Novel genes for autism implicate both excitatory and inhibitory cell lineages in risk

F. Kyle Satterstrom[1,2,3,*], Jack A. Kosmicki[1,2,3,4,5,*], Jiebiao Wang[6,*], Michael S. Breen[7,8,9], Silvia De Rubeis[7,8,9,†], Joon-Yong An[10], Minshi Peng[6], Ryan Collins[5,11], Jakob Grove[12,13,14], Lambertus Klei[15], Christine Stevens[1,3,4,5], Jennifer Reichert[7,8], Maureen S. Mulhern[7,8], Mykyta Artomov[1,3,4,5], Sherif Gerges[1,3,4,5], Brooke Sheppard[10], Xinyi Xu[7,8], Aparna Bhaduri[16,17], Utku Norman[18], Harrison Brand[5], Grace Schwartz[10], Rachel Nguyen[19], Elizabeth E. Guerrero[20], Caroline Dias[21,22], Branko Aleksic[23], Richard Anney[24], Mafalda Barbosa[9], Somer Bishop[10], Alfredo Brusco[25,26], Jonas Bybjerg-Grauholm[12,27], Angel Carracedo[28,29], Marcus C.Y. Chan[30], Andreas G. Chiocchetti[31], Brian H. Y. Chung[30], Hilary Coon[32,33], Michael L. Cuccaro[34], Aurora Currò[35], Bernardo Dalla Bernardina[36], Ryan Doan[21], Enrico Domenici[37], Shan Dong[10], Chiara Fallerini[35], Montserrat Fernández-Prieto[28,38], Giovanni Battista Ferrero[39], Christine M. Freitag[31], Menachem Fromer[3], J. Jay Gargus[19], Daniel Geschwind[40,41], Elisa Giorgio[25], Javier González-Peñas[42], Stephen Guter[43], Danielle Halpern[7,8], Emily Hansen-Kiss[44], Xin He[45], Gail E. Herman[44], Irva Hertz-Picciotto[20], David M. Hougaard[12,27], Christina M. Hultman[46], Iuliana Ionita-Laza[47], Suma Jacob[43], Jesslyn Jamison[7,8], Astanand Jugessur[48], Miia Kaartinen[49], Gun Peggy Knudsen[48], Alexander Kolevzon[7,8,50], Itaru Kushima[23], So Lun Lee[30], Terho Lehtimäki [51], Elaine T. Lim[4], Carla Lintas[52], W. Ian Lipkin[47], Diego Lopergolo[35], Fátima Lopes[53], Yunin Ludena[20], Patricia Maciel[53], Per Magnus[48], Behrang Mahjani[7,8], Nell Maltman[43], Dara S. Manoach[54], Gal Meiri[55], Idan Menashe[56], Judith Miller[32,33], Nancy Minshew[15], Eduarda Montenegro M. de Souza[57], Danielle Moreira[57], Eric M. Morrow[58], Ole Mors[12,59], Preben Bo Mortensen[12,13,60,61], Matthew Mosconi[43], Pierandrea Muglia[62], Benjamin Neale[1,2,3], Merete Nordentoft[12,63], Norio Ozaki[23], Aarno Palotie[2,3,64,65], Mara Parellada[42], Maria Rita Passos-Bueno[57], Margaret Pericak-Vance[34], Antonio Persico[66,67], Isaac Pessah[20], Kaija Puura[49], Abraham Reichenberg[7,8,9,68], Alessandra Renieri[35,69], Evelise Riberi[39], Elise B. Robinson[1,2,3], Kaitlin E. Samocha[70], Sven Sandin[7,8,46], Susan L. Santangelo[71,72], Gerry Schellenberg[73], Stephen W. Scherer[74,75], Sabine Schlitt[31], Rebecca Schmidt[20], Lauren Schmitt[43], Isabela Maya W. Silva[57], Tarjinder Singh[1,2,3,4,5], Paige M. Siper[7,8,9], Moyra Smith[19], Gabriela Soares[76], Camilla Stoltenberg[48], Pål Suren[48], Ezra Susser[77,78], John Sweeney[43], Peter Szatmari[79], Lara Tang[7,8], Flora Tassone[20,80], Karoline Teufel[31], Elisabetta Trabetti[81], Maria del Pilar Trelles[7,8], Christopher Walsh[21], Lauren A. Weiss[10], Thomas Werge[12,82,83], Donna Werling[10], Emilie M. Wigdor[1,2,3], Emma Wilkinson[7,8], Jeremy A. Willsey[10], Tim Yu[21], Mullin H.C. Yu[30], Ryan Yuen[74], Elaine Zachi[57], iPSYCH consortium , Catalina Betancur[84,†], Edwin H. Cook[43,†], Louise Gallagher[85,†], Michael Gill[85,†], Thomas Lehner[86], Geetha Senthil[86], James S. Sutcliffe[87,88,†], Audrey Thurm[86], Michael E. Zwick[89,†], Anders D. Børglum[12,13,14,90], Matthew W. State[10], A. Ercument Cicek[6,18], Michael E. Talkowski[5], David J. Cutler[89], Bernie Devlin[15,†,#], Stephan J. Sanders[10,#,§], Kathryn Roeder[6,91,†,#,§], Joseph D. Buxbaum[7,8,9,92,93,94,†,#,§], Mark J. Daly[1,2,3,4,5,64,†,#,§]

[1]Stanley Center for Psychiatric Research, Broad Institute, Cambridge, Massachusetts, USA. [2]Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA. [3]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. [4]Harvard Medical School, Boston, Massachusetts, USA. [5]Center for Genomic Medicine, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA. [6]Department of Statistics,

Carnegie Mellon University, Pittsburgh, Pennsylvania, USA. [7]Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, New York, USA. [8]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York, USA. [9]The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, New York, USA. [10]Department of Psychiatry, UCSF Weill Institute for Neurosciences, University of California San Francisco, San Francisco, California, USA. [11]Program in Bioinformatics and Integrative Genomics, Harvard Medical School, Boston, Massachusetts, USA. [12]The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Denmark. [13]Center for Genomics and Personalized Medicine, Aarhus, Denmark. [14]Department of Biomedicine - Human Genetics, Aarhus University, Aarhus, Denmark. [15]Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA. [16]Department of Neurology, University of California San Francisco, San Francisco, California, USA. [17]The Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research, University of California San Francisco, San Francisco, California, USA. [18]Computer Engineering Department, Bilkent University, Ankara, Turkey. [19]Center for Autism Research and Translation, University of California Irvine, Irvine, California, USA. [20]MIND (Medical Investigation of Neurodevelopmental Disorders) Institute, University of California Davis, Davis, California, USA. [21]Division of Genetics, Boston Children's Hospital, Boston, Massachusetts, USA. [22]Division of Developmental Medicine, Boston Children's Hospital, Boston, Massachusetts, USA. [23]Department of Psychiatry, Graduate School of Medicine, Nagoya University, Nagoya, Japan. [24]Division of Psychological Medicine & Clinical Neurosciences, MRC Centre for Neuropsychiatric Genetics & Genomics, Cardiff University, Cardiff, United Kingdom. [25]Department of Medical Sciences, University of Torino, Turin, Italy. [26]Medical Genetics Unit, Città della Salute e della Scienza Hospital, Turin, Italy. [27]Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark. [28]Grupo de Medicina Xenómica, Centro de Investigaciónen en Red de Enfermedades Raras (CIBERER), Universidade de Santiago de Compostela, Santiago de Compostela, Spain. [29]Fundación Pública Galega de Medicina Xenómica, Servicio Galego de Saúde (SERGAS), Santiago de Compostela, Spain. [30]Department of Pediatrics & Adolescent Medicine, Duchess of Kent Children's Hospital, The University of Hong Kong, Hong Kong Special Administrative Region, China. [31]Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, Goethe University Frankfurt, Frankfurt, Germany. [32]Department of Internal Medicine, University of Utah, Salt Lake City, Utah, USA. [33]Department of Psychiatry, University of Utah, Salt Lake City, Utah, USA. [34]John P Hussman Institute for Human Genomics, University of Miami, Miami, USA. [35]Medical Genetics, University of Siena, Siena, Italy. [36]Child Neuropsychiatry Department, Department of Surgical Sciences, Dentistry, Gynecology and Pediatrics. University of Verona, Verona, Italy. [37]Centre for Integrative Biology, University of Trento, Trento, Italy. [38]Neurogenetics Group, Instituto de Investigación Sanitaria de Santiago (IDIS-SERGAS), Santiago de Compostela, Spain. [39]Department of Public Health and Pediatrics, University of Torino, Turin, Italy. [40]Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California, USA. [41]Center for Autism Research and Treatment and Program in Neurobehavioral Genetics, Semel Institute, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California, USA. [42]Department of Child and Adolescent Psychiatry, Hospital General Universitario Gregorio Marañón, IiSGM, CIBERSAM, School of Medicine Complutense University, Madrid, Spain. [43]Institute for Juvenile Research, Department of Psychiatry, University of Illinois at Chicago, Chicago, Illinois, USA. [44]The Research Institute at Nationwide Children's Hospital, Columbus, Ohio, USA. [45]Department of Human Genetics, University of Chicago, Chicago, Illinois, USA. [46]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [47]Department of Biostatistics, Columbia University, New York, New York,

USA. [48]Norwegian Institute of Public Health, Oslo, Norway. [49]Department of Child Psychiatry, University of Tampere and Tampere University Hospital, Tampere, Finland. [50]Department of Pediatrics, Icahn School of Medicine at Mount Sinai, New York, New York, USA. [51]Department of Clinical Chemistry, Fimlab Laboratories and Finnish Cardiovascular Research Center-Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland. [52]Campus Bio-medico, Rome, Italy. [53]Life and Health Sciences Research Institute, School of Medicine, University of Minho, Campus de Gualtar, Braga, Portugal. [54]Department of Psychiatry, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA. [55]Pre-School Psychiatry Unit, Soroka University Medical Center, Beer Sheva, Israel. [56]Department of Public Health, Ben-Gurion University of the Negev, Beer-Sheva, Israel. [57]Centro de Pesquisas sobre o Genoma Humano e Células tronco, Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil. [58]Department of Molecular Biology, Cell Biology and Biochemistry and Department of Psychiatry and Human Behavior, Brown University, Providence, Rhode Island, USA. [59]Psychosis Research Unit, Aarhus University Hospital, Risskov, Denmark. [60]National Centre for Register-Based Research, Aarhus University, Aarhus, Denmark. [61]Centre for Integrated Register-based Research, Aarhus University, Aarhus, Denmark. [62]UCB Pharma, Belgium. [63]Mental Health Services in the Capital Region of Denmark, Mental Health Center Copenhagen, University of Copenhagen, Copenhagen, Denmark. [64]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland. [65]Psychiatric & Neurodevelopmental Genetics Unit, Department of Psychiatry, Massachusetts General Hospital, Boston, Massachusetts, USA. [66]Interdepartmental Program "Autism 0-90", Gaetano Martino University Hospital, University of Messina, Messina, Italy. [67]Mafalda Luce Center for Pervasive Developmental Disorders, Milan, Italy. [68]Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, New York, USA. [69]Genetica Medica, Azienda Ospedaliera Universitaria Senese, Siena, Italy. [70]The Wellcome Trust Sanger Institute, Cambridge, United Kingdom. [71]Maine Medical Center and Maine Medical Center Research Institute, Portland, Maine, USA. [72]Department of Psychiatry, Tufts University School of Medicine, Boston, Massachusetts, USA. [73]Department of Pathology and Laboratory Medicine, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA. [74]Program in Genetics and Genome Biology, The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Ontario, Canada. [75]McLaughlin Centre, University of Toronto, Toronto, Ontario, Canada. [76]Center for Medical Genetics Dr. Jacinto Magalhães, National Health Institute Dr. Ricardo Jorge, Porto, Portugal. [77]New York State Psychiatric Institute, New York, New York, USA. [78]Department of Epidemiology, Mailman School of Public Health, Columbia University, Epidemiology, Mailman School of Public Health, Columbia University, New York, New York, USA. [79]Department of Psychiatry and Behavioural Neurosciences, Offord Centre for Child Studies, McMaster University, Hamilton, Ontario, Canada. [80]Department of Biochemistry and Molecular Medicine, University of California Davis, School of Medicine, Davis, CA, USA. [81]Department of Neurosciences, Biomedicine and Movement Sciences, Section of Biology and Genetics, University of Verona, Verona, Italy. [82]Institute of Biological Psychiatry, MHC Sct. Hans, Mental Health Services Copenhagen, Roskilde, Denmark. [83]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark. [84]Sorbonne Université, INSERM, CNRS, Neuroscience Paris Seine, Institut de Biologie Paris Seine, Paris, France. [85]Department of Psychiatry, School of Medicine, Trinity College Dublin, Dublin, Ireland. [86]National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland, USA. [87]Vanderbilt Genetics Institute, Vanderbilt University School of Medicine, Nashville, Tennessee, USA. [88]Department of Molecular Physiology and Biophsysics and Psychiatry, Vanderbilt University School of Medicine, Nashville, Tennessee, USA. [89]Department of Human Genetics, Emory University School of Medicine, Atlanta, Georgia, USA. [90]Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark. [91]Computational Biology Department, Carnegie Mellon University, Pittsburgh,

Pennsylvania, USA. [92]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York, USA. [93]Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, New York, USA. [94]Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, New York, USA.

* Equal contribution
§ Equal contribution
† Lead member of standing committee of the ASC
# ASC Principal Investigators

Abstract

We present the largest exome sequencing study to date focused on rare variation in autism spectrum disorder (ASD) (n=35,584). Integrating *de novo* and case-control variation with an enhanced Bayesian framework incorporating evolutionary constraint against mutation, we implicate 99 genes in ASD risk at a false discovery rate (FDR) ≤ 0.1. Of these 99 risk genes, 46 show higher frequencies of disruptive *de novo* variants in individuals ascertained for severe neurodevelopmental delay, while 50 show higher frequencies in individuals ascertained for ASD, and comparing ASD cases with disruptive mutations in the two groups shows differences in phenotypic presentation. Expressed early in brain development, most of the risk genes have roles in neuronal communication or regulation of gene expression, and 12 fall within recurrent copy number variant loci. In human cortex single-cell gene expression data, expression of the 99 risk genes is also enriched in both excitatory and inhibitory neuronal lineages, implying that disruption of these genes alters the development of both neuron types. Together, these insights broaden our understanding of the neurobiology of ASD.

