Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism

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,†], James S. Sutcliffe[86,87,†], Audrey Thurm[88], 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,†,#,§], Mark J. Daly[1,2,3,4,5,64,†,#,§], Joseph D. Buxbaum[7,8,9,92,93,94,†,#,§].

[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]Vanderbilt Genetics Institute, Vanderbilt University School of Medicine, Nashville, Tennessee, USA. [87]Department of Molecular Physiology and Biophsysics and Psychiatry, Vanderbilt University School of Medicine, Nashville, Tennessee, USA. [88]National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland, 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

iPSYCH consortium banner authorship

Esben Agerbo, Thomas Damm Als, Vivek Appadurai, Marie Bækved-Hansen, Rich Belliveau, Anders D. Børglum, Alfonso Buil, Jonas Bybjerg-Grauholm, Caitlin E. Carey, Felecia Cerrato, Kimberly Chambert, Claire Churchhouse, Søren Dalsgaard, Mark J. Daly, Ditte Demontis, Ashley Dumont, Jacqueline Goldstein, Jakob Grove, Christine S. Hansen, Mads Engel Hauberg, Mads V. Hollegaard, David M. Hougaard, Daniel P. Howrigan, Hailiang Huang, Julian Maller, Alicia R. Martin, Joanna Martin, Manuel Mattheisen, Jennifer Moran, Ole Mors, Preben Bo Mortensen, Benjamin M. Neale, Merete Nordentoft, Jonatan Pallesen, Duncan S. Palmer, Carsten Bøcker Pedersen, Marianne Giørtz Pedersen, Timothy Poterba, Jesper Buchhave Poulsen, Stephan Ripke, Elise B. Robinson, F. Kyle Satterstrom, Andrew J. Schork, Christine Stevens, Wesley K. Thompson, Patrick Turley, Raymond K. Walters, Thomas Werge, Emilie M. Wigdor

Summary

We present the largest exome sequencing study of autism spectrum disorder (ASD) to date (n=35,584 total samples, 11,986 with ASD). Using an enhanced Bayesian framework to integrate *de novo* and case-control rare variation, we identify 102 risk genes at a false discovery rate ≤ 0.1. Of these genes, 49 show higher frequencies of disruptive *de novo* variants in individuals ascertained for severe neurodevelopmental delay, while 53 show higher frequencies in individuals ascertained for ASD; comparing ASD cases with mutations in these groups reveals phenotypic differences. Expressed early in brain development, most of the risk genes have roles in regulation of gene expression or neuronal communication (i.e., mutations effect neurodevelopmental and neurophysiological changes), and 13 fall within loci recurrently hit by copy number variants. In human cortex single-cell gene expression data, expression of risk genes is enriched in both excitatory and inhibitory neuronal lineages, consistent with multiple paths to an excitatory/inhibitory imbalance underlying 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, affects more than 1% of individuals (Baio et al., 2018). Multiple studies have demonstrated high heritability, much of it due to common variation (Gaugler et al., 2014), although rare inherited and *de novo* variants are major contributors to individual risk (De Rubeis et al., 2014; Iossifov et al., 2014; Sanders et al., 2015). When this rare variation disrupts a gene in individuals with ASD more often than expected by chance, it implicates that gene in risk (He et al., 2013). ASD risk genes, in turn, provide insight into the underpinnings of ASD, both individually (Ben-Shalom et al., 2017; Bernier et al., 2014) and *en masse* (De Rubeis et al., 2014; Ruzzo et al., 2018; Sanders et al., 2015; Willsey et al., 2013). However, fundamental questions about the altered neurodevelopment and altered neurophysiology in ASD—including when it occurs, where, and in what cell types—remain poorly resolved.

Here we present the largest exome sequencing study in ASD to date. Through an international collaborative effort and the willingness of thousands of participating families, we assembled a cohort of 35,584 samples, including 11,986 with ASD. We introduce an enhanced Bayesian analytic framework, which incorporates recently developed gene- and variant-level scores of evolutionary constraint of genetic variation, and we use it to identify 102 ASD-associated genes (FDR $\leq$ 0.1). Because ASD is often one of a constellation of symptoms of neurodevelopmental delay (NDD), we identify subsets of the 102 ASD-associated genes that have disruptive *de novo* variants more often in NDD-ascertained or ASD-ascertained cohorts. We also consider the

cellular function of ASD-associated genes and, by examining extant data from single cells in the developing human cortex, show that their expression is enriched in maturing and mature excitatory and inhibitory neurons from midfetal development onwards, confirm their role in neuronal communication or regulation of gene expression, and show that these functions are separable. Together, these insights form an important step forward in elucidating the neurobiology of ASD.

