiles during fetal corticogenesis[49], defining genes physically interacting with intergenic or intronic SNPs (**Figure S4.5.1**).

Spatiotemporal transcriptomic atlas of human brain was obtained from Kang et al[83]. We used transcriptomic profiles of multiple brain regions with developmental epochs that span prenatal (6-37 post-conception week, PCW) and postnatal (4 months-42 years) periods. Expression values were log-transformed and centered to the mean expression level for each sample using a *scale(center=T, scale=F)+1* function in R. ASD candidate genes identified by Hi-C analyses (**Figure S4.5.1)** were selected for each sample and their average centered expression values were calculated and plotted.

Summary statistics

The summary statistics are available for download the iPSYCH download page http://ipsych.au.dk/downloads/ and at the PGC download site: https://www.med.unc.edu/pgc/results-and-downloads.

Acknowledgements

The iPSYCH project is funded by the Lundbeck Foundation (grant numbers 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 MJD), and NIMH (5U01MH094432-02 to MJD). 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 Michael O'Donovan and ADB). 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 ADB). Dr SD Rubeis was supported by NIH grant MH097849 (to JDB), MH111661 (to JDB), and the Seaver Foundation (to SDR and JDB). Dr J Martin was supported by the Wellcome Trust (grant no: 106047). We thank the research participants and employees of 23andMe for making this work possible.

Members of the PGC-MDD Group and the 23andMe Research Team

See list in the supplementary file.

Author contributions

Analysis: JG, SR, TDA, MM, RW, HW, JP, FB, JHC, CC, KD, SDR, BD, SD, MEH, SH, DH, HH, LK, JM, JM, ARM, MN, TN, DSP, TP, BSP, PQ, JR, EBR, KR, PR, ES, SS, FKS, SS, PFS, PT, GBW, XX, DG.

JG, BMN, MJD, and ADB supervised and coordinated the analyses.

Sample and/or data provider and processing: JG, SR, MM, RW, EA, OA, RA, RB, JDB, JBG, MBH, FC, KC, DD, AD, JG, CH, MVH, CH, JM, AP, CP, MGP, JBP, KR, AR, GDS, HS, KS, CS, PGC-ASD, BUPGEN, PGC-MDD, 23andMeResearch Team, DH, OM, PBM, BMN, MJD, ADB.

Core PI group: DG, MN, DH, TW, OM, PBM, BMN, MJD, ADB.

Core writing group: JG, MJD, ADB.

Direction of study: MJD, ADB.

Box1. Selected loci and candidates (ordered by chromosome).

Gene	Locus* and supporting evidence	Gene function
NEGR1	Chr1:72,729,142 Shared ASD-MDD locus Locus also significant in depression[29,78], obesity and BMI[84-88] NEGR1 is the only protein-coding gene in the locus NEGR1 is supported by brain Hi-C and eQTL analyses[29]	NEGR1 (neuronal growth regulator 1) is an adhesion molecule modulating synapse formation in hippocampal neurons[89,90] and neurite outgrowth[91,92]. It is member of the IgLON protein family implicated in synaptic plasticity and axon extension[93-95]. Predominantly expressed (and developmentally upregulated) in hippocampus and cortex[96] and also hypothalamus[97].
PTBP2	Chr1:96,561,801 ASD locus Locus also significant in BMI[84,86,88] and weight[84]. In schizophrenia the locus show p-value of $6.5x10^{-6}$ [15] PTBP2 is the nearest protein-coding gene, approx. 625kb from index SNP. De novo and rare variants in PTBP2 have been reported in ASD cases[1,3,98]. PTBP2 is supported by Hi-C results in this study (Fig. 5d)	PTBP2 is also known as nPTB (neuronal PTB) or brPTB (brain PTB) and is a splicing regulator. PTBP1 and its paralog PTBP2 bind to intronic polypyrimidine tracts in pre-mRNAs and target large sets of exons to coordinate alternative splicing programs during development[99]. Several switches in the expression of PTBP1 and PTBP2 regulate alternative splicing during neurogenesis and neuronal differentiation [100-103].
CADPS	Chr3:62,481,063 Shared ASD-Educational attainment locus Locus also significant in study of cognitive decline rate[104] 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 brain and pituitary (gtexportal), immunoreactive CAPS-1 is localized in neural and various endocrine tissues[105]. In hippocampal synapses, CADPS regulates the pool of readily releasable vesicles at pre-synaptic terminals[106,107]
KCNN2	Chr5: 113,801,423 ASD locus (gene-wise analysis) Locus also significant in educational attainment[26,108]. KCNN2 synaptic levels are regulated by the E3 ubiquitin ligase UBE3A[109], of which overexpression has been linked to ASD risk[109,110].	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 CNS neurons, activation of KCNN2 modulates neuronal excitability by causing membrane hyperpolarization[111]. Hippocampal KCNN2 has roles in the formation of new memory[112], encoding and consolidation of contextual fear[113], and in drug-induced plasticity[114].
KMT2E	Chr7:104,744,219 ASD locus Locus also significant in schizophrenia[15,115] and in ASD-schizophrenia meta-analysis[5]. KMT2E de novo mutations are associated with ASD at FDR<0.1[116] A KMT2E credible SNP is a loss-of-function variant (Table S3.4.1)	KMT2E encodes Histone-lysine N-methyltransferase 2E and forms a family together with SETD5[117,118]. Evidence suggest that recognition of the histone H3K4me3 mark by the KMT2E PHD finger can facilitate the recruitment of KMT2E to transcription-active chromatin regions[119,120]. KMT2E has been implicated in chromatin regulation, control of cell cycle progression, and maintaining genomic stability[121].