Introduction

Autism spectrum disorder (ASD), a childhood-onset neurodevelopmental condition characterized by deficits in social communication and restricted, repetitive patterns of behavior or interests (*1*), affects more than 1% of children (*2*). Multiple studies have demonstrated high heritability, indicating that genetic factors play an important, causal role (*3*). Although common genetic variants, which are present to a greater or lesser degree in everyone, account for the majority of the observed heritability (*4*), rare inherited variants and newly arising, or *de novo*, mutations are major contributors to individual risk (*5-14*). When this rare variation disrupts a gene in individuals with ASD more often than expected by chance, it implicates that gene in risk (*5, 10, 11, 15, 16*). Such genes, in turn, can provide insight into the atypical neurodevelopment underlying ASD, both individually (*17, 18*) and *en masse* (*5, 10, 19*). Fundamental questions about the nature of this disrupted neurobiological development – including when it occurs, where, and in what cell types – remain unanswered. Here we present the largest exome sequencing study in ASD to date, greatly expanding the list of genes significantly associated with ASD, and combine these results with functional genomic data to gain novel insights into the neurobiology of ASD.

Building on previous Autism Sequencing Consortium (ASC) studies (*5, 10, 20*), we analyze 35,584 samples, including 11,986 ASD cases split almost evenly between family-based cohorts (6,430 cases ["probands"] with both parents sequenced, enabling *de novo* mutations to be detected) and case-control cohorts (5,556 cases with 8,809 ancestry-matched controls). We introduce an enhanced Bayesian analytic framework, which leverages recently developed gene- and variant-level scores of evolutionary constraint of genetic variation, to implicate genes in

ASD more rigorously than previous studies. In this way, we identify 99 genes likely to play a role in ASD risk (false discovery rate [FDR] ≤ 0.1) and confirm that they are strongly enriched for genes involved in gene expression regulation (GER) or neuronal communication (NC). Furthermore, by analysis of extant gene expression data, we show that many of the GER genes are expressed in multiple tissues throughout the body and reach a peak of cortical expression in early fetal development, whereas many NC genes are expressed predominantly in the brain and reach a peak of cortical expression in late fetal and perinatal development. Considering data from single cells in the developing human cortex, most of the 99 ASD genes are highly expressed from midfetal development onwards, and both the GER and NC sets are enriched in maturing and mature excitatory and inhibitory neurons.

The symptoms of ASD often occur in tandem with comorbidities. In at least a third of individuals, ASD is one of a constellation of symptoms of neurodevelopmental delay (NDD), alongside intellectual disability (*2*) and motor impairments (*21*). Unsurprisingly, many ASD genes are also associated with NDD (*22-26*). By comparing disruptive *de novo* variants in our study to those from NDD cohorts, we split the 99 genes into those with a higher frequency in ASD-ascertained subjects ("ASD-predominant" or "ASD_P") and those with a higher frequency in NDD-ascertained subjects ("ASD_{NDD}"). We show that disruptive variants in ASD_{NDD} genes result in higher rates of neurodevelopmental comorbidities even in ASD-ascertained subjects, suggesting extreme selective pressure, while disruptive variants in ASD_P genes yield phenotypes closer to ASD cases without ASD-associated variants, suggesting more modest selective pressures. These distinctions suggest complex genotype-phenotype correlations across neurodevelopmental domains, similar to those observed across tissues in well-defined genetic syndromes.

Results

Data generation and quality control

Our primary goal is to associate genes with risk for ASD by examining the distribution of genetic variation found in them. To do this, we integrated whole-exome sequence (WES) data from several sources. After reported family structures were verified and stringent filters were applied for sample, genotype, and variant quality, we included 35,584 samples (11,986 ASD cases) in our analyses. These WES data included 21,219 family-based samples (6,430 ASD cases, 2,179 sibling controls, and both of their parents) and 14,365 case-control samples (5,556 ASD cases, 8,809 controls) (Fig. S1; Table S1). Read-level WES data were processed for 24,022 samples (67.5%), including 6,197 newly sequenced ASC samples, using BWA (*27*) to perform alignment and GATK (*28*) to perform joint variant calling (Fig. S1). These data were integrated with variant- or gene-level counts from an additional 11,562 samples (Fig. S1), including 10,025 samples from the Danish iPSYCH study which our consortium had not previously incorporated (*29*).

From this cohort, we identified a set of 10,552 rare *de novo* variants in protein-coding exons (allele frequency ≤ 0.1% in our dataset as well as the non-psychiatric subsets of the reference databases ExAC and gnomAD (*30*)), with 70% of probands and 67% of unaffected offspring carrying at least one *de novo* variant (4,521 out of 6,430 and 1,468 out of 2,179, respectively; Table S2; Fig. S1). For rare inherited and case-control variant analyses, we included variants with an allele count no greater than five in our dataset and in the non-psychiatric subset of ExAC

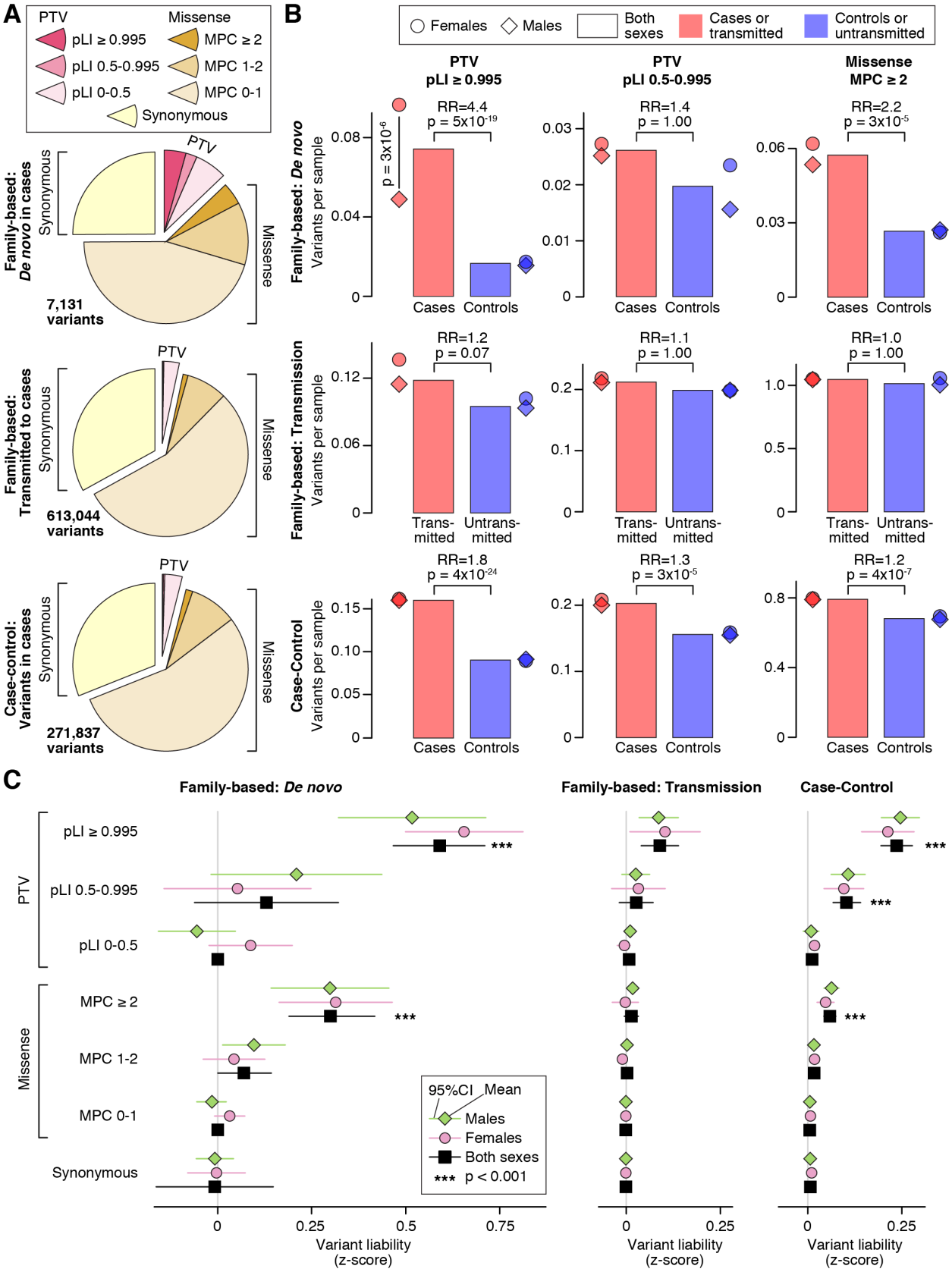
(*30, 31*). Analyses of inherited variation use only the family-based data, specifically comparing variants that were transmitted or untransmitted from parents to their affected offspring.

Impact of genetic variants on ASD risk

Exonic variants can be divided into groups based on their predicted functional impact. For any such group, the differential burden of variants carried by cases versus controls reflects the average liability that these variants impart for ASD. This ASD liability, along with the mutation rate per gene, can be used to determine the number of mutations required to demonstrate ASD association for a specific gene (*5, 10, 11*). For example, because protein-truncating variants (PTVs, consisting of nonsense, frameshift, and essential splice site variants) show a much greater difference in burden between ASD cases and controls than missense variants, their average impact on liability must be larger (*15*). Recent analyses have shown that additional measures of functional severity, such as the "probability of loss-of-function intolerance" (pLI) score (*30, 31*) and the integrated "missense badness, PolyPhen-2, constraint" (MPC) score (*32*), can further delineate specific variant classes with a higher burden in ASD cases.

We divided the list of rare autosomal genetic variants into seven tiers of predicted functional severity. Three tiers for PTVs by pLI score (≥0.995, 0.5-0.995, 0-0.5), in order of decreasing expected impact on liability; three tiers for missense variants by MPC score (≥2, 1-2, 0-1), also in order of decreasing impact; and a single tier for synonymous variants, which should have minimal impact on liability. We also divided the variants into three bins by their inheritance pattern: *de novo*, inherited, and case-control, with the latter reflecting a mixture of *de novo* and inherited variants that cannot be distinguished directly without parental data. Unlike inherited variants, newly arising *de novo* mutations are exposed to minimal selective pressure and,

accordingly, have the potential to mediate substantial risk to severe disorders that limit fecundity, such as ASD (*33*). This expectation is borne out by the substantially higher proportions of all three PTV tiers and the two most severe missense variant tiers in *de novo* variants compared to inherited variants (Fig. 1A). *De novo* mutations are also extremely rare, with 1.23 variants per subject distributed over the 17,484 genes assessed, so the overall proportions of variants in the case-control data are similar to those of inherited variants (Fig. 1A).


A
PTV
pLI ≥ 0.995
pLI 0.5-0.995
pLI 0-0.5
Missense
MPC ≥ 2
MPC 1-2
MPC 0-1
Synonymous
Family-based: De novo in cases
Synonymous
PTV
Missense
7,131 variants
Family-based: Transmitted to cases
613,044 variants
Case-control: Variants in cases
271,837 variants
B
Females
Males
Both sexes
Cases or transmitted
Controls or untransmitted
PTV pLI ≥ 0.995
PTV pLI 0.5-0.995
Missense MPC ≥ 2
Family-based: De novo
Variants per sample
p = 3x10-6
RR=4.4
p = 5x10-19
RR=1.4
p = 1.00
RR=2.2
p = 3x10-5
Cases
Controls
Family-based: Transmission
RR=1.2
p = 0.07
RR=1.1
p = 1.00
RR=1.0
p = 1.00
Trans-mitted
Untrans-mitted
Case-Control
RR=1.8
p = 4x10-24
RR=1.3
p = 3x10-5
RR=1.2
p = 4x10-7
C
Family-based: De novo
Family-based: Transmission
Case-Control
PTV
pLI ≥ 0.995
pLI 0.5-0.995
pLI 0-0.5
Missense
MPC ≥ 2
MPC 1-2
MPC 0-1
Synonymous
95%CI
Mean
Males
Females
Both sexes
*** p < 0.001
Variant liability (z-score)

1

Figure 1. Distribution of rare autosomal protein-coding variants in ASD cases and controls. *__A__, The proportion of rare autosomal genetic variants split by predicted functional consequences, represented by color, is displayed for family-based data (split into de novo and inherited variants) and case-control data. PTVs and missense variants are split into three tiers of predicted functional severity, represented by shade, based on the pLI and MPC metrics, respectively. __B__, The relative difference in variant frequency (i.e. burden) between ASD cases and controls (top and bottom) or transmitted and untransmitted parental variants (middle) is shown for the top two tiers of functional severity for PTVs (left and center) and the top tier of functional severity for missense variants (right). Next to the bar plot, the same data are shown divided by sex. __C__, The relative difference in variant frequency shown in 'B' is converted to a trait liability z-score, split by the same subsets used in 'A'. For context, a z-score of 2.18 would shift an individual from the population mean to the top 1.69% of the population (equivalent to an ASD threshold based on 1 in 68 children (34)). No significant difference in liability was observed between males and females for any analysis. Statistical tests: B, C: Binomial Exact Test (BET) for most contrasts; exceptions were "both" and "case-control", for which Fisher's method for combining BET p-values for each sex and, for case-control, each population, was used; p-values corrected for 168 tests are shown. Abbreviations: PTV: protein-truncating variant; pLI: probability loss-of-function intolerant; MPC: missense badness, PolyPhen-2, and constraint; RR: relative risk.*

Comparing affected probands to unaffected siblings, we observe a 4.4-fold enrichment for *de novo* PTVs in the 1,447 autosomal genes with a pLI ≥ 0.995 (366 in 6,430 cases versus 36 in 2,179 controls; 0.074 vs. 0.017 variants per sample (vps); $p=5x10^{-19}$; Fig. 1B). A less pronounced difference in burden is observed for rare inherited PTVs in these genes, with a 1.2-fold enrichment for transmitted versus untransmitted alleles (695 transmitted versus 557 untransmitted in 5,869 parents; 0.12 vs. 0.10 vps; p=0.07; Fig. 1B). The relative burden in the case-control data falls between the estimates for *de novo* and inherited data in these most severe PTVs, with a 1.8-fold enrichment in cases versus controls (874 in 5,556 cases versus 759 in 8,809 controls; 0.16 vs. 0.09 vps; $p=4x10^{-24}$; Fig. 1B). Analysis of the middle tier of PTVs (0.5 ≤ pLI < 0.995) shows a similar, but muted, pattern (Fig. 1B), while the lowest tier of PTVs (pLI < 0.5) shows no case enrichment (Table S3).

De novo missense variants are observed more frequently than *de novo* PTVs and, *en masse*, they show only marginal enrichment over the rate expected by chance (*5*) (Fig. 1). However, the most

severe missense variants (MPC ≥ 2) occur at a similar frequency to *de novo* PTVs, and we observe a 2.2-fold case enrichment (354 in 6,430 cases versus 58 in 2,179 controls; 0.057 vs. 0.027 vps; $p=3x10^{-5}$; Fig. 1B), with a consistent 1.2-fold enrichment in the case-control data (4,277 in 5,556 cases versus 6,149 in 8,809 controls; 0.80 vs. 0.68 vps; $p=4x10^{-7}$; Fig. 1B). Of note, this top tier of missense variation shows stronger enrichment in cases than the middle tier of PTVs. Consistent with prior expectations, the other two tiers of missense variation were not enriched in cases (Table S3).