Results

Dataset

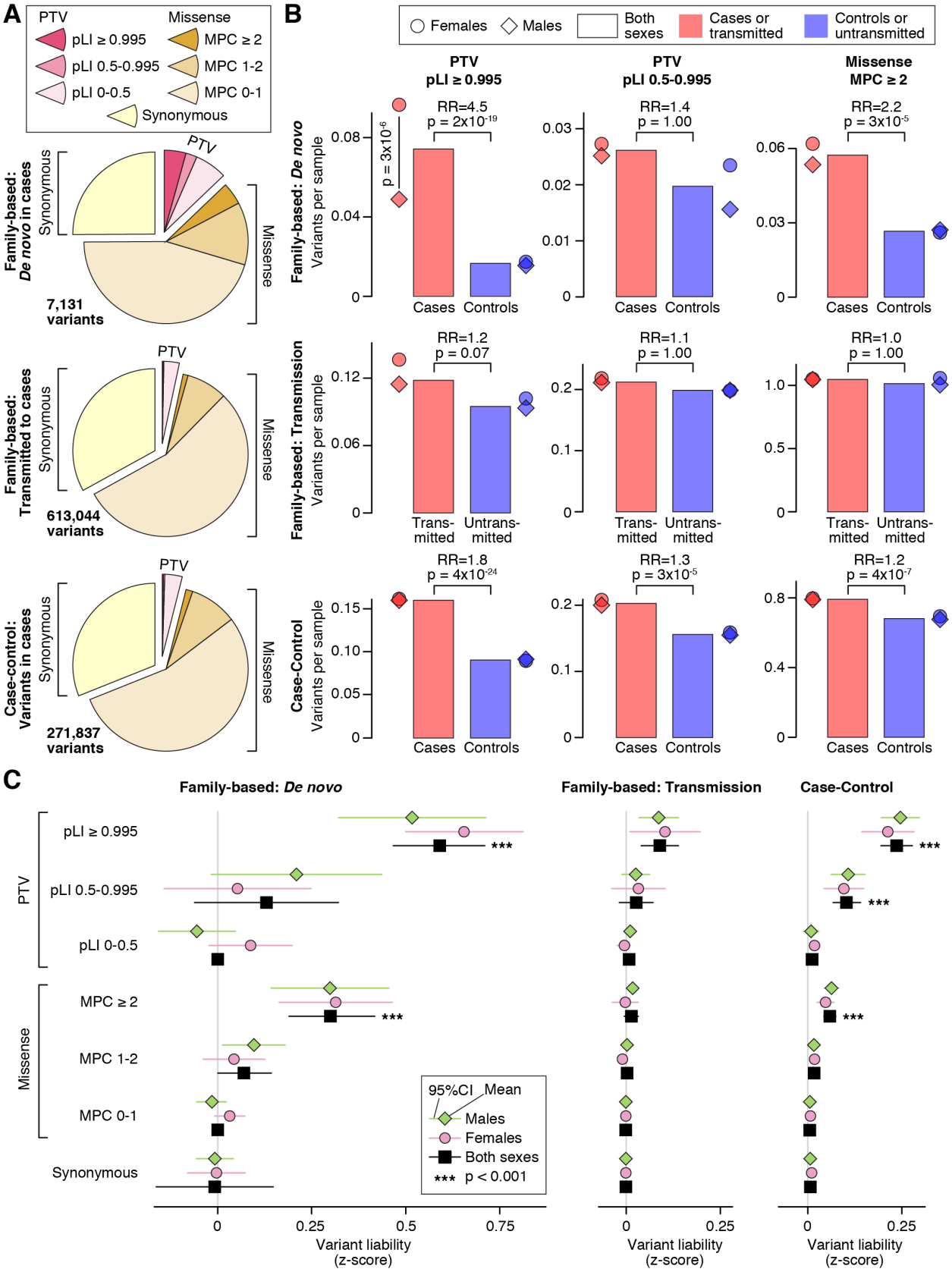
We analyzed whole-exome sequence data from 35,584 samples that passed our quality control procedures (STAR Methods). This included 21,219 family-based samples (6,430 ASD cases, 2,179 unaffected controls, and both of their parents) and 14,365 case-control samples (5,556 ASD cases, 8,809 controls) (Fig. S1; Table S1). Of these, 17,462 samples were either newly sequenced by our consortium (6,197 samples: 1,908 probands with parents; 274 ASD cases; 25 controls) or newly incorporated (11,265 samples: 416 probands with parents; 4,811 ASD cases and 5,214 controls from the Danish iPSYCH study (Satterstrom et al., 2018)).