MACROD2	Chr20: 14836243 ASD locus Locus found significant in previous ASD GWAS[61] but not supported in larger study[122] MACROD2 is the only protein-coding gene in the locus	MACROD2 is a nuclear enzyme that binds to mono-ADP-ribosylated (MARylated) proteins and functions as an eraser of mono-ADP-ribosylation[123]. Intracellular MARylated histones and GSK3β are substrates of MACROD2, and the removal of MAR from GSK3β is responsible for reactivating of its kinase activity[123]. This gene is expressed in lung and multiple regions of the brain. Low or no expression across most other tissue (https://www.gtexportal. org).

*position of index SNP is listed.

Table 1. Genome-wide significant loci. Independent loci ($r^2 < 0.1$, distance > 400kb) with index variant (Index var), chromosome (CHR), chromosomal position (BP), alleles (A1/A2), allele frequency of A1 (FRQ), estimate of effect (β) with respect to A1, standard error of β (SE), and the association p-value of the index variant (P). The "Analysis" column shows the analysis that generated the particular result. The column "Support from other scans" lists the analyses that also support the locus. For the ASD scan results, this shows the genome wide significant results in the locus from the other scans, and for results from any other analysis, it shows the result from the ASD scan. An * indicate different lead SNP. The column "Nearest genes" lists nearest genes from within 50kb of the $r^2 \geq 0.6$ LD friends of the index variant (intergenic variants marked with a dash). **a.** Loci from the ASD scan and the combined analysis with the follow-up sample. Column "Analysis" indicates which results stem from the original scan and which comes from the combined. **b.** Loci from the three MTAG analyses with schizophrenia (SCZ)[15], educational attainment (Edu)[26] and major depression (MD)[29] – loci not already represented in section a.

	Index var	CHR	BP	P	β	SE	A1/A2	FRQ	Analysis	Support from other scans			Nearest genes
										Scan	P	β	
a	rs910805	20	21248116	2.04×10^{-9}	-0.096	0.016	A/G	0.76	ASD	ASD-SCZ	1.5×10^{-10}	-0.069	*KIZ, XRN2, NKX2-2, NKX2-4*
										ASD-Edu*	2.0×10^{-8}	-0.061	
	rs10099100	8	10576775	1.07×10^{-8}	0.084	0.015	C/G	0.331	ASD	Comb ASD	9.6×10^{-9}	0.078	*C8orf74, SOX7, PINX1*
										ASD-Edu	1.6 x 10-8	0.056	
	rs201910565	1	96561801	2.48×10^{-8}	-0.077	0.014	A/AT	0.689	Comb ASD	ASD	3.4×10^{-7}	-0.033	*LOC102723661, PTBP2*
	rs71190156	20	14836243	2.75×10^{-8}	-0.078	0.014	GTTTT TTT/G	0.481	ASD	Comb ASD	3.0 x 10-8	-0.072	*MACROD2*
										ASD-Edu	1.2×10^{-8}	0.053	
	rs111931861	7	104744219	3.53×10^{-8}	-0.216	0.039	A/G	0.966	Comb ASD	ASD	1.1×10^{-7}	-0.094	*KMT2E, SRPK2*
b	rs2388334	6	98591622	3.34×10^{-12}	-0.065	0.009	A/G	0.517	ASD-Edu	ASD	1.0×10^{-6}	-0.068	*MMS22L, POU3F2*
	rs325506	5	104012303	3.26×10^{-11}	0.057	0.009	C/G	0.423	ASD-MD	ASD	3.5×10^{-7}	0.071	*NUD12*
	rs11787216	8	142615222	1.99×10^{-9}	-0.058	0.010	T/C	0.364	ASD-Edu	ASD	2.6×10^{-6}	-0.030	*MROH5*
	rs1452075	3	62481063	3.17×10^{-9}	0.061	0.010	T/C	0.721	ASD-Edu	ASD	2.1×10^{-7}	0.035	*CADPS*
	rs1620977	1	72729142	6.66×10^{-9}	0.056	0.010	A/G	0.26	ASD-MD	ASD	1.2×10^{-4}	0.062	*NEGR1*
	rs10149470	14	104017953	8.52×10^{-9}	-0.049	0.008	A/G	0.487	ASD-MD	ASD	8.5×10^{-5}	-0.056	*MARK3, CKB, TRMT61A, BAG5, APOPT1, KLC1, XRCC3*
	rs16854048	4	42123728	1.29×10^{-8}	0.069	0.012	A/C	0.858	ASD-MD	ASD	5.9×10^{-5}	0.082	*SLC30A9, BEND4, TMEM33, DCAF4L1*