Sex differences in ASD risk

The prevalence of ASD is consistently higher in males than females, usually by a factor of three or more (*2*). Females diagnosed with ASD carry a higher burden of genetic risk factors, including *de novo* copy number variants (CNVs) (*9, 10*), *de novo* PTVs (*5, 31*), and *de novo* missense variants (*5*). Here we observe a similar result, with a 2-fold enrichment of *de novo* PTVs in highly constrained genes in affected females versus affected males ($p=3x10^{-6}$) and similar non-significant trends in other categories with large differences between cases and controls (Fig. 1B; Table S3). The excess of genetic risk we observe in females is consistent with a model dubbed the female protective effect (FPE) that postulates females being more resilient to ASD and consequently requiring an increased genetic load (in this case, deleterious variants of larger effect) to reach the threshold for a diagnosis (*35, 36*). The converse hypothesis is that risk variation has larger effects in males than in females so that females require a higher genetic burden to reach the same diagnostic threshold as males.

To discern between these two possibilities, we assessed the ASD trait liability in males and females using sex-specific estimates of ASD prevalence (*34*). Relative to the general population,

ASD can be conceptualized as the extreme tail of a normally distributed quantitative trait, termed "liability," with individuals who cross a liability threshold receiving the diagnostic label of ASD. The threshold is determined by ASD prevalence, estimated at 2.38% in males and 0.53% in females (*34*). Using this model and the relative burden of variants in cases and controls (Table S3), we can estimate the impact that different classes of genetic variants would have on liability (Supplemental Online Methods (SOM)) and, in theory, all sources of risk can be calibrated to this common metric. For context, the observed ASD prevalence maps onto a trait liability threshold with a z-score of 1.98 in males and 2.56 in females. Across all classes of genetic variants, we observed no significant sex differences in trait liability, consistent with the FPE model (Fig. 1C).

Differences in ASD liability

In the absence of sex-specific differences in liability, we estimated the liability across both sexes together. PTVs in any of the 1,447 genes with a pLI $\geq$ 0.995 have a liability z-score of 0.59 when *de novo*, compared to 0.24 in case-control populations and 0.09 for inherited variants (Fig. 1C; Table S3). These liability z-scores, reflecting a higher ratio of true ASD risk variants to variants with minimal or neutral impact on ASD risk in *de novo* variants compared to the other two groups, can be leveraged to enhance gene discovery.

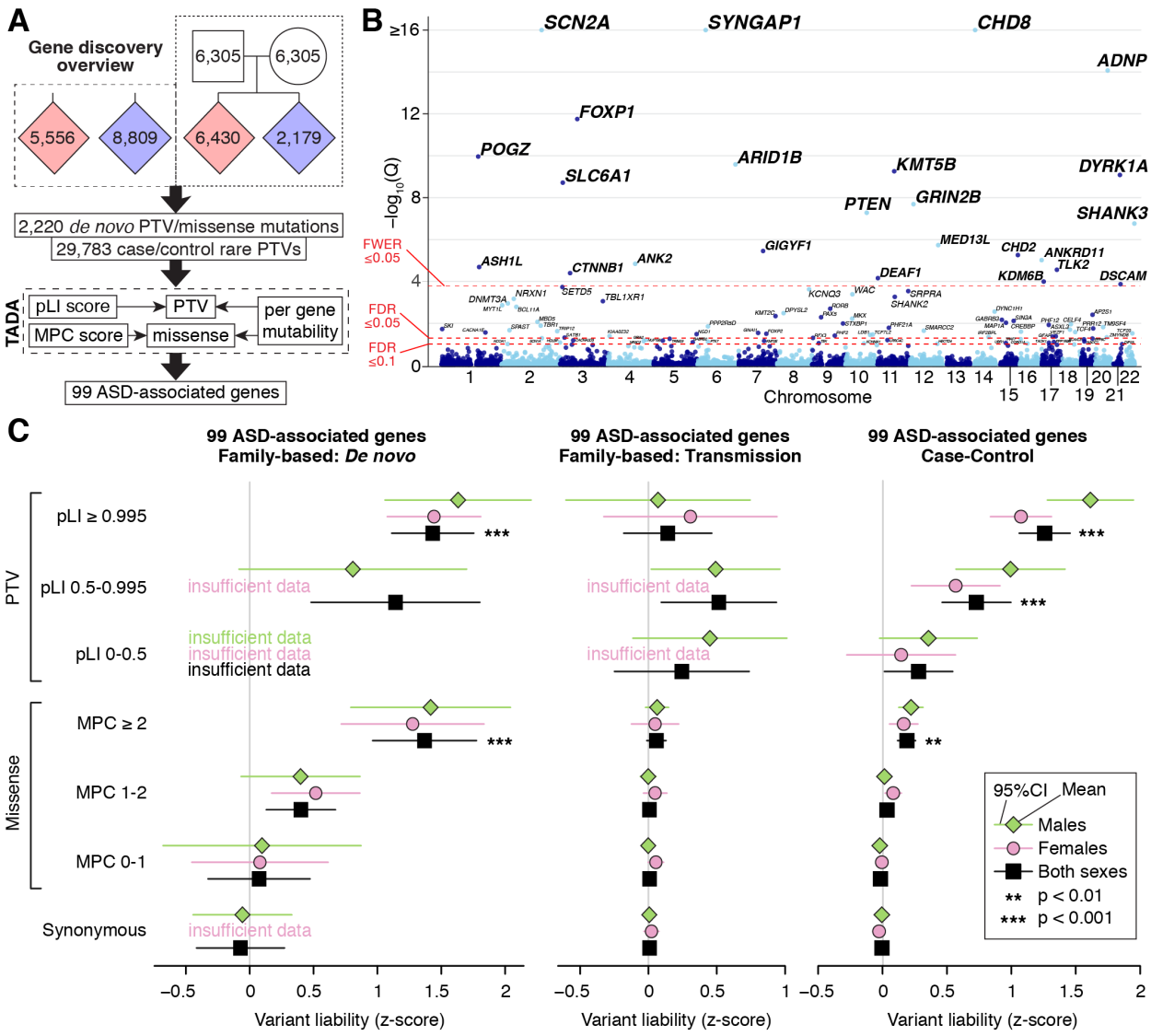


Figure 2. Gene discovery in the ASC cohort. A*, An overview of gene discovery. Whole exome sequencing data from 35,584 samples is entered into a Bayesian analysis framework (TADA) that incorporates pLI score for PTVs and MPC score for missense variants.* ***B****, The model identifies 99 autosomal genes associated with ASD at a false discovery rate (FDR) threshold of ≤ 0.1, which is shown on the y-axis of this Manhattan plot with each point representing a gene. Of these, 75 exceed the threshold of FDR ≤ 0.05 and 25 exceed the threshold family-wise error rate (FWER) ≤ 0.05.* ***C****, Repeating our ASD trait liability analysis (Fig. 1C) restricted to variants observed within the 99 ASD-associated genes only. Statistical tests: B, TADA; C, Binomial Exact Test (BET) for most contrasts; exceptions were "both" and "case-control", for which Fisher's method for combining BET p-values for each sex and, for case-control, each population, was used; p-values corrected for 168 tests are shown. Abbreviations: PTV: protein-truncating variant; pLI: probability loss-of-function intolerant; MPC: missense badness, PolyPhen-2, and constraint.*

ASD gene discovery

An ASD-associated gene can be identified by an excess of variants in affected individuals compared to the expected count, which can be based on the per-gene mutation rates and sample size for *de novo* mutations or the relative frequency of classes of variants in controls. The average risk carried by variants of a particular type (e.g., PTVs) is conveyed by the relative liabilities (Fig. 1). For our earlier published work, we used the Transmitted And *De novo* Association (TADA) model (*15*) to integrate missense and PTVs that are *de novo*, inherited, or from case-control populations to stratify autosomal genes by FDR for association (*5, 10*). Here, we update the TADA model to include pLI score as a continuous metric for PTVs and MPC score in two tiers (≥2, 1-2) for missense variants (Supplemental Methods and Fig. S2, S3). In family data we include *de novo* PTVs as well as *de novo* missense variants in the model, while for case-control we include only PTVs, which show the largest liability; we do not include inherited variants due to the limited liabilities observed (Fig. 1C). These modifications result in an enhanced TADA model that has greater sensitivity and accuracy than the original model (Supplementary Methods).

Considering only *de novo* variants observed in WES data from our previous publication (*10*), the original TADA model identifies 31 genes at FDR ≤ 0.1. Keeping this FDR threshold constant, applying the original TADA model to the *de novo* variants of the new ASC cohort of 35,584 samples identifies 65 ASD-associated genes. Integrating the pLI and MPC scores into the enhanced TADA model boosts this to 85 genes. Finally, integrating the case-control data identifies 99 ASD-associated genes at FDR ≤ 0.1, of which 75 meet the more stringent threshold of FDR ≤ 0.05, while 25 are significant after Bonferroni correction (Fig. 2B; Table S4). Three additional genes reach FDR ≤ 0.1 (*KDM5B*, *RAI1*, and *EIF3G*) but are excluded from our high

confidence lists because they demonstrate an excess of *de novo* PTVs in unaffected siblings, suggesting the possibility that the mutational model may underestimate their true mutation rate (Supplementary Methods). Of note, however, heterozygous loss of *RAI1* expression is known to cause the neurodevelopmental disorder Smith-Magenis syndrome (*37*).

By simulation experiments (described in the Supplementary Methods), we demonstrate the reliable performance of the refined TADA model, in particular showing that our risk gene list, with FDR ≤ 0.1, is properly calibrated (Fig. S2). Of the 99 ASD-associated genes, 58 were not discovered by our earlier analyses. The patterns of liability seen for these 99 genes are similar to that seen over all genes (compare Fig. 2C versus Fig. 1C), although the effects of variants are uniformly larger, as would be expected for this selected list of putative risk genes that would be enriched for true risk variants. Note that, in keeping with the theory underlying the "winner's curse," we would expect liability to be overestimated for some of these genes, specifically those with the least evidence for association.

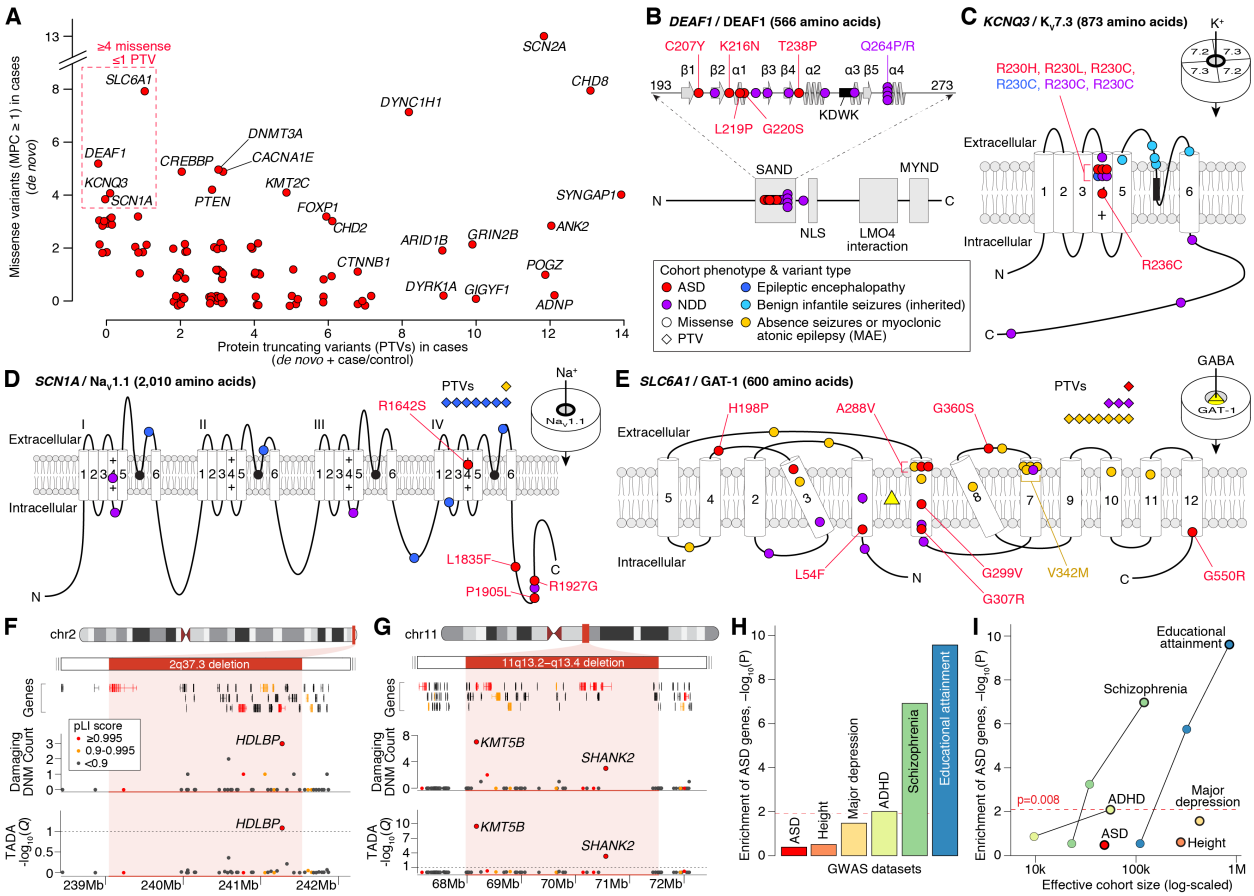


Figure 3. Genetic characterization of ASD genes. ***A****, Count of PTVs versus missense variants (MPC ≥ 1) in cases for each ASD-associated gene (red points, selected genes labeled). These counts reflect the data used by TADA for association analysis: de novo and case/control data for PTVs; de novo only for missense.* ***B****, Location of ASD de novo missense variants in the DEAF1 transcription factor. The five ASD mutations (marked in red) are in the SAND DNA-binding domain (amino acids 193 to 273, spirals show alpha helices, arrows show beta sheets, KDWK is the DNA binding-motif (38)) alongside ten variants observed in NDD, several of which have been shown to reduce DNA binding, including Q264P and Q264R (39-41).* ***C****, The location of four ASD missense variants are shown in the gene KCNQ3, which encodes the protein $K_V7.3$, which forms a neuronal voltage-gated potassium channel in combination with KCNQ2/$K_V7.2$. All four ASD variants were located in the voltage sensor (fourth of six transmembrane domains), with three in the same residue (R230), including the gain-of-function R230C mutation (42) observed in NDD (41). Five inherited variants observed in benign infantile seizures are shown in the pore loop (43, 44).* ***D****, The location of four ASD missense variants in SCN1A, which encodes the voltage-gated sodium channel $Na_V1.1$, alongside 17 de novo variants in NDD and epilepsy (41).* ***E****, The location of eight ASD missense variants in SLC6A1, which encodes the neuronal GABA-transporter GAT-1, alongside 31 de novo variants in NDD and epilepsy (41, 45).* ***F****, Subtelomeric 2q37 deletions are associated with facial dysmorphisms, brachydactyly, high BMI, neurodevelopmental delay, and ASD (46). While three genes within the locus have a pLI score ≥ 0.995, only HDLBP, which encodes an RNA-binding protein, is associated with ASD.* ***G****, Deletions at the 11q13.2−q13.4 locus have been observed in NDD, ASD, and otodental dysplasia (47-49). Five genes within the locus have a pLI score ≥ 0.995, including two ASD-*

associated genes: KMT5B and SHANK2. ***H****, Assessment of gene-based enrichment, via MAGMA, of 99 ASD-associated genes against genome-wide significant common variants from six genome-wide association studies (GWAS).* ***I****, Gene-based enrichment of 99 ASD-associated genes in multiple GWAS as a function of effective cohort size. The GWAS used for each disorder in 'I' has a black outline. Statistical tests: F, G, TADA; H, I, MAGMA. Abbreviations: ADHD: attention deficit hyperactivity disorder, C: C-terminus; LMO4: LIM domain only 4; MYND: Myeloid translocation protein 8, Nervy, and DEAF1; N: N-terminus; NDD: neurodevelopmental delay; NLS: Nuclear localization signal; SAND: Sp-100, AIRE, NucP41/ 75, and DEAF1.*