From this cohort, we identified a set of 9,345 rare *de novo* variants in protein-coding exons (allele frequency ≤ 0.1% in our dataset as well as in the non-psychiatric subsets of the reference databases ExAC and gnomAD, with 63% of probands and 59% of unaffected offspring carrying at least one such rare coding *de novo* variant—4,073 out of 6,430 and 1,294 out of 2,179, respectively; Table S2; Fig. S1). For rare inherited and case-control analyses, we included

variants with an allele count no greater than five in our dataset and in the non-psychiatric subset of ExAC (Kosmicki et al., 2017; Lek et al., 2016).

Impact of genetic variants on ASD risk

The differential burden of genetic variants carried by cases versus controls reflects the average liability they impart for ASD. For example, because protein-truncating variants (PTVs, consisting of nonsense, frameshift, and essential splice site variants) show a greater difference in burden between ASD cases and controls than missense variants, their average impact on liability must be larger (He et al., 2013). Recent analyses have shown that measures of functional severity, such as the "probability of loss-of-function intolerance" (pLI) score (Kosmicki et al., 2017; Lek et al., 2016) and the integrated "missense badness, PolyPhen-2, constraint" (MPC) score (Samocha et al., 2017), can further delineate variant classes with higher burden. Therefore, 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; 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, expected to have minimal impact. We further divided the variants by their inheritance pattern: *de novo*, inherited, and case-control. Unlike inherited variants, *de novo* mutations are exposed to minimal selective pressure and have the potential to mediate substantial risk to disorders that limit fecundity, including ASD (Power et al., 2013). 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).


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=4.5
p = 2x10-19
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 (Christensen et al., 2016)). 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.*

Comparing probands to unaffected siblings, we observe a 3.5-fold enrichment of *de novo* PTVs in the 1,447 autosomal genes with a pLI ≥ 0.995 (366 in 6,430 cases versus 35 in 2,179 controls; 0.057 vs. 0.016 variants per sample (vps); $p=4x10^{-17}$, two-sided Poisson exact test; Fig. 1B). A less pronounced difference is observed for rare inherited PTVs in these genes, with a 1.2-fold enrichment of transmitted versus untransmitted alleles (695 transmitted versus 557 untransmitted in 5,869 parents; 0.12 vs. 0.10 vps; p=0.07, binomial exact test; 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; $p=4x10^{-24}$, binomial exact test; 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 (De Rubeis et al., 2014) (Fig.

1). However, the most severe *de novo* missense variants (MPC ≥ 2) occur at a frequency similar to *de novo* PTVs, and we observe a 2.1-fold case enrichment (354 in 6,430 cases versus 58 in 2,179 controls; 0.055 vs. 0.027 vps; $p=3x10^{-8}$, two-sided Poisson exact test; 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; $p=4x10^{-7}$, binomial exact test; Fig. 1B). Of note, in the *de novo* data, this top tier of missense variation shows stronger enrichment in cases than the middle tier of PTVs. The other two tiers of missense variation are not significantly enriched in cases (Table S3).

Sex differences in ASD risk

The prevalence of ASD is higher in males than females. In line with previous observations of females with ASD carrying a higher genetic burden than males (De Rubeis et al., 2014), we observe a 2-fold enrichment of *de novo* PTVs in highly constrained genes in affected females (n=1,097) versus affected males (n=5,333) ($p=3x10^{-6}$, two-sided Poisson exact test; Fig. 1B; Table S3). This result is consistent with the female protective effect (FPE) model, which postulates that females require an increased genetic load (in this case, high-liability PTVs) to reach the threshold for a diagnosis (Werling, 2016). 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; however, across all classes of genetic variants, we observed no significant sex differences in trait liability, consistent with the FPE model (STAR Methods; Fig. 1C). In the absence of sex-specific differences in liability, we estimated the liability z-scores for different classes of variants across both sexes together (Fig. 1C; Table S3) and leveraged them to enhance gene discovery.