Figure legends

Figure 1. Manhattans plots. a: The main ASD scan with the results of the combined analysis with the follow-up sample in yellow in the foreground. GWS clumps are painted green with index SNPs as diamonds. **b-d**: Manhattan plots for three MTAG scans of ASD together with, respectively, schizophrenia[15], educational attainment[26] and major depression[29] (see Figures **S.2.71-74** for full size plots). In all panels the results of the composite of the five analyses (consisting for each marker of the minimal p-value of the five) is shown in grey in the background.

Figure 2. **Genetic correlation with other traits**. Significant genetic correlations between ASD and other traits after Bonferroni correction for testing a total of 234 traits available at LDhub with the addition of a handful of new phenotypes. The results here corresponds to the following GWAS analyses: IQ[43], educational attainment[26], college[124], self-reported tiredness[125], neuroticism[27], subjective well-being[27], schizophrenia[15], major depression[29], depressive symptoms[27], ADHD[30], and chronotype[44]. See Table **S3.3.2** for the full output of this analysis.
* Indicates that the values are from in-house analyses of new summary statistics not yet included in LD Hub.

Figure 3. Profiling PRS load across distinct ASD sub-groups for 8 different phenotypes (schizophrenia[15], major depression[29], educational attainment[26], human intelligence[43], subjective well-being[27], chronotype[44], neuroticism[27] and BMI[88]. The bars show coefficients from multivariate regression of the 8 normalized scores on the distinct ASD sub-types (adjusting for batches and PCs). The subtypes are the hierarchically defined subtypes for childhood autism (hCHA), atypical autism (hATA), Asperger's (hAsp), and the lumped pervasive disorders developmental group (hPDM). Beware that the orientation 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 can be seen in Figure **S4.4.9**, and with ID as a subtype in Figure **S4.4.8**. 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) (**Figure S4.4.10**).

Figure 4. Decile plots (Odds Ratio (OR) by PRS within each decile): **a.** Decile plot with 95%-CI for the internally trained ASD score (P-value threshold is 0.1). **b.** Decile plots on a weighted sums

of PRSs starting with the ASD score of panel a and successively adding the scores for major depression[29], subjective well-being[27], schizophrenia[15], educational attainment[26], and chronotype[44=44]. In all instances the P-value threshold for the score used is the one with the highest Nagelkerke's R^2. Figures **S4.4.5** and **S4.4.7** show the stability across leave-one out groups that was used to create these combined results.

Figure 5. Chromatin interactions identify putative target genes of ASD loci. a-d. Chromatin interaction maps of credible SNPs to the 1Mb flanking region, providing putative candidate genes that physically interact with credible SNPs. Gene Model is based on Gencode v19 and putative target genes are marked in red; Genomic coordinate for a credible SNP is labeled as GWAS; -log10(P-value), significance of the interaction between a SNP and each 10kb bin, grey dotted line for FDR=0.01; TAD borders in CP and GZ. **e-f.** Developmental expression trajectories of ASD candidate genes show highest expression in prenatal periods. **g**. ASD candidate genes are highly expressed in the developing cortex as compared to other brain regions.

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* (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* (2015).
5. Anney, R. J. L. *et al.* 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. *Molecular 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. *Annals of human genetics* **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 research* **1380**, 78–84 (2011).
8. Anney, R. *et al.* Individual common variants exert weak effects on the risk for autism spectrum disorderspi. *Hum Mol Genet* **21**, 4781–4792 (2012).
9. Turley, P. *et al.* Mtag: Multi-trait analysis of gwas. *bioRxiv* (2017).
10. Pedersen, C. B. *et al.* The ipsych2012 case-cohort sample: new directions for unravelling the genetic and environmental architecture of severe mental disorders. *Molecular Psychiatry* (2017).
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. *Scandinavian journal of public health* **39**, 54–57 (2011).
13. Hollegaard, M. *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. Delaneau, O., Marchini, J. & Zagury, J.-F. A linear complexity phasing method for thousands of genomes. *Nature methods* **9**, 179–181 (2011).
17. Howie, B., Marchini, J. & Stephens, M. Genotype imputation with thousands of genomes. *G3 (Bethesda)* **1**, 457–470 (2011).
18. Cross-Disorder Group of the Psychiatric Genomics Consortium *et al.* Genetic relationship between five psychiatric disorders estimated from genome-wide snps. *Nature genetics* **45**, 984–994 (2013).
19. Gratten, J., Wray, N. R., Keller, M. C. & Visscher, P. M. Large-scale genomics unveils the genetic architecture of psychiatric disorders. *Nature neuroscience* **17**, 782–790 (2014).
20. Purcell, S. M. *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).
21. Hansen, S. N., Overgaard, M., Andersen, P. K. & Parner, E. T. Estimating a population cumulative incidence under calendar time trends. *BMC medical research methodology* **17**, 7 (2017).
22. Begum, F., Ghosh, D., Tseng, G. C. & Feingold, E. Comprehensive literature review and statistical considerations for gwas meta-analysis. *Nucleic acids research* **40**, 3777–3784 (2012).