Patterns of mutations in ASD genes

Within the set of observed mutations, the ratio of PTVs to missense mutations varies substantially between genes (Fig. 3A). Some genes, such as *ADNP*, reach our association threshold through PTVs alone, and three genes stand out as having an excess of PTVs, relative to missense mutations, based on gene mutability: *SYNGAP1*, *DYRK1A*, and *ARID1B* (binomial test, $p < 0.0005$). Because of the increased cohort size, availability of the MPC metric, and integration of these into the enhanced TADA model, we are able for the first time to associate genes with ASD based on *de novo* missense variation alone, as in the case of *DEAF1*. While we expect PTVs to act primarily through haploinsufficiency, missense variants can both reduce or alter gene function, often referred to as loss-of-function and gain-of-function, respectively. When missense variants cluster in protein domains, they can provide insight into the direction of functional effect and reveal genotype-phenotype correlations (*5, 17*). We therefore considered the location of variants within four genes with four or more *de novo* missense variants and one or no PTVs (Fig. 3A; Table S5).

We observe five *de novo* missense variants and no PTVs in *DEAF1*, which encodes a self-dimerizing transcription factor involved in neuronal differentiation (*38*). Consistent with the idea that ASD risk from *DEAF1* is primarily mediated by missense variation, multiple PTVs are present in *DEAF1* in the ExAC control population (*30*), resulting in a pLI score of 0 and indicating that heterozygous PTVs are likely benign. All five missense variants are in the SAND

domain (Fig. 3B), which is critical for both dimerization and DNA binding (*38, 50*). A similar pattern of SAND domain missense enrichment and no PTVs is observed in individuals with intellectual disability, speech delay, and behavioral abnormalities (*39-41*). Functional analyses of several SAND domain missense variants reported reduced DNA binding (*39, 51*) rather than gain-of-function effects, although given that haploinsufficiency via PTVs does not appear to phenocopy this result, there may be an unforeseen gain of function or dominant negative impact.

Four *de novo* missense variants and no PTVs are observed in the gene *KCNQ3*, which encodes the $K_V7.3$ subunits of a neuronal voltage-gated potassium channel. All four cases have comorbid intellectual disability. The $K_V7.2$ subunits are encoded by the gene *KCNQ2*, with four $K_V7.2$ or $K_V7.3$ subunits forming a channel (Fig. 3C). This family of potassium channels is responsible for the M-current, which reduces neuronal excitability following action potentials. Loss-of-function missense variants in both *KCNQ2* and *KCNQ3* are associated with benign familial neonatal epilepsy (BFNE), while gain-of-function variants with persistent current have been associated with NDD and/or epileptic encephalopathy (*42*). All four *de novo* missense variants in ASD cases are within six residues of each other in the voltage-sensing fourth transmembrane domain, with three at a single residue previously characterized as gain-of-function in NDD (R230C, Fig. 3C) (*42*). All the variants replace one of the critical positively charged arginine residues, significantly reducing the domain's net positive charge and therefore its attraction to the electronegative cell interior. This makes a compelling case for an etiological role in the gain-of-function phenotype, and our data extend this gain-of-function phenotype to include ASD. Furthermore, the observation of seizures in loss-of-function and risk for the ASD-NDD spectrum in gain-of-function of these hyperpolarizing potassium channels is almost the opposite of that observed in *SCN2A*, which encodes the depolarizing voltage-gated sodium channel $Na_V1.2$. In

SCN2A, mild gain-of-function variants lead to BFNE, while PTVs and loss-of-function missense variants, expected to be hyperpolarizing, lead to ASD and NDD (*5, 11, 17*). Considering the other genes strongly associated with BFNE, we observe one *de novo* PTV in *KCNQ2* (FDR=0.48) and no putative risk-mediating variants in *PRRT2*.

SCN1A, which encodes the voltage-gated sodium channel $Na_V1.1$ and is a paralogue of *SCN2A (52)*, is strongly associated with Dravet syndrome, a form of progressive epileptic encephalopathy including febrile, myoclonic, and/or generalized seizures, EEG abnormalities, and NDD (*53*). Previous studies observed that up to 67% of children with Dravet syndrome also meet diagnostic criteria for ASD (*54-56*). In keeping with these findings, we observe statistical association between *SCN1A* and ASD in our cohort (FDR=0.05, TADA). Four cases have *de novo* missense variants with MPC ≥ 1 in *SCN1A* (Fig. 3A; Table S5), with three of these being located in the C-terminus (*57*); all four cases are reported to have seizures, though details of seizure onset, severity, or type are not available. In epileptic encephalopathy cohorts, PTVs are the predominant mutation type; in contrast, missense variants are the more common type in ASD and NDD (Fig. 3D). Electrophysiological analysis would be required to distinguish mild loss-of-function from gain-of-function effects for these ASD-ascertained variants.

The gene *SLC6A1* encodes GAT-1, a voltage-gated GABA transporter. *SLC6A1* was previously associated with developmental delay and cognitive impairment (*23, 41*), while a case series highlighted its role in myoclonic atonic epilepsy (MAE) and absence seizures (*45*). Here, we extend the phenotypic spectrum to include ASD, through the observation of eight *de novo* missense variants (MPC ≥ 1) and one PTV in the case-control cohort (Fig. 3E). Four of these missense variants are in the highly conserved sixth transmembrane domain, with one being

recurrent in two independent cases (A288V). Of the six ASD cases with seizure status available, five have seizures reported; four of the ASD cases have data available on cognitive performance and all four have intellectual disability. In cases ascertained for MAE, PTVs account for 54% (7 PTV, 6 missense) of observed *de novo* variants (*45*), and several of the missense variants reduce GABA transport (*58*). By contrast, in our cases ascertained for ASD, only 11% are PTVs (1 PTV, 8 missense), while cases ascertained for developmental delay fall in between (30%; 3 PTV, 7 missense) (*41*). This trend may reflect underlying correlations between genotype, protein function, and phenotype correlations, although further functional assessment is required to confirm this.

ASD genes within recurrent copy number variants (CNVs)

Large CNVs encompassing certain genomic loci represent another important source of risk for ASD (*e.g.*16p11.2 microdeletions) (*10*). However, these genomic disorder (GD) segments can include dozens of genes, which has impeded the identification of discrete dominant-acting ("driver") gene(s) within these regions. We sought to determine whether the 99 TADA-defined genes could nominate dosage-sensitive genes within GD regions. We first curated a consensus GD list from nine sources, totaling 823 protein-coding genes in 51 autosomal GD loci associated with ASD or ASD-related phenotypes, including NDD (Table S6).

Within the 51 GDs were 12/99 (12.1%) ASD genes that localized to 11/51 (21.6%) GD loci (after excluding *RAI1*, as described above; Table S6). Using multiple permutation strategies (see Supplementary Methods), we found that this observed result was greater than expected by chance when simultaneously controlling for number of genes, PTV mutation rate, and brain expression levels per gene (2.2-fold increase; $p=5.8x10^{-3}$). These 11 GD loci that encompassed a

TADA gene divided into three groups: 1) the overlapping TADA gene matched the consensus driver gene, e.g., *SHANK3* for Phelan-McDermid syndrome (22q13.3 deletion) or *NSD1* for Sotos syndrome (5q35.2 deletion) (*59, 60*); 2) a TADA gene emerged that did not match the previously predicted driver gene(s) within the region, such as *HDLBP* at 2q37.3 (Fig. 3F), where *HDAC4* has been hypothesized as a driver gene (*61, 62*); and 3) no previous gene had been established within the GD locus, such as *BCL11A* at 2p15-p16.1. One GD locus, 11q13.2-q13.4, had two genes with independent ASD associations in this study (*SHANK2* and *KMT5B*, Fig. 3G), highlighting that many GDs are the consequence of risk conferred by multiple genes within the CNV segment, including many genes likely exerting small effects that our current sample sizes are not sufficiently powered to detect (*10*).

Relationship of ASD genes with GWAS signal

Common variation plays an important role in ASD risk (*4*), as expected given the high heritability (*3*). While the specific common variants influencing risk remain largely unknown, recent genome-wide association studies (GWAS) have revealed a handful of associated loci (*63*). What has become apparent from other GWAS studies, especially those relating GWAS findings to the genes they might influence, is that risk variants commonly influence expression of nearby genes (*64*). Thus, we asked if there was evidence that common genetic variation within or near the 99 identified genes (within 10 Kb) influences ASD risk or other traits related to ASD risk. Note that among the first five genome-wide significant ASD hits from the current largest GWAS (*63*), *KMT2E* is a "double hit" – clearly implicated by the GWAS and also in the list of 99 FDR ≤ 0.1 genes described here.

To explore this question more thoroughly, we ran a gene set enrichment analysis of our 99 TADA genes against GWAS summary statistics using MAGMA (*65*) to integrate the signal for those statistics over each gene using brain-expressed protein-coding genes as our background. We used results from six GWAS datasets: ASD, schizophrenia (SCZ), major depressive disorder (MDD), and attention deficit hyperactive disorder (ADHD), which are all positively genetically correlated with ASD and with each other; educational attainment (EA), which is positively correlated with ASD and negatively correlated with schizophrenia and ADHD; and, as a negative control, human height (Table S7) (*63, 66-77*). Correcting for six analyses, we observed significant enrichment emerging from the SCZ and EA GWAS results only (Fig. 3H). Curiously, the ASD and ADHD GWAS signals were not enriched in the 99 ASD genes. Although in some ways these results are counterintuitive, one obvious confounder is power (Fig. 3I). Effective cohort sizes for the SCZ, EA, and height GWAS dwarf that for ASD, and the quality of GWAS signal strongly increases with sample size. Thus, for results from well-powered GWAS, it is reassuring that there is no signal for height, yet clearly detectable signal for two traits genetically correlated to ASD: SCZ and EA.

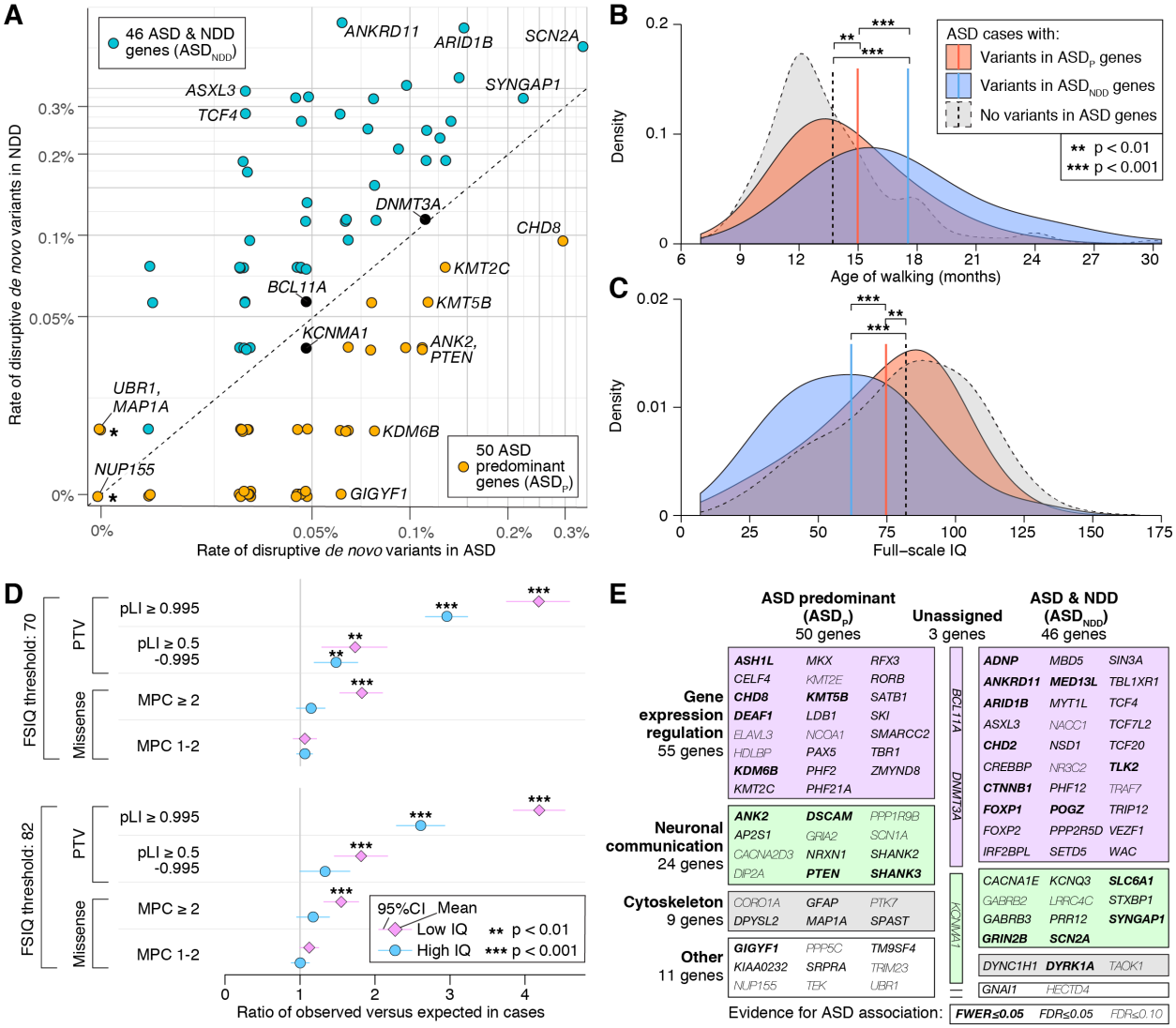


Figure 4. Phenotypic and functional categories of ASD-associated genes. A*, The frequency of disruptive de novo variants (e.g. PTVs or missense variants with MPC ≥ 1) in ascertained ASD and ascertained NDD cases (Table S4) is shown for the 99 ASD-associated genes (selected genes labeled). Fifty genes with a higher frequency in ASD are designated ASD-predominant (ASD_P), while the 46 genes more frequently mutated in NDD are designated as ASD_{NDD}. Three genes with similar frequency in the two disorders are unassigned (BCL11A, DNMT3A, and KCNMA1). Three genes marked with a star (UBR1, MAP1A, and NUP155) are included in the ASD_P category on the basis of case-control data (Table S4), which are not shown in this plot.* ***B****, ASD cases with disruptive de novo variants in ASD genes show delayed walking compared to ASD cases without such de novo variants, and the effect is greater for those with disruptive de novo variants in ASD_{NDD} genes.* ***C****, Similarly, cases with disruptive de novo variants in ASD_{NDD} genes and, to a lesser extent, ASD_P genes have a lower full-scale IQ than other ASD cases.* ***D****, Despite the association between de novo variants in ASD genes and cognitive impairment shown in 'C', an excess of disruptive de novo variants is observed in cases without intellectual disability (FSIQ ≥ 70) or with an IQ above the cohort median (FSIQ ≥ 82).* ***E****, Along with the phenotypic division (A), genes can also be classified functionally into four groups (gene expression regulation*

(GER), neuronal communication (NC), cytoskeleton, and other) based on gene ontology and research literature. The 99 ASD risk genes are shown in a mosaic plot divided by gene function and, from 'A', the ASD vs. NDD variant frequency, with the area of each box proportional to the number of genes. Statistical tests: B, C, t-test; D, chi-square with 1 degree of freedom. Abbreviations: PTV: protein-truncating variant; pLI: probability loss-of-function intolerant; MPC: missense badness, PolyPhen-2, and constraint; FSIQ: full-scale IQ.

Relationship between ASD and other neurodevelopmental disorder genes

Sibling studies yield high heritability estimates in ASD (*3*), suggesting a high contribution from inherited genetic risk factors, but comparable estimates of heritability in severe NDD, often including intellectual disability, are low (*78*). Consistent with a genomic architecture characterized by few inherited risk factors, exome studies identify an even higher frequency of gene-disrupting *de novo* variants in severe NDD than in ASD (*23, 26*). As with ASD, these *de novo* variants converge on a small number of genes, enabling numerous NDD-associated genes to be identified (*22-26*). Because at least 30% of ASD subjects have comorbid intellectual disability and/or other NDD, it is unsurprising that many genes confer risk to both disorders, as documented previously (*79*) and in this dataset (Fig. 4A). Distinguishing genes that, when disrupted, lead to ASD more frequently than NDD might shed new light on how atypical neurodevelopment maps onto the relative deficits in social dysfunction and the repetitive and restrictive behaviors in ASD.

To partition the 99 ASD genes in this manner, we compiled data from 5,264 trios ascertained for severe NDD (Table S8). Considering disruptive *de novo* variants – which we define here as *de novo* PTVs or missense variants with MPC ≥ 1 – we compared the relative frequency, R, of *de novo* variants in ASD- or NDD-ascertained trios. Genes with $R > 1.25$ were classified as ASD-predominant (ASD_P, 47 genes), while those with $R < 0.8$ were classified as ASD with NDD (ASD_{NDD}, 46 genes). An additional three genes were assigned to the ASD_P group on the basis of case-control data, while three were unassigned (Fig. 4A). For this partition, we then evaluated

transmission of rare PTVs (relative frequency < 0.001) from parents to their affected offspring: for ASD_P genes, 51 such PTVs were transmitted and 23 were not (transmission disequilibrium test, p=0.001), whereas, for ASD_{NDD} genes, 16 were transmitted and 13 were not (p=0.25). Note that the frequency of PTVs in parents is also markedly greater in ASD_P genes (1.48 per gene) than in ASD_{NDD} genes (0.80 per gene) and these frequencies are significantly different (p=0.005, binomial test), while the frequency of *de novo* PTVs in probands is not markedly different (92 in ASD_P genes, 114 in ASD_{NDD} genes, p=0.14, binomial test with probability of success = 0.498 [PTV in ASD_P gene]). Thus, the count of PTVs in ASD_P genes in parents and their segregation to ASD offspring strongly supports this classification, whereas the count of *de novo* PTVs shows only a trend for higher frequencies in ASD_{NDD} genes.

Consistent with this partition, ASD subjects who carry disruptive *de novo* variants in ASD_{NDD} genes walk 2.4 ± 1.2 months later (Fig. 4B; $p=1.6x10^{-4}$, t-test, df=238) and have an IQ 11.9 ± 6.1 points lower (Fig. 4C; $p=1.7x10^{-4}$, t-test, df=265), on average, than ASD subjects who carry disruptive *de novo* variants in ASD_P genes (Table S9). Both sets of subjects differ significantly from the rest of the ASD cohort with respect to IQ and age of walking (Fig. 4B, 4C; Table S9). While the data support some overall distinction between the genes identified in ASD and NDD *en masse*, we cannot definitively identify which specific genes are distinct at present.

<u>Burden of mutations in ASD as a function of IQ</u>

Within the set of 6,430 family-based ASD cases, 3,010 had a detected *de novo* variant and either a recorded full-scale IQ or a clinical assessment of ID. We partitioned these subjects into those with IQ ≥ 70 (69.4%) versus those with IQ < 70 (30.6%), then characterized the burden of *de novo* variants within these groups. ASD subjects in the lower IQ group carry a greater burden of

de novo variants, relative to both expectation and the high IQ group, in the two top tiers of PTVs and the top tier of missense variants (Fig. 4D). Excess burden, however, is not concentrated solely in the low IQ group, but also observed in the two top PTV tiers for the high IQ group (Fig. 4D). Similar patterns were observed if we repeat the analysis partitioning the sample at IQ < 82 (46.3%) versus IQ ≥ 82 (53.7%), which was the mean IQ after removing affected subjects who carry disruptive variants in the 99 ASD genes (Fig. 4C). Finally, considering the 99 ASD-associated genes only, there are significant contributions to the association signal from the high IQ group, as documented by model-driven simulations accounting for selection bias due to an FDR threshold (Supplementary Methods). Thus, the signal for association, mediated by mutation, is not solely limited to the low IQ subjects, supporting the idea that *de novo* variants do not solely impair cognition (*80*).

Functional dissection of ASD genes

Given the substantial increase in ASD gene discovery compared to our previous analyses, we leveraged the ASD-associated gene list to provide high-level functional insight into ASD neurobiology. Past analyses have identified two major functional groups of ASD-associated genes: those involved in gene expression regulation (GER), including chromatin regulators and transcription factors, and those involved in neuronal communication (NC), including synaptic function (*5, 10*). A simple gene ontology analysis with the new list of 99 ASD genes replicates this finding, identifying 16 genes in category GO:0006357 "regulation of transcription from RNA polymerase II promoter" (5.7-fold enrichment, FDR=6.2×10^{-6}) and 9 genes in category GO:0007268: "synaptic transmission" (5.0-fold enrichment, FDR=3.8×10^{-3}). To assign genes to the GER and NC categories for further analyses, we used a combination of gene ontology and primary literature research as described in the Supplementary Methods (Table S10 and Fig. 4E).

Considering the 20 genes not assigned to the GER and NC categories, we see the emergence of a new functional group of nine "cytoskeleton genes", based on annotation with the gene ontology term GO:0007010 "cytoskeleton organization" or related child terms. The remaining 11 genes are described as "Other" (Table S10 and Fig. 4E), many of which have roles in signaling cascades and/or ubiquitination.

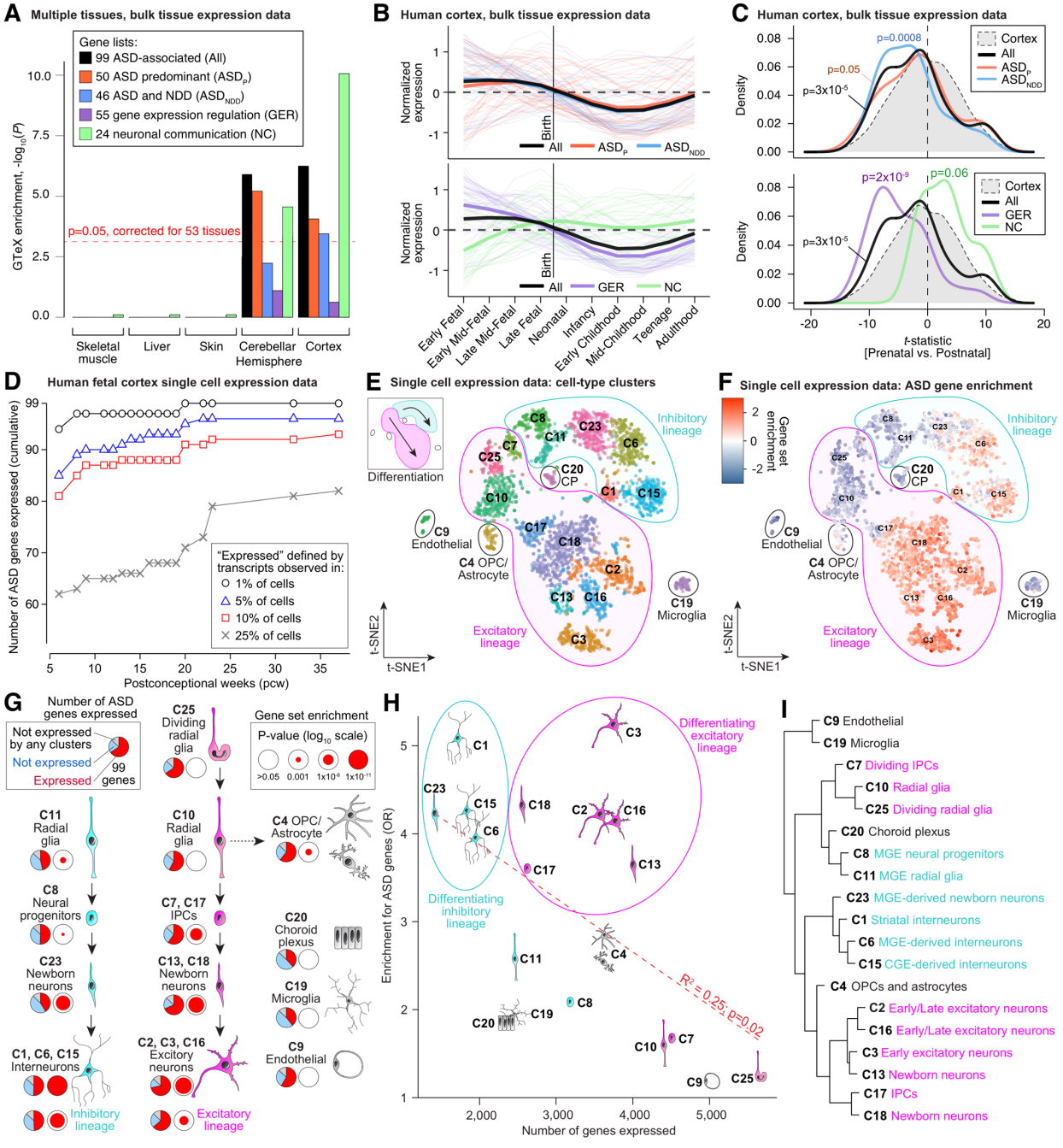


Figure 5. Analysis of 99 ASD-associated genes in the context of gene expression data. A, *GTEx bulk RNA-seq data from 53 tissues was processed to identify genes enriched in specific tissues. Gene set enrichment was performed for the 99 ASD genes and four subsets (ASD$_P$, ASD$_{NDD}$, GER, NC) for each tissue. Five representative tissues are shown here, including cortex, which showed the greatest degree of enrichment (OR=3.7; p=3.0×10^{-6}).* ***B****, BrainSpan (81) bulk RNA-seq data across 10 developmental stages was used to plot the normalized expression of the 98 brain-expressed ASD genes across development, split by the four subsets.* ***C****, A t-statistic was calculated comparing prenatal to postnatal expression in the BrainSpan data. The t-statistic distribution of the 99 ASD-associated genes shows a prenatal bias (p=3x10^{-5}), which is*

*especially pronounced for GER genes ($p=2\times10^{-9}$), while NC genes are postnatally biased ($p=0.06$). **D**, The cumulative number of ASD-associated genes expressed in RNA-seq data for 4,261 cells collected from human forebrain across prenatal development (82). **E**, t-SNE analysis identifies 19 clusters with unambiguous cell type in this single-cell expression data. **F**, The enrichment of the 99 ASD-associated genes within cells of each type is represented by color. The most consistent enrichment is observed in maturing and mature excitatory (bottom center) and inhibitory (top right) neurons. **G**, The developmental relationships of the 19 clusters is indicated by black arrows, with the inhibitory lineage shown on the left (cyan), excitatory lineage in the middle (magenta), and non-neuronal cell types on the right (grey). The proportion of the 99 ASD-associated genes observed in at least 25% of cells within the cluster is shown by the pie chart, while the log-transformed p-value of gene set enrichment is shown by the size of the red circle. **H**, The relationship between the number of cells in the cluster (x-axis) and the p-value for ASD gene enrichment (y-axis) is shown for the 19 cell type clusters. Linear regression indicates that clusters with few expressed genes (e.g. C23 newborn inhibitory neurons) have higher p-values than clusters with many genes (e.g. C25 radial glia). **I**, The relationship between the 19 cell type clusters using hierarchical clustering based on the 10% of genes with the greatest variability among cell types. Statistical tests: A, t-test; C, Wilcoxon test; E, F, Fisher Exact Test; H, I, Fisher Exact Test. Abbreviations: GTEx: Genotype-Tissue Expression; CP: choroid plexus; OPC: oligodendrocyte progenitor cells; MGE: medial ganglionic eminence; CGE: Caudal ganglionic eminence; IPC: intermediate progenitor cell; t-SNE: t-distributed stochastic neighbor embedding.*

<u>ASD genes are expressed early in brain development</u>

The 99 ASD-associated genes can thus be subdivided by functional role (55 GER genes and 24 NC genes) and phenotypic impact (50 ASD_P genes, 46 ASD_{NDD} genes) to give five gene sets (including the set of all 99). Gene expression data provide the opportunity to evaluate where and when these genes are expressed and can be used as a proxy for where and when neurobiological alterations ensue in ASD. We first evaluated enrichment for these five gene sets in the 53 tissues with bulk RNA-seq data in the Genotype-Tissue Expression (GTEx) resource. To focus on the genes that provide the most insight into tissue type, we selected genes that were expressed in a tissue at a significantly higher level than the remaining 52 tissues, specifically log fold-change > 0.5 and FDR<0.05 (*t*-test, R package limma). Subsequently, we assessed over-representation of each ASD gene set within 53 sets of genes expressed in each tissue relative to a background of all tissue-specific genes in GTEx. At a threshold of $p \leq 9\text{x}10^{-4}$, reflecting 53 tissues, enrichment was observed in 11 of the 13 samples of brain tissue, with the strongest enrichment in cortex

(∩=30 genes; $p=3\times10^{-6}$; OR=3.7; Fig. 5A) and cerebellar hemisphere (∩=41 genes; $p=3\times10^{-6}$; OR=2.9; Fig. 5A). Of the four gene subsets, NC genes were the most highly enriched in cortex (FDR=2×10^{-10}; OR=22.1; Fig. 5A), while GER genes were the least enriched (FDR=0.63; OR=1.7; Fig. 5A).