23. Bulik-Sullivan, B. K. *et al.* Ld score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291–295 (2015).
24. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nature Genetics* 1061–4036 (2015).
25. 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).
26. Okbay, A. *et al.* Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539–542 (2016).
27. Okbay, A. *et al.* Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nature genetics* **48**, 624–633 (2016).
28. Clarke, T.-K. *et al.* Common polygenic risk for autism spectrum disorder (asd) is associated with cognitive ability in the general population. *Mol Psychiatry* (2015).
29. Wray, N. R. *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depressive disorder. *bioRxiv* (2017).
30. Demontis, D. *et al.* Discovery of the first genome-wide significant risk loci for adhd. *Nature Genetics* (2017).
31. Pourcain, B. S. *et al.* Asd and schizophrenia show distinct developmental profiles in common genetic overlap with population-based social communication difficulties (2017).
32. 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).
33. Parikshak, N. N. *et al.* Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* **155**, 1008–1021 (2013).
34. Samocha, K. E. *et al.* A framework for the interpretation of de novo mutation in human disease. *Nat Genet* **46**, 944–950 (2014).
35. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
36. Pardiñas, A. F. *et al.* Common schizophrenia alleles are enriched in mutation-intolerant genes and maintained by background selection. *bioRxiv* (2016).
37. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. the gene ontology consortium. *Nature genetics* **25**, 25–29 (2000).
38. Consortium, G. O. Gene ontology consortium: going forward. *Nucleic acids research* **43**, D1049–D1056 (2015).
39. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 15545–15550 (2005).
40. 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).
41. Yang, J. *et al.* Common snps explain a large proportion of the heritability for human height. *Nat Genet* **42**, 565–569 (2010).
42. 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).
43. Sniekers, S. *et al.* Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nature genetics* **49**, 1546–1718 (2017).
44. Jones, S. E. *et al.* Genome-wide association analyses in 128,266 individuals identifies new morningness and sleep duration loci. *PLoS genetics* **12**, e1006125 (2016).

45. Robinson, E. B. *et al.* Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. *Nature Genetics* **48**, 552–555 (2016).
46. Weiner, D. *et al.* Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders (2017).
47. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature Genetics* **47**, 1228–1235 (2015).
48. Shlyueva, D., Stampfel, G. & Stark, A. Transcriptional enhancers: from properties to genome-wide predictions. *Nature reviews. Genetics* **15**, 272–286 (2014).
49. Won, H. *et al.* Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* **538**, 523–527 (2016).
50. 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).
51. Kosmicki, J. A. *et al.* Refining the role of de novo protein-truncating variants in neurodevelopmental disorders by using population reference samples. *Nature genetics* **49**, 504–510 (2017).
52. Robinson, E. B. *et al.* Autism spectrum disorder severity reflects the average contribution of de novo and familial influences. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 15161–15165 (2014).
53. Reichenberg, A. *et al.* Discontinuity in the genetic and environmental causes of the intellectual disability spectrum. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 1098–1103 (2016).
54. Polderman, T. J. C. *et al.* Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature genetics* **47**, 702–709 (2015).
55. Sadakata, T. *et al.* Autistic-like phenotypes in cadps2-knockout mice and aberrant cadps2 splicing in autistic patients. *The Journal of clinical investigation* **117**, 931–943 (2007).
56. Parikshak, N. N. *et al.* Genome-wide changes in lncrna, splicing, and regional gene expression patterns in autism. *Nature* **540**, 423–427 (2016).
57. Børglum, A. D. *et al.* Genome-wide study of association and interaction with maternal cytomegalovirus infection identifies new schizophrenia loci. *Molecular Psychiatry* **19**, 325–333 (2014). Published online 29 January 2013.
58. Illumina gencall data analysis software. Tech. Rep., Illumina, Inc. (2005).
59. Korn, J. M. *et al.* Integrated genotype calling and association analysis of snps, common copy number polymorphisms and rare cnvs. *Nature genetics* **40**, 1253–1260 (2008).
60. Goldstein, J. I. *et al.* zcall: a rare variant caller for array-based genotyping: genetics and population analysis. *Bioinformatics (Oxford, England)* **28**, 2543–2545 (2012).
61. Anney, R. *et al.* A genome-wide scan for common alleles affecting risk for autism. *Hum Mol Genet* **19**, 4072–4082 (2010).
62. Lajonchere, C. M. & Consortium, A. Changing the landscape of autism research: the autism genetic resource exchange. *Neuron* **68**, 187–191 (2010).
63. Geschwind, D. H. *et al.* The autism genetic resource exchange: a resource for the study of autism and related neuropsychiatric conditions. *American journal of human genetics* **69**, 463–466 (2001).
64. Gauthier, J. *et al.* Autism spectrum disorders associated with x chromosome markers in french-canadian males. *Molecular psychiatry* **11**, 206–213 (2006).
65. Fischbach, G. D. & Lord, C. The simons simplex collection: a resource for identification of autism genetic risk factors. *Neuron* **68**, 192–195 (2010).