We next leveraged the BrainSpan human neocortex bulk RNA-seq data (*83*) to assess enrichment of ASD genes across development (Fig. 5B, 5C). Of the 17,484 autosomal protein-coding genes assessed for ASD-association, 13,735 genes (78.5%) were expressed in the neocortex (RPKM ≥ 0.5 in 80% of samples of at least one neocortical region and developmental period). Of the 99 ASD-associated genes, only the cerebellar transcription factor *PAX5 (81)* was not expressed in the cortex (78 expected; $p=1x10^{-9}$, binomial test). Compared to other genes expressed in the cortex, the remaining 98 ASD genes are expressed at higher levels during prenatal development, but at lower levels during postnatal development (Fig. 5B). To quantify this pattern, we developed a *t*-statistic that assesses the relative prenatal vs. postnatal expression of each of the 13,735 protein-coding genes. Using this metric, the 98 cortex-expressed ASD-associated genes showed enrichment in the prenatal cortex ($p=3\times10^{-5}$, Wilcoxon test; Fig. 5C). The ASD_P and ASD_{NDD} gene sets showed similar patterns (Fig. 5B), though the prenatal bias *t*-statistic was slightly more pronounced for the ASD_{NDD} group (p=0.0008; Fig. 5C). In contrast, the functional subdivisions reveal distinct patterns, with the GER genes reaching their highest levels during early to late fetal development (Fig. 5B) with a marked prenatal bias ($p=2\times10^{-9}$; Fig. 5C), while NC genes are highest between late mid-fetal development and infancy (Fig. 5B) and show a trend towards postnatal bias (p=0.06; Fig. 5C). Thus, supporting their role in ASD risk and in keeping with prior analyses (*19, 84-86*), the ASD genes show higher expression in human neocortex and are expressed early in brain development. The differing expression patterns of

GER and NC genes may reflect two distinct periods of ASD susceptibility during development or a single susceptibility period when both functional gene sets are highly expressed in mid- to late fetal development.

ASD genes are enriched in maturing and mature inhibitory and excitatory neurons

Prior analyses have implicated excitatory glutamatergic neurons in the cortex and medium spiny neurons in the striatum in ASD (*19, 84-86*) using a variety of systems analytical approaches, including gene co-expression. Here, we exploit the 99 ASD-associated genes to perform a more direct assessment, leveraging existing single-cell RNA-seq data from 4,261 cells collected from the prenatal human cortex (*82*), ranging from 6 to 37 post-conception weeks (pcw) with an average of 16.3 pcw (Table S11).

Following the logic that only genes that were expressed could mediate ASD risk when disrupted, we divided the 4,261 cells into 17 bins by developmental stage and assessed the cumulative distribution of expressed genes by developmental endpoint (Fig. 5D). For each endpoint, a gene was defined as expressed if at least one transcript mapped to this gene in 25% or more of cells for at least one pcw stage. By definition, more genes were expressed as fetal development progressed, with 4,481 genes expressed by 13 pcw and 7,171 genes expressed by 37 pcw. While the majority of ASD-associated genes were expressed at the earliest developmental stages (e.g. 66 of 99 at 13 pcw), the most dramatic increase in the number of genes expressed occurred during midfetal development (68 by 19 pcw, rising to 79 by 23 pcw), consistent with the BrainSpan bulk-tissue data (Fig. 5B, 5C). More liberal thresholds for gene expression resulted in higher numbers of ASD-associated genes expressed (Fig. 5D), but the patterns of expression were similar across definitions and when considering gene function or cell type (Fig. S4).

To investigate the cell types implicated in ASD, we considered 25 cell type clusters identified by t-distributed stochastic neighbor embedding (t-SNE) analysis, of which 19 clusters, containing 3,839 cells, were unambiguously associated with a cell type (*82*) (Fig. 5E, Table S11) and were used for enrichment analysis. Within each cell type cluster, a gene was considered expressed if at least one of its transcripts was detected in 25% or more cells; 7,867 protein coding genes met this criterion in at least one cluster. From cells of each type, by contrasting one cell type to the others, we observed enrichment for the 99 ASD-associated genes in maturing and mature neurons of the excitatory and inhibitory lineages (Fig. 5F, 5G) but not in non-neuronal cells. Early excitatory neurons (C3) expressed the most ASD genes ($\cap$=71 genes, $p < 1\times10^{-10}$), while choroid plexus (C20) expressed the fewest ASD genes ($\cap$=38 genes, p=0.006); 13 genes were not expressed in any cluster (Fig. 5G). Within the major neuronal lineages, early excitatory neurons (C3) and striatal interneurons (C1) showed the greatest degree of gene set enrichment ($\cap$=71 and $\cap$=50 genes, $p < 1\times10^{-10}$; Fig. 5F, 5G; Table S11). Overall, maturing and mature neurons in the excitatory and inhibitory lineages showed a similar degree of enrichment, while those in the excitatory lineage expressed the most ASD genes; this difference is due to the larger numbers of genes expressed in excitatory lineage cells (Fig. 5H). The only non-neuronal cell type with significant enrichment for ASD genes was oligodendrocyte progenitor cells (OPCs) and astrocytes (C4; $\cap$=60 genes, $p=1x10^{-5}$). To assess the validity of the t-SNE clusters, we selected 10% of the expressed genes showing the greatest variability among the cell types and performed hierarchical clustering (Fig. 5I). This recaptured the division of these clusters by lineage (excitatory vs. inhibitory) and by development stage (radial glia and progenitors vs. neurons).

Thus, based on the intersection of the ASD-associated genes and three gene expression datasets, we show that all 99 ASD-associated genes are brain expressed; the bulk of these genes show high expression during fetal development, especially during mid-to-late fetal periods; and the vast majority of these genes are expressed in both excitatory and inhibitory neuronal lineages. Enrichment of ASD-associated genes strongly implicates both excitatory and inhibitory neurons in ASD during their maturation in mid-to-late fetal development.

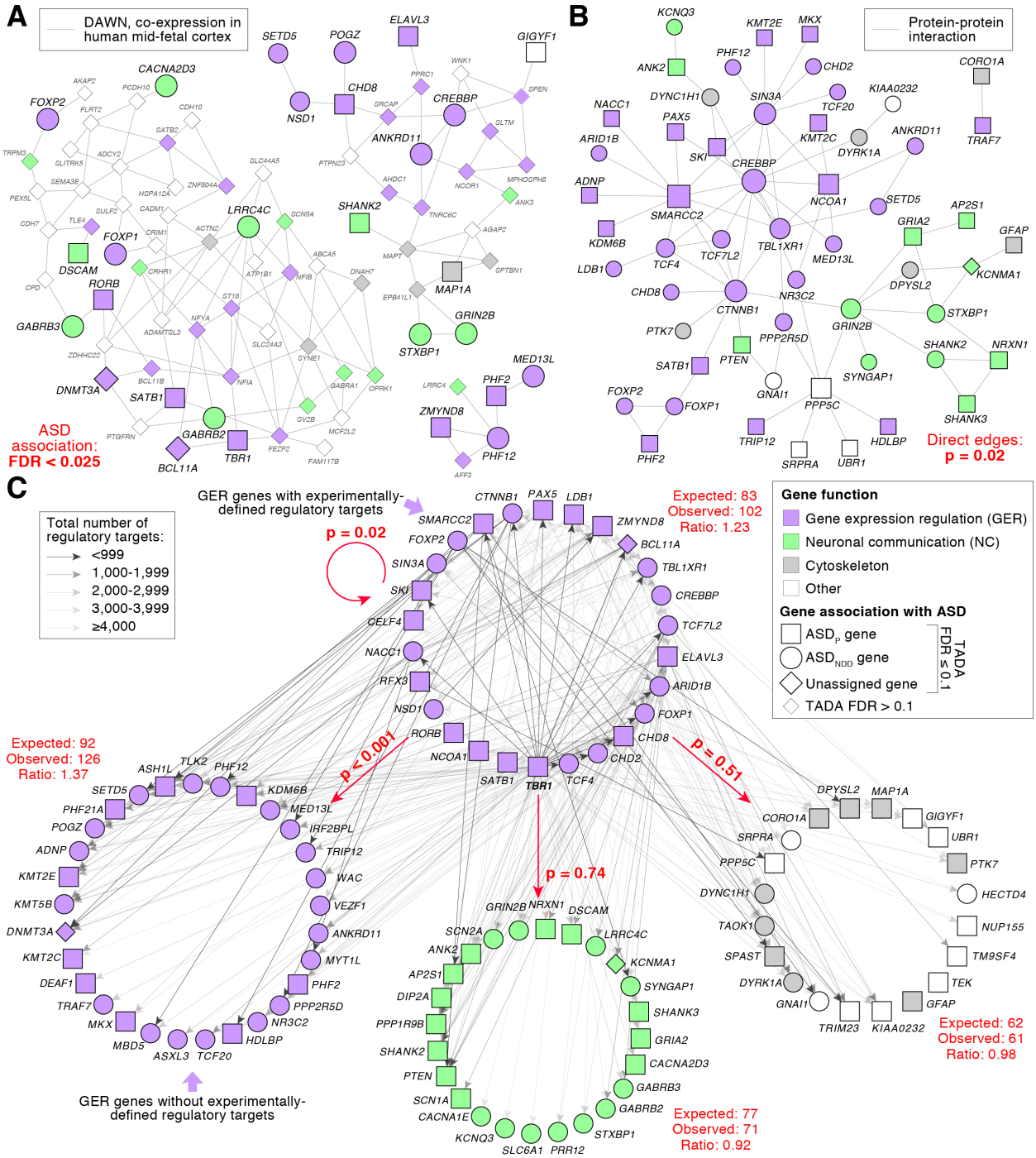


***Figure 6. Functional relationships of ASD risk genes. A**, ASD association data from TADA (Table S4) is integrated with co-expression data from the midfetal human brain to implicate additional genes in ASD. The top 100 genes that share edges are shown (FDR ≤ 0.025). **B**, ASD-associated genes form a single protein-protein interaction network with more edges than expected by chance (p=0.02). **C**, Experimental data, obtained using ChIP and CLIP methods across multiple species and a wide range of neuronal and non-neuronal tissues types, identified the regulatory targets of 26 GER genes (top circle). These data were used to assess whether three functionally-defined groups of ASD-associated genes were enriched for regulatory targets,*

represented as arrows, weighted by the total number of regulatory targets for the GER gene. The expected number of targets in each functional group was estimated by permutation, controlling for brain expression, de novo PTV mutation rate, and pLI. Statistical tests: A, DAWN; B, Permutation; C, Permutation. Abbreviations: DAWN: Discovering Association With Networks.

Functional relationships among ASD genes and prediction of novel risk genes

The ASD-associated genes show convergent functional roles (Fig. 4E) and expression patterns in the human cortex (Fig. 5B). It is therefore reasonable to hypothesize that genes co-expressed with these ASD genes might have convergent or auxiliary function and thus contribute to risk. The Discovering Association With Networks (DAWN) approach integrates ASD association with gene co-expression data to identify clusters of genes with highly correlated co-expression, some of which also show strong association signal from TADA (*87*). Our previous DAWN analysis identified 160 putative ASD risk genes, 146 of which were not highlighted by the ASD association data alone (*5*). Of these 146, 11 are in our list of 99 ASD-associated genes, reflecting highly significant enrichment ($p=7.9x10^{-10}$, FET). Here, we leveraged the DAWN model using our new TADA results (Table S12) and BrainSpan gene co-expression data from the midfetal human cortex, as implicated in our analyses (Fig. 5B, 5E), to look for additional genes plausibly implicated in risk. DAWN yields 100 genes (FDR ≤ 0.025), including 40 that are captured in the 99 TADA ASD genes and 60 that are not (Fig. 6A). Of these 60 genes, three are associated with NDD (*23*) and another 15 have been associated with rare genetic disorders (*88, 89*); of note, six of these have autosomal recessive inheritance (Table S12). If these 60 novel genes impact ASD risk, we would predict the set would be highly enriched in the excitatory and inhibitory cell types (Fig. 5E-5H). This expectation is supported with 38 out of 60 genes being expressed in excitatory cell types ($p < 1.6x10^{-4}$, FET), 25 of which are also expressed in inhibitory cell types ($p < 7.9\times10^{-4}$, FET). Furthermore, many of these 60 genes play a role in GER or NC (Fig. 6A).

We also sought to interpret gene co-expression and enrichment across a broader range of early developmental samples using a common analytical tool, Weighted Gene Coexpression Network Analysis (WGCNA). With WCGNA, we analyzed spatiotemporal co-expression from 177 high-quality BrainSpan samples aged 8 pcw to 1 year, yielding 27 early developmental co-expression modules (Fig. S5, Table S13). If a module captures ASD-related biology, then we would expect to see ASD genes mapping therein. We identified significant over-representation in two modules after correction for multiple testing (Fig. S5, Table S13): M4 contained a significant over-representation of the NC gene set (p=0.002, FET, OR=13.7, ∩=5 genes); and M25 contained a significant over-representation across all 99 ASD genes ($p=3x10^{-11}$, FET, OR=12.1, ∩=17 genes), driven by the GER gene set ($p=9x10^{-16}$, OR=26.2, ∩=17 genes). With regard to single-cell gene expression, genes in NC-specific M4 showed greatest enrichment in maturing and mature neurons, both excitatory and inhibitory ($p < 0.001$ for each of 6 neuronal cell types, FET), whereas genes in M25 showed enrichment across all 19 cell types ($p < 0.001$ for all cell types, FET).