66. 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).
67. 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).
68. 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).
69. Consortium, . G. P. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
70. Sudmant, P. H. *et al.* An integrated map of structural variation in 2,504 human genomes. *Nature* **526**, 75–81 (2015).
71. 1000 Genomes Project Consortium *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
72. Price, A. L. *et al.* Long-range ld can confound genome scans in admixed populations. *Am J Hum Genet* **83**, 132–5; author reply 135–9 (2008).
73. Purcell, S. & Chang, C. Plink verion 1.9 (2015).
74. Chang, C. C. *et al.* Second-generation plink: rising to the challenge of larger and richer datasets. *GigaScience* **4**, 7 (2015). Original DateCompleted: 20150227.
75. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet* **2**, e190 (2006).
76. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904–909 (2006).
77. Willer, C. J., Li, Y. & Abecasis, G. R. Metal: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics (Oxford, England)* **26**, 2190–2191 (2010).
78. Hyde, C. L. *et al.* Identification of 15 genetic loci associated with risk of major depression in individuals of european descent. *Nature genetics* **48**, 1031–1036 (2016).
79. Consortium, I. H. . *et al.* Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010).
80. Ernst, J. & Kellis, M. Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues. *Nature biotechnology* **33**, 364–376 (2015).
81. Consortium, R. E. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
82. McLean, C. Y. *et al.* Great improves functional interpretation of cis-regulatory regions. *Nature biotechnology* **28**, 495–501 (2010).
83. Kang, H. J. *et al.* Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483–489 (2011).
84. Thorleifsson, G. *et al.* Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nature genetics* **41**, 18–24 (2009).
85. Willer, C. J. *et al.* Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nature genetics* **41**, 25–34 (2009).
86. Speliotes, E. K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature genetics* **42**, 937–948 (2010).
87. Berndt, S. I. *et al.* Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. *Nature genetics* **45**, 501–512 (2013).
88. Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).

89. 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 research* **1224**, 1–11 (2008).
90. Hashimoto, T., Maekawa, S. & Miyata, S. Iglon cell adhesion molecules regulate synaptogenesis in hippocampal neurons. *Cell biochemistry and function* **27**, 496–498 (2009).
91. Pischedda, F. *et al.* A cell surface biotinylation assay to reveal membrane-associated neuronal cues: Negr1 regulates dendritic arborization. *Molecular & cellular proteomics : MCP* **13**, 733–748 (2014).
92. Pischedda, F. & Piccoli, G. The iglon family member negr1 promotes neuronal arborization acting as soluble factor via fgfr2. *Frontiers in molecular neuroscience* **8**, 89 (2015).
93. Marg, A. *et al.* Neurotractin, a novel neurite outgrowth-promoting ig-like protein that interacts with cepu-1 and lamp. *The Journal of cell biology* **145**, 865–876 (1999).
94. Funatsu, N. *et al.* Characterization of a novel rat brain glycosylphosphatidylinositol-anchored protein (kilon), a member of the iglon cell adhesion molecule family. *The Journal of biological chemistry* **274**, 8224–8230 (1999).
95. 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. *The Journal of biological chemistry* **290**, 4330–4342 (2015).
96. 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. *Molecular and cellular neurosciences* **29**, 580–590 (2005).
97. 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).
98. Doan, R. N. *et al.* Mutations in human accelerated regions disrupt cognition and social behavior. *Cell* **167**, 341–354.e12 (2016).
99. Vuong, J. K. *et al.* Ptbp1 and ptbp2 serve both specific and redundant functions in neuronal pre-mrna splicing. *Cell reports* **17**, 2766–2775 (2016).
100. Boutz, P. L. *et al.* A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons. *Genes & development* **21**, 1636–1652 (2007).
101. 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. *Molecular cell* **27**, 435–448 (2007).
102. Spellman, R., Llorian, M. & Smith, C. W. J. Crossregulation and functional redundancy between the splicing regulator ptb and its paralogs nptb and rod1. *Molecular cell* **27**, 420–434 (2007).
103. Zheng, S. *et al.* Psd-95 is post-transcriptionally repressed during early neural development by ptbp1 and ptbp2. *Nature neuroscience* **15**, 381–8, S1 (2012).
104. Li, Q. S., Parrado, A. R., Samtani, M. N., Narayan, V. A. & Initiative, A. D. N. Variations in the fra10ac1 fragile site and 15q21 are associated with cerebrospinal fluid aβ1-42 level. *PloS one* **10**, e0134000 (2015).
105. Wassenberg, J. J. & Martin, T. F. J. Role of caps in dense-core vesicle exocytosis. *Annals of the New York Academy of Sciences* **971**, 201–209 (2002).
106. Shinoda, Y. *et al.* Caps1 stabilizes the state of readily releasable synaptic vesicles to fusion competence at ca3-ca1 synapses in adult hippocampus. *Scientific reports* **6**, 31540 (2016).
107. Farina, M. *et al.* Caps-1 promotes fusion competence of stationary dense-core vesicles in presynaptic terminals of mammalian neurons. *eLife* **4** (2015).

108. Rietveld, C. A. *et al.* Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *PNAS* **111**, 13790–13794 (2014).
109. Sun, J. *et al.* Ube3a regulates synaptic plasticity and learning and memory by controlling sk2 channel endocytosis. *Cell reports* **12**, 449–461 (2015).
110. Cook, E. H. *et al.* Autism or atypical autism in maternally but not paternally derived proximal 15q duplication. *American journal of human genetics* **60**, 928–934 (1997).
111. Lin, M. T., Luján, R., Watanabe, M., Adelman, J. P. & Maylie, J. Sk2 channel plasticity contributes to ltp at schaffer collateral-ca1 synapses. *Nature neuroscience* **11**, 170–177 (2008).
112. Hammond, R. S. *et al.* Small-conductance ca2+-activated k+ channel type 2 (sk2) modulates hippocampal learning, memory, and synaptic plasticity. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **26**, 1844–1853 (2006).
113. Murthy, S. R. K. *et al.* Small-conductance ca2+-activated potassium type 2 channels regulate the formation of contextual fear memory. *PloS one* **10**, e0127264 (2015).
114. 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. *Biological psychiatry* **75**, 105–114 (2014).
115. Goes, F. S. *et al.* Genome-wide association study of schizophrenia in ashkenazi jews. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics* **168**, 649–659 (2015).
116. Sanders, S. *et al.* Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. *Neuron* **87**, 1215–1233 (2015).
117. 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).
118. Sun, X.-J. *et al.* Genome-wide survey and developmental expression mapping of zebrafish set domain-containing genes. *PloS one* **3**, e1499 (2008).
119. Ali, M. *et al.* Molecular basis for chromatin binding and regulation of mll5. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 11296–11301 (2013).
120. Lemak, A. *et al.* Solution nmr structure and histone binding of the phd domain of human mll5. *PloS one* **8**, e77020 (2013).
121. Zhang, X., Novera, W., Zhang, Y. & Deng, L.-W. Mll5 (kmt2e): structure, function, and clinical relevance. *Cellular and molecular life sciences : CMLS* **74**, 2333–2344 (2017).
122. Torrico, B. *et al.* Lack of replication of previous autism spectrum disorder gwas hits in european populations. *Autism research : official journal of the International Society for Autism Research* **10**, 202–211 (2017).
123. Feijs, K. L. H., Forst, A. H., Verheugd, P. & Lüscher, B. Macrodomain-containing proteins: regulating new intracellular functions of mono(adp-ribosyl)ation. *Nature reviews. Molecular cell biology* **14**, 443–451 (2013).
124. Davies, G. *et al.* Genome-wide association study of cognitive functions and educational attainment in uk biobank (n=112,151). *Molecular psychiatry* **21**, 758–767 (2016).
125. Deary, V. *et al.* Genetic contributions to self-reported tiredness. *Molecular psychiatry* (2017).