GER and NC gene sets play a prominent role in risk for ASD despite their disparate functions, patterns of expression (Fig. 5B), and early developmental co-expression (Fig. 6A and Table S12); however, the manner in which these two gene sets converge on the ASD phenotype remains unclear. We considered whether these genes might have additional, previously unrecognized interactions at the protein-level, for example, an extranuclear role for GER genes. Protein-protein interaction (PPI) analysis (Fig. 6B, Table S14) identified a significant excess of interactions between all ASD genes (∩=82 genes, p=0.02), GER genes (∩=49 genes, p=0.006), and NC genes (∩=12 genes, p=0.03), but not between GER and NC genes (∩=2 genes, p=1.00). We therefore evaluated whether the GER genes regulate the NC genes. To perform this analysis,

we collated experimentally derived ChIP- and CLIP-seq data identified by searching ChEA, ENCODE, and PubMed (Table S15). We identified at least one dataset of regulatory targets for 26 of the 55 GER genes across multiple tissue types (neural tissue in 31% (∩=8) of genes) and three species (human tissue in 54% (∩=14) of genes, with mouse/rat accounting for the remainder). Across the 26 genes, 14,925 protein-coding genes were targeted. The regulatory targets of the 26 GER genes were enriched for the same 26 GER genes (1.2-fold over expectation; p=0.02) and the other 29 GER genes (1.3-fold over expectation; p<0.001), but not NC genes or genes with other functions (Fig. 6C).

These results raise the possibility that the GER genes do not regulate the NC genes directly, but rather potentially converge with NC genes in downstream processes in maturing neurons. However, these findings must be interpreted with some degree of caution, due to the non-human and non-neural tissues and heterogeneous methodologies. A similar caveat holds for the PPI analysis (Fig. 6b); studies of protein interaction from brain tissue are limited. Therefore, to address these caveats and to provide additional human brain-specific support for this hypothesis, we examined whether GER and NC gene sets relate to well-curated binding sites for *CHD8*, a GER hub gene (Fig. 6C), using human brain-specific ChIP sequencing data from two independent studies. Strong and consistent enrichment for *CHD8* binding sites (Fig. S6) were observed amongst GER genes for *CHD8* sites derived from the human mid-fetal brain at 16-19 pcw (p=0.001) as well as *CHD8* sites derived from human neural progenitor cells (p=0.001), however we did not observe significant enrichment for NC genes (p=0.10, p=0.25, respectively).

Discussion

We explore rare *de novo* and inherited coding variation from 35,584 individuals, including 11,986 ASD cases – the largest number of cases analyzed to date – and implicate 99 genes in risk for ASD at FDR ≤ 0.1 (Fig. 2). The evidence for several of the 99 genes is driven by missense variants, including confirmed gain-of-function mutations in the potassium channel *KCNQ3* and patterns that may similarly reflect gain-of-function or altered function in *DEAF1*, *SCN1A*, and *SLC6A1* (Fig. 3). Twelve of the 99 ASD-associated genes fall in established genomic disorder (GD) loci, a greater number than expected by chance, despite these two data sources being independent. Similarly, we observe substantial overlap with common variants associated with schizophrenia and educational attainment (Fig. 3). Collectively, many of the genes implicated herein provide important new insights into the functional pathways, tissues, cell types, and developmental timing involved in ASD risk, as well as specificity for ASD versus broader NDD phenotypes.

By comparing mutation frequencies in ASD cases in our study to other studies in which subjects were ascertained for severe neurodevelopmental delay (NDD), we partitioned the 99 ASD-associated genes into two groups: those that occur at a higher frequency in our ASD subjects, 50 ASD_P genes, and those that occur at a higher frequency in NDD subjects, 46 ASD_{NDD} genes (Fig. 4A). Two additional lines of evidence support the partition: first, cognitive impairment and motor delay are more frequently observed in our subjects (all ascertained for ASD) with mutations in ASD_{NDD} than in ASD_P genes, in keeping with the wider neurodevelopmental impact of the ASD_{NDD} genes (Fig. 4B, 4C); second, while *de novo* PTVs were observed at a similar frequency in ASD_P and ASD_{NDD} genes in ASD subjects, their parents more frequently carried PTVs in ASD_P genes than in ASD_{NDD} genes, and they transmitted them to their offspring far more often. Together, these observations indicate that ASD-associated genes are distributed

across a spectrum of phenotypes and selective pressure. At one extreme, gene haploinsufficiency leads to global developmental delay, with impaired cognitive, social, and gross motor skills leading to extreme negative selection (e.g. *ANKRD11* or *ARID1B*). At the other extreme, gene haploinsufficiency leads to ASD, but there is a more modest involvement of other developmental phenotypes and selective pressure (e.g. *GIGYF1* or *ANK2*). This distinction has important ramifications for clinicians, geneticists, and neuroscientists, since it suggests that clearly delineating the impact of these genes across neurodevelopmental dimensions may offer a route to deconvolve the social dysfunction and repetitive behaviors that define ASD from more general neurodevelopmental impairment.

Observing the convergence of both GER and NC genes in maturing and mature neurons (Fig. S4) raises the question of how they interact. This interaction does not appear to be at the level of direct protein contact (Fig. 6B), despite both gene sets being represented in the PPI dataset. The bias of GER genes towards earlier expression than NC (Fig. 5c), alongside their functional role, raises the hypothesis that the GER genes act through regulation of downstream NC genes. Testing the regulatory targets of 26 GER genes speaks against this simple relationship, since we observed enrichment for the regulation of other GER genes, but not of NC genes. While the heterogeneous data sources and tissue types limit this analysis, if GER genes mediate risk by regulating NC genes, we would expect a clear and strong enrichment signal, which we do not see. Focusing on *CHD8*, a GER gene strongly associated with ASD (Fig. 2) for which regulatory targets have been defined in neuronal tissues including human fetal cortex (*90, 91*), we show that enrichment of ASD-associated genes in these targets is exclusive to GER genes (Fig. S6). Experimental validation of this surprising result in neural, ideally human, tissues is critical.

Analyses of the 99 ASD-associated genes in the context of single-cell gene expression data from the developing human cortex (*82*) implicated mid-to-late fetal development and maturing and mature neurons in both excitatory and inhibitory lineages (Fig. 5). Non-neuronal cells did not show substantial enrichment, with the exception of astrocytes and OPCs that expressed 60 ASD genes (2.7-fold enrichment; p=0.0002). Of these 60 genes, 58 overlapped with radial glia, which may reflect shared developmental origins rather than an independent enrichment signal. In contrast to post-mortem findings in ASD brains (*92, 93*), no enrichment was observed in microglia. These findings validate and extend prior network analyses (*19, 84-86*) by leveraging a substantially larger ASD gene set and gene expression at single-cell resolution in developing human brains.

Our enrichment tests (Fig. 5F-H) implicitly assume that the functional consequences of haploinsufficiency are greatest at higher levels of expression, as required for 25% of cells to express the gene. Alternatively, if haploinsufficiency leads to functional consequences even at low levels of gene expression, then earlier developmental stages and more cell types are involved in ASD neurobiology. Because many ASD-associated genes show high expression across a variety of developmental stages, all early in neurodevelopment, we predict that damaging mutations to any one of them alters neurodevelopmental trajectory, perhaps uniquely. Moreover, most of these genes could impact the trajectories of both the excitatory and inhibitory lineages, implying they have a remarkable range of impacts on both excitatory and inhibitory development. If true, this has broad implications for the neurobiology of ASD, including the hopelessness of grasping its nature by studying the impact of one gene in one cell type and in one developmental context at a time. Rather, ASD must arise by some commonality amongst diverse neurodevelopmental trajectories. Any such hypothesis needs to explain convergence on ASD

phenotype – with its readily recognizable impairments in social communication and restricted or repetitive behaviors or interests – based on our ASD-associated set of genes. Two related and very general hypotheses are compatible, and they involve inappropriate crosstalk between excitatory and inhibitory neurons: an excitatory-inhibitory imbalance (*94*) and failed homeostatic control over cortical circuits (*95*).

Conclusion

Through an international collaborative effort, the willingness of thousands of families to volunteer, and the integration of data from several large-scale genomic collaborations, we have assembled a cohort of 35,584 samples from which we identify 99 ASD-associated genes (FDR ≤ 0.1; Fig. 2), including some acting through gain-of-function missense variants (Fig. 3). We observe phenotypic distinctions, identifying a group of 50 ASD genes (ASD_P, Fig. 4) that are enriched for ASD features, distinct from cognitive or motor impairments, and consequently subject to more modest selective pressures. We also observe functional distinctions, with 55 genes regulating the expression of other genes (GER), 24 genes implicated in neuronal communication (NC), and the remainder enriched for genes that play a role in the cytoskeleton. These functional distinctions are mirrored in gene expression, gene co-expression, and protein-protein interaction data (Fig. 5, 6), but not phenotypically (Fig. S7), although both sets of genes are enriched in maturing and mature excitatory and inhibitory neurons in the fetal brain (Fig. S4). While these gene sets converge in cell type and overlap in expression trajectories, based on currently available data, the NC genes do not appear to be enriched as regulatory targets of the GER genes (Fig. 5 and S6). Identifying the nature of this convergence, especially in ASD-

enriched genes, is likely to hold the key to understanding the neurobiology that underlies the ASD phenotype.

Acknowledgements

We thank the families who participated in this research, without whose contributions genetic studies would be impossible. This study was supported by the National Institute of Mental Health (U01s: MH100209 (to B.D.), MH100229 (to M.J.D.), MH100233 (to J.D.B), & MH100239 (to M.W.S.); U01s: MH111658 (to B.D.), MH111660 (to M.J.D.), MH111661 (to J.D.B), & MH111662 (to S.J.S. and M.W.S.); Supplement to U01 MH100233 (MH100233-03S1 to J.D.B.); R37 MH057881 (to B.D.); R01 MH109901 (to S.J.S. M.W.S., A.J.W.); R01 MH109900 (to K.R.); and, R01 MH110928 (to S.J.S., M.W.S., A.J.W.)), National Human Genome Research Institute (HG008895), Seaver Foundation, Simons Foundation (SF402281 to S.J.S., M.W.S., B.D., K.R.), and Autism Science Foundation (to S.J.S., S.L.B., E.B.R.). M.E.T. is supported by R01 MH115957 and Simons Foundation (SF573206) and R.C. is supported by NHGRI T32 HG002295-14 and NSF GRFP 2017240332. S.D.R. is supported by the Seaver Foundation. 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. The Danish National Biobank resource at the Statens Serum Institut was supported by the Novo Nordisk Foundation. Sequencing of iPSYCH samples was supported by grants from the Simons Foundation (SFARI 311789 to M.J.D) and the Stanley Foundation. Other support for this study was received from the NIMH (5U01MH094432-02 to M.J.D). Computational resources for handling and statistical analysis of iPSYCH data on the GenomeDK and Computerome HPC facilities were provided by, respectively, Centre for Integrative Sequencing,

iSEQ, Aarhus University, Denmark (grant to A.D.B), and iPSYCH. The iPSYCH study was approved by the Regional Scientific Ethics Committee in Denmark and the Danish Data Protection Agency. The Norwegian Mother and Child Cohort Study is supported by the Norwegian Ministry of Health and Care Services and the Ministry of Education and Research, NIH/NINDS (grant no.1 UO1 NS 047537-01 and grant no.2 UO1 NS 047537-06A1). We are grateful to all the participating families in Norway who take part in this ongoing cohort study and the Autism Birth Cohort Study. This work was also supported by the Research Council Norway grant number 185476 and Wellcome Trust grant number (098051]). For the collection of the cohort in Turin, the Italian Ministry for Education, University and Research (Ministero dell'Istruzione, dell'Università e della Ricerca - MIUR) funded the Department of Medical Sciences under the program "Dipartimenti di Eccellenza 2018 – 2022" Project code D15D18000410001. We also thank the Associazione "Enrico e Ilaria sono con noi" ONLUS and the Fondazione FORMA. The collection of the cohort in Santiago de Compostela, Spain (Angel Carracedo) was supported by the Fundación María José Jove. The collection of the cohort in Madrid, Spain (Mara Parellada) was funded by Instituto de Salud Carlos III (C.0001481, PI14/02103, PI17/00819) and IiSGM. The collection in Japan (Branko Aleksic) was supported by AMED under grant No. JP18dm0107087 and No. JP18dm0207005. The collection in Hong Kong (Brian H.Y. Chung) was supported by the Society for the Relief of Disabled Children, Hong, Kong. For the collection in Germany (Andreas Chiocchetti and Christine M. Freitag), we thank S. Lindlar, J. Heine, and H. Jelen for technical assistance, and H. Zerlaut and C. Lemler for database management. The collection was supported by Saarland University (T6 03 10 00-45 to Christine M Freitag); German Research Association DFG (Po 255/17-4 to Fritz Poustka); EC FP6-LIFESCIHEALTH (512158; AUTISM MOLGEN to Annemarie Poustka, and Fritz Poustka); BMBF ERA-NET NEURON project: EUHFAUTISM (EUHFAUTISM-

01EW1105 to Christine M Freitag); Landes-Offensive zur Entwicklung wissenschaftlich-ökonomischer Exzellenz (LOEWE): Neuronal Coordination Research Focus Frankfurt (NeFF, to Christine M Freitag), and EC IMI initiative AIMS-2-TRIALS (777394-2 to Christine M Freitag). During the last 3 years, Christine M Freitag has been consultant to Desitin and Roche, receives royalties for books on ASD, ADHD, and MDD, and has been granted research funding by the European Commission (EC), Deutsche Forschungsgemeinschaft (DFG), and the German Ministry of Science and Education (BMBF). The collection in Utah was supported by R01 MH094400 (to Hilary Coon). The collection in Siena (Alessandra Renieri) was supported by the "Cell lines and DNA bank of Rett Syndrome, X-linked mental retardation and other genetic diseases", member of the Telethon Network of Genetic Biobanks (project no. GTB12001), funded by Telethon Italy, and of the EuroBioBank network. The PAGES collection in Sweden was supported by R01 MH097849 (to J.D.B.), Supplement to R01 MH097849 (MH097849-02S1 to J.D.B.), and R01 MH097849 (to J.D.B.). The collection at University of Pittsburgh (Nancy Minshew) was supported by the Trees Charitable Trust. The collection in Brazil (Maria Rita Passos-Bueno) was supported by Fundação de apoio a pesquisa do estado de São Paulo (FAPESP)/CEPID 2013/08028-1, Conselho Nacional do Desenvolvimento Tecnológico (CNPq)466651/2014-7. For the collection in Finland (Kaija Puura), we thank The Academy of Finland (grant 286284 to T.L.); Competitive State Research Financing of the Expert Responsibility area of Tampere University Hospital (to T.L., K.P.); Signe and Ane Gyllenberg Foundation (to T.L.); Tampere University Hospital Supporting Foundation (to T.L.), European Union (The GEBACO Project no. 028696, to K.P. and M.K.), the Medical Research Fund of Tampere University Hospital (to K.P.), The Child Psychiatric Research Foundation (Finland, to M.K.) and The Emil Aaltonen Foundation (to M.K.). For the collection at UCSF (Lauren A. Weiss), we acknowledge funding sources NIH Exploratory/Developmental Research Grant

Award (R21) HD065273 (to L.A.W.), Simons Foundation Autism Research Initiative (SFARI) 136720 (to L.A.W.) as well as IMHRO and UCSF-Research Evaluation and Allocation Committee (REAC) support (to L.A.W.). The collection at UIC (Edwin H. Cook) was supported by NICHD P50 HD055751 (to E.H.C), and the sequencing was funded through X01 HG007235. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. DDG2P data used for the analyses described in this manuscript were obtained from http://www.ebi.ac.uk/gene2phenotype/. The funders played no role in the design of the study, in the collection, analysis, and interpretation of data, or in writing the manuscript. We thank Tom Nowakowski (UCSF) for facilitating access to the single-cell gene expression data.