Figure 1. Manhattans plots

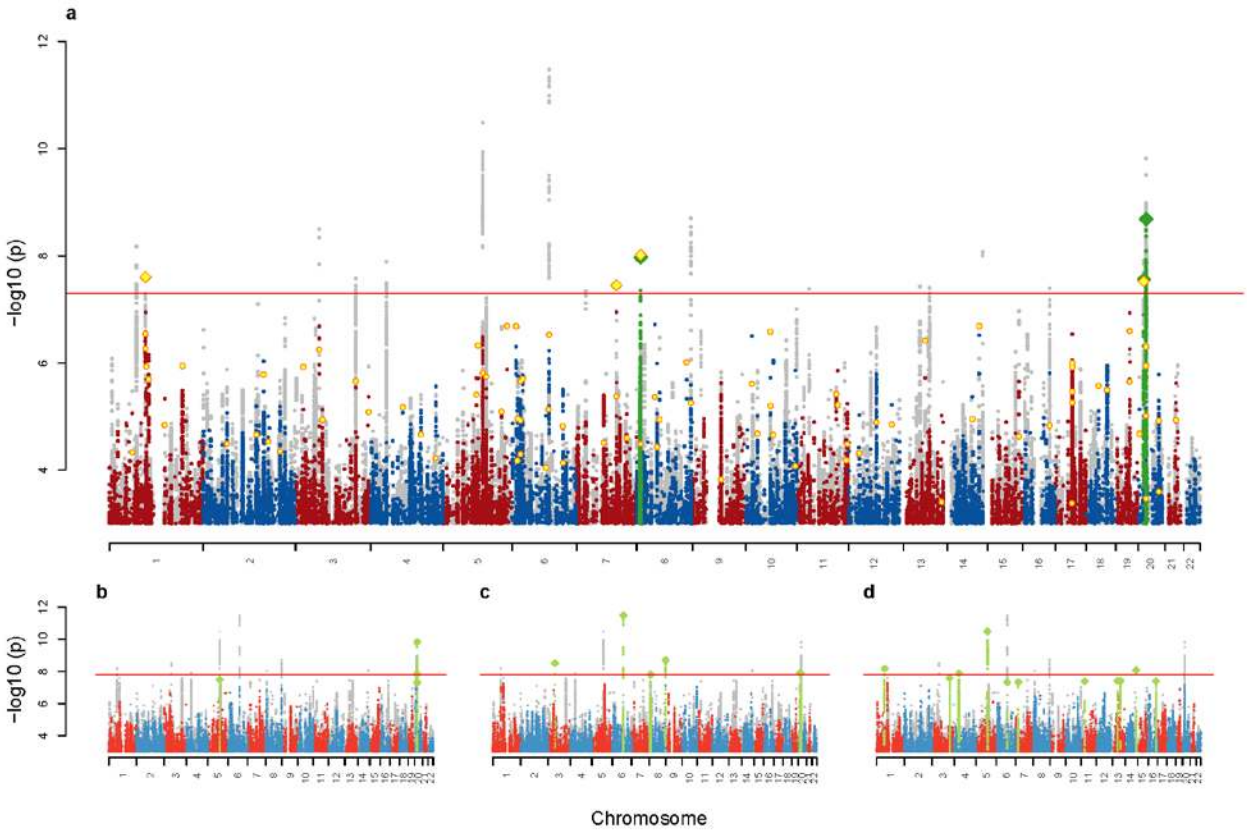

Figure 2. Genetic correlation with other traits

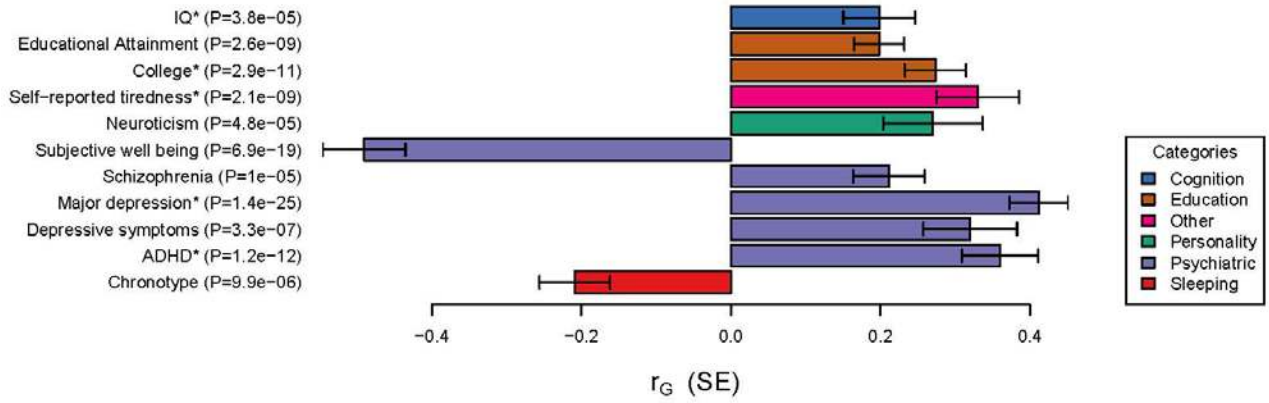

Figure 3 Profiling PRS load across distinct ASD sub-groups for 8 different phenotypes

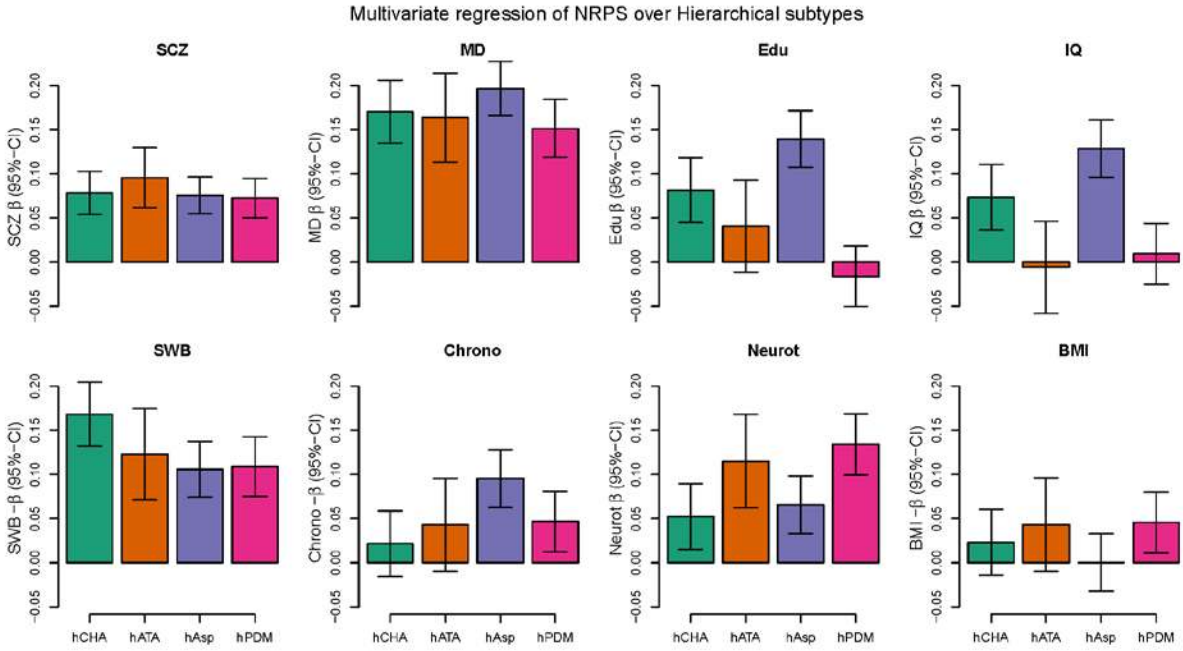

Figure 4. Decile plots

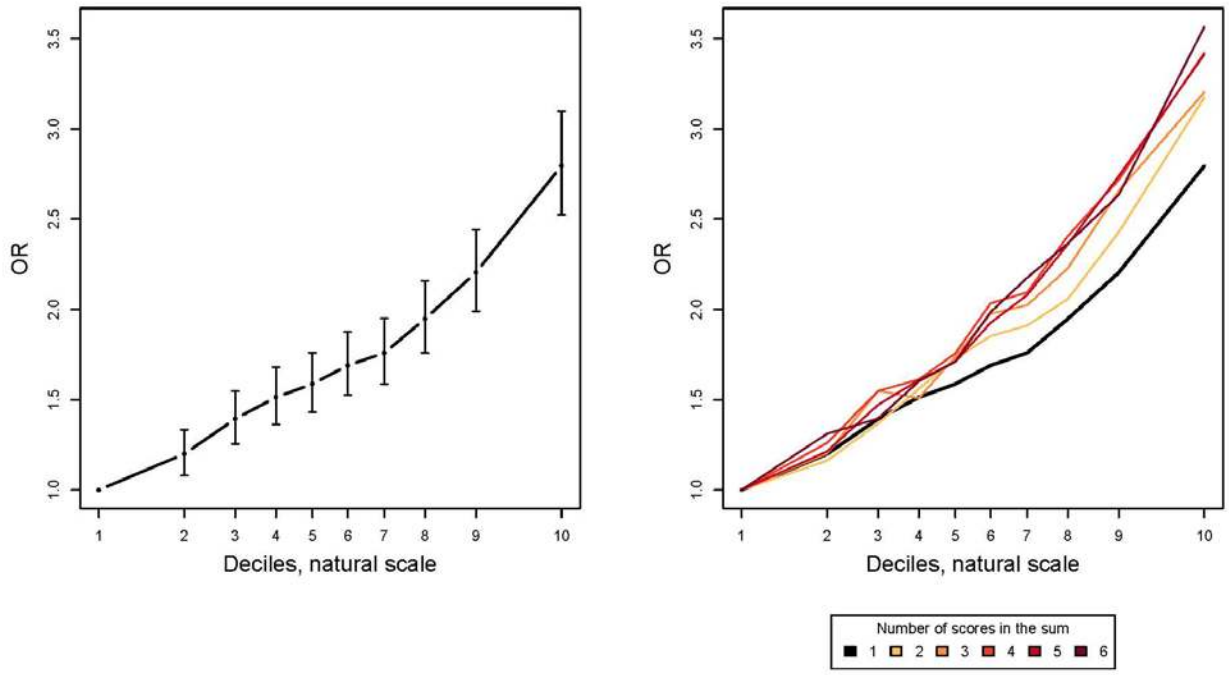

Figure 5. Chromatin interactions identify putative target genes of ASD loci