References

1. *Diagnostic and statistical manual of mental disorders : DSM-5*. (Fifth edition. Arlington, VA : American Psychiatric Publishing, ©2013, 2013).
2. J. Baio *et al.*, Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years—Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014. *MMWR Surveillance Summaries* **67**, 1 (2018).
3. B. H. K. Yip *et al.*, Heritable variation, with little or no maternal effect, accounts for recurrence risk to autism spectrum disorder in Sweden. *Biological psychiatry* **83**, 589-597 (2018).
4. T. Gaugler *et al.*, Most genetic risk for autism resides with common variation. *Nature genetics* **46**, 881-885 (2014).
5. S. De Rubeis *et al.*, Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209-215 (2014).
6. I. Iossifov *et al.*, The contribution of de novo coding mutations to autism spectrum disorder. *Nature* **515**, 216-221 (2014).
7. I. Iossifov *et al.*, De novo gene disruptions in children on the autistic spectrum. *Neuron* **74**, 285-299 (2012).
8. B. J. O'Roak *et al.*, Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* **485**, 246-250 (2012).
9. Stephan J. Sanders *et al.*, Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams Syndrome region, are strongly associated with autism. *Neuron* **70**, 863-885 (2011).
10. S. J. Sanders *et al.*, Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. *Neuron* **87**, 1215-1233 (2015).
11. S. J. Sanders *et al.*, De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**, 237-241 (2012).
12. S. Dong *et al.*, De novo insertions and deletions of predominantly paternal origin are associated with autism spectrum disorder. *Cell Rep* **9**, 16-23 (2014).
13. D. Levy *et al.*, Rare de novo and transmitted copy-number variation in autistic spectrum disorders. *Neuron* **70**, 886-897 (2011).
14. J. Sebat *et al.*, Strong association of de novo copy number mutations with autism. *Science (New York, N.Y.)* **316**, 445-449 (2007).
15. X. He *et al.*, Integrated model of de novo and inherited genetic variants yields greater power to identify risk genes. *PLoS genetics* **9**, e1003671 (2013).
16. K. E. Samocha *et al.*, A framework for the interpretation of de novo mutation in human disease. *Nature genetics* **46**, 944-950 (2014).
17. R. Ben-Shalom *et al.*, Opposing effects on NaV1.2 function underlie differences between SCN2A variants observed in individuals with autism spectrum disorder or infantile seizures. *Biological psychiatry* **82**, 1-9 (2017).
18. R. Bernier *et al.*, Disruptive CHD8 mutations define a subtype of autism early in development. *Cell* **158**, 263-276 (2014).
19. A. J. Willsey *et al.*, Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* **155**, 997-1007 (2013).
20. J. D. Buxbaum *et al.*, The autism sequencing consortium: large-scale, high-throughput sequencing in autism spectrum disorders. *Neuron* **76**, 1052-1056 (2012).

21. S. L. Bishop *et al.*, Identification of developmental and behavioral markers associated with genetic abnormalities in autism spectrum disorder. *American Journal of Psychiatry* **174**, appi.aj 2017.2011 (2017).
22. J. de Ligt *et al.*, Diagnostic exome sequencing in persons with severe intellectual disability. *The New England journal of medicine* **367**, 1921-1929 (2012).
23. S. Deciphering Developmental Disorders, Prevalence and architecture of de novo mutations in developmental disorders. *Nature* **542**, 433-438 (2017).
24. C. Gilissen *et al.*, Genome sequencing identifies major causes of severe intellectual disability. *Nature* **511**, 344-347 (2014).
25. A. Rauch *et al.*, Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet (London, England)* **380**, 1674-1682 (2012).
26. S. H. Lelieveld *et al.*, Meta-analysis of 2,104 trios provides support for 10 new genes for intellectual disability. *Nature neuroscience* **19**, 1194-1196 (2016).
27. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)* **25**, 1754-1760 (2009).
28. A. McKenna *et al.*, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* **20**, 1297-1303 (2010).
29. F. K. Satterstrom *et al.*, ASD and ADHD have a similar burden of rare protein-truncating variants. *bioRxiv*, (2018).
30. M. Lek *et al.*, Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-291 (2016).
31. J. A. Kosmicki *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).
32. K. E. Samocha *et al.*, Regional missense constraint improves variant deleteriousness prediction. *bioRxiv*, (2017).
33. R. A. Power *et al.*, Fecundity of patients with schizophrenia, autism, bipolar disorder, depression, anorexia nervosa, or substance abuse vs their unaffected siblings. *JAMA psychiatry* **70**, 22-30 (2013).
34. D. L. Christensen *et al.*, Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years - Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2012. *Morbidity and mortality weekly report. Surveillance summaries* **65**, 1-23 (2016).
35. A. K. Halladay *et al.*, Sex and gender differences in autism spectrum disorder: Summarizing evidence gaps and identifying emerging areas of priority. *Molecular autism* **6**, 36 (2015).
36. D. M. Werling, The role of sex-differential biology in risk for autism spectrum disorder. *Biology of Sex Differences* **7**, 1-18 (2016).
37. R. E. Slager, T. L. Newton, C. N. Vlangos, B. Finucane, S. H. Elsea, Mutations in RAI1 associated with Smith-Magenis syndrome. *Nature genetics* **33**, 466-468 (2003).
38. M. J. Bottomley *et al.*, The SAND domain structure defines a novel DNA-binding fold in transcriptional regulation. *Nature structural biology* **8**, 626-633 (2001).
39. A. T. Vulto-van Silfhout *et al.*, Mutations affecting the SAND domain of DEAF1 cause intellectual disability with severe speech impairment and behavioral problems. *Am J Hum Genet* **94**, 649-661 (2014).

40. L. Chen *et al.*, Functional analysis of novel DEAF1 variants identified through clinical exome sequencing expands DEAF1-associated neurodevelopmental disorder (DAND) phenotype. *Human mutation* **38**, 1774-1785 (2017).
41. H. O. Heyne *et al.*, De novo variants in neurodevelopmental disorders with epilepsy. *Nature genetics* **50**, 1048-1053 (2018).
42. F. Miceli *et al.*, Early-onset epileptic encephalopathy caused by gain-of-function mutations in the voltage sensor of Kv7.2 and Kv7.3 potassium channel subunits. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **35**, 3782-3793 (2015).
43. S. Maljevic *et al.*, Novel KCNQ3 mutation in a large family with benign familial neonatal epilepsy: A rare cause of neonatal seizures. *Molecular Syndromology* **7**, 189-196 (2016).
44. M. J. Landrum *et al.*, ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic acids research* **42**, D980-D985 (2014).
45. K. M. Johannesen *et al.*, Defining the phenotypic spectrum of SLC6A1 mutations. *Epilepsia* **59**, 389-402 (2018).
46. C. Leroy *et al.*, The 2q37-deletion syndrome: an update of the clinical spectrum including overweight, brachydactyly and behavioural features in 14 new patients. *Eur J Hum Genet* **21**, 602-612 (2013).
47. B. P. Coe *et al.*, Refining analyses of copy number variation identifies specific genes associated with developmental delay. *Nature genetics* **46**, 1063-1071 (2014).
48. G. M. Cooper *et al.*, A copy number variation morbidity map of developmental delay. *Nature genetics* **43**, 838-846 (2011).
49. G. B. Schaefer, N. J. Mendelsohn, Clinical genetics evaluation in identifying the etiology of autism spectrum disorders: 2013 guideline revisions. *Genetics in medicine : official journal of the American College of Medical Genetics* **15**, 399-407 (2013).
50. P. J. Jensik, J. I. Huggenvik, M. W. Collard, Identification of a nuclear export signal and protein interaction domains in deformed epidermal autoregulatory factor-1 (DEAF-1). *Journal of Biological Chemistry* **279**, 32692-32699 (2004).
51. Y. Chen *et al.*, Modeling Rett syndrome using TALEN-edited MECP2 mutant cynomolgus monkeys. *Cell* **169**, 945-955 (2017).
52. S. J. Sanders *et al.*, Progress in understanding and treating SCN2A-mediated disorders. *Trends in neurosciences* **41**, 442-456 (2018).
53. L. Claes *et al.*, De novo mutations in the sodium-channel gene SCN1A cause severe myoclonic epilepsy of infancy. *Am J Hum Genet* **68**, 1327-1332 (2001).
54. B. M. Li *et al.*, Autism in Dravet syndrome: Prevalence, features, and relationship to the clinical characteristics of epilepsy and mental retardation. *Epilepsy and Behavior* **21**, 291-295 (2011).
55. C. Rosander, T. Hallbook, Dravet syndrome in Sweden: a population-based study. *Developmental medicine and child neurology* **57**, 628-633 (2015).
56. F. Ragona *et al.*, Dravet syndrome: Early clinical manifestations and cognitive outcome in 37 Italian patients. *Brain and Development* **32**, 71-77 (2010).
57. J. C. Mulley *et al.*, SCN1A mutations and epilepsy. *Human mutation* **25**, 535-542 (2005).
58. K. A. Mattison *et al.*, SLC6A1 variants identified in epilepsy patients reduce gamma-aminobutyric acid transport. *Epilepsia* **59**, e135-e141 (2018).
59. L. Soorya *et al.*, Prospective investigation of autism and genotype-phenotype correlations in 22q13 deletion syndrome and SHANK3 deficiency. *Molecular autism* **4**, 18 (2013).

60. S. Turkmen *et al.*, Mutations in NSD1 are responsible for Sotos syndrome, but are not a frequent finding in other overgrowth phenotypes. *Eur J Hum Genet* **11**, 858-865 (2003).
61. P. Villavicencio-Lorini *et al.*, Phenotypic variant of Brachydactyly-mental retardation syndrome in a family with an inherited interstitial 2q37.3 microdeletion including HDAC4. *Eur J Hum Genet* **21**, 743-748 (2013).
62. S. R. Williams *et al.*, Haploinsufficiency of HDAC4 causes brachydactyly mental retardation syndrome, with brachydactyly type E, developmental delays, and behavioral problems. *Am J Hum Genet* **87**, 219-228 (2010).
63. J. Grove *et al.*, Common risk variants identified in autism spectrum disorder. *bioRxiv*, (2017).
64. M. E. Hauberg *et al.*, Large-scale identification of common trait and disease variants affecting gene expression. *American journal of human genetics* **100**, 885-894 (2017).
65. C. A. de Leeuw, J. M. Mooij, T. Heskes, D. Posthuma, MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* **11**, e1004219 (2015).
66. J. Zheng *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 (Oxford, England)* **33**, 272-279 (2017).
67. D. Demontis *et al.*, Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nature genetics*, (2018).
68. S. Ripke *et al.*, Genome-wide association study identifies five new schizophrenia loci. *Nature genetics* **43**, 969--976 (2011).
69. A. Okbay *et al.*, Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539 (2016).
70. C. Schizophrenia Working Group of the Psychiatric Genomics, Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421-427 (2014).
71. S. Ripke *et al.*, A mega-analysis of genome-wide association studies for major depressive disorder. *Molecular psychiatry* **18**, 497-511 (2013).
72. N. R. Wray *et al.*, Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature genetics* **50**, 668-681 (2018).
73. L. Yengo *et al.*, Meta-analysis of genome-wide association studies for height and body mass index in ∼700000 individuals of European ancestry. *Human molecular genetics* **27**, 3641-3649 (2018).
74. B. M. Neale *et al.*, Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry* **49**, 884-897 (2010).
75. J. J. Lee *et al.*, Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature genetics* **50**, 1112-1121 (2018).
76. C. A. Rietveld *et al.*, GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* **340**, 1467-1471 (2013).
77. S. Ripke *et al.*, Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nature genetics* **45**, 1150-1159 (2013).
78. A. Reichenberg *et al.*, Discontinuity in the genetic and environmental causes of the intellectual disability spectrum. *Proceedings of the National Academy of Sciences* **113**, 1098-1103 (2016).
79. D. Pinto *et al.*, Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* **466**, 368-372 (2010).

80. E. B. Robinson *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).
81. H. J. Kang *et al.*, Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483-489 (2011).
82. T. J. Nowakowski *et al.*, Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science* **358**, 1318-1323 (2017).
83. M. J. Hawrylycz *et al.*, An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391-399 (2012).
84. N. N. Parikshak *et al.*, Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* **155**, 1008-1021 (2013).
85. X. Xu, A. B. Wells, D. R. O'Brien, A. Nehorai, J. D. Dougherty, Cell type-specific expression analysis to identify putative cellular mechanisms for neurogenetic disorders. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **34**, 1420-1431 (2014).
86. J. Chang, S. R. Gilman, A. H. Chiang, S. J. Sanders, D. Vitkup, Genotype to phenotype relationships in autism spectrum disorders. *Nature neuroscience* **18**, 191-198 (2014).
87. L. Liu *et al.*, DAWN: a framework to identify autism genes and subnetworks using gene expression and genetics. *Molecular autism* **5**, 1-18 (2014).
88. Online Mendelian Inheritance in Man, OMIM.
89. C. F. Wright *et al.*, Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. *Lancet (London, England)* **385**, 1305-1314 (2015).
90. A. Sugathan *et al.*, CHD8 regulates neurodevelopmental pathways associated with autism spectrum disorder in neural progenitors. *Proceedings of the National Academy of Sciences of the United States of America* **111**, E4468-4477 (2014).
91. J. Cotney *et al.*, The autism-associated chromatin modifier CHD8 regulates other autism risk genes during human neurodevelopment. *Nat Commun* **6**, 6404 (2015).
92. M. J. Gandal *et al.*, Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science (New York, N.Y.)* **359**, 693-697 (2018).
93. I. Voineagu *et al.*, Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* **474**, 380-384 (2011).
94. J. L. Rubenstein, M. M. Merzenich, Model of autism: increased ratio of excitation/inhibition in key neural systems. *Genes, brain, and behavior* **2**, 255-267 (2003).
95. S. B. Nelson, V. Valakh, Excitatory/inhibitory balance and circuit homeostasis in autism spectrum disorders. *Neuron* **87**, 684-698 (2015).