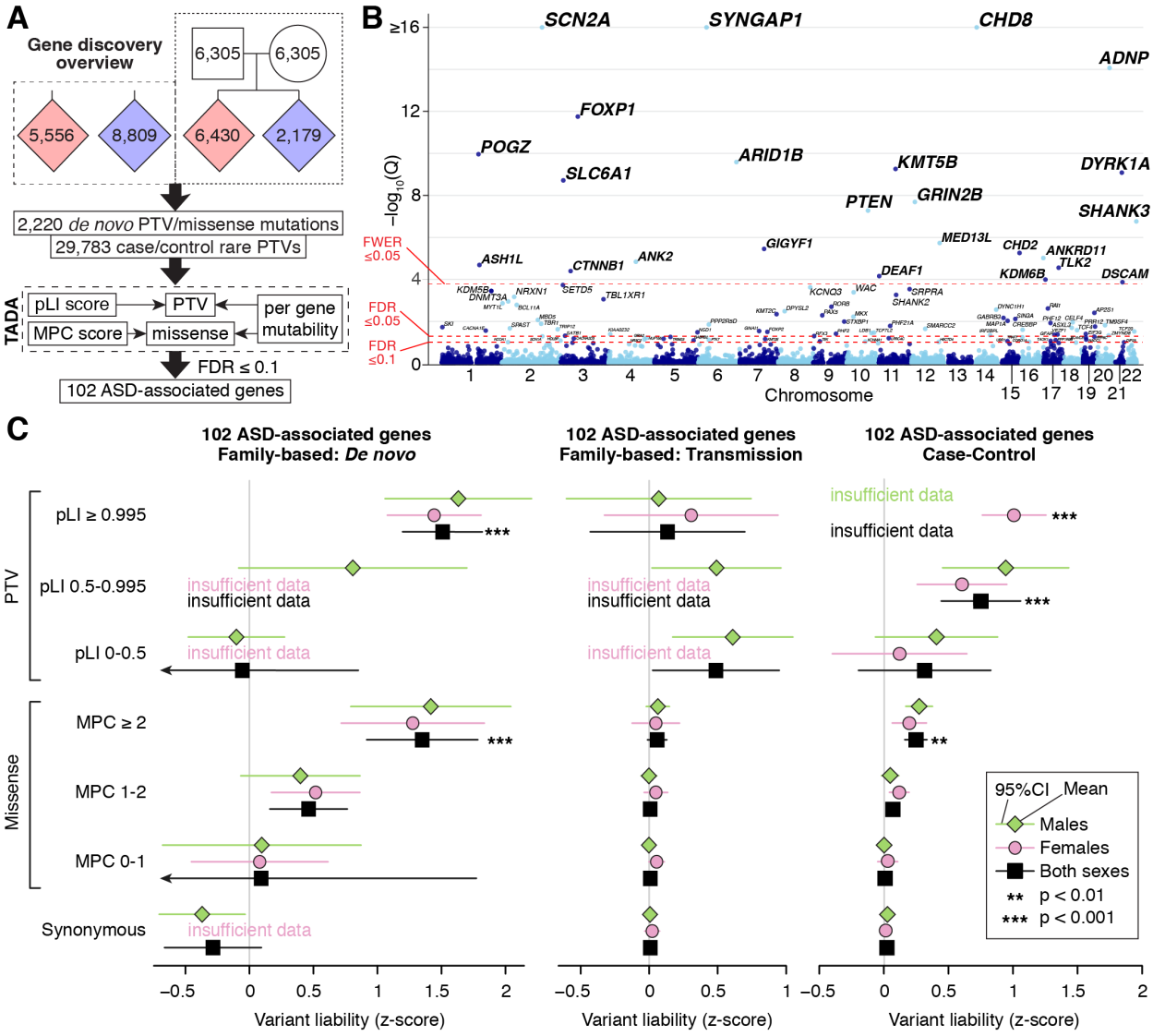


Figure 2. Gene discovery in the ASC cohort. ***A****, WES data from 35,584 samples are entered into a Bayesian analysis framework (TADA) that incorporates pLI score for PTVs and MPC score for missense variants.* ***B****, The model identifies 102 autosomal genes associated with ASD at a false discovery 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, 78 exceed the threshold of FDR ≤ 0.05 and 26 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 102 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.*

ASD gene discovery

In previous risk gene discovery efforts, we used the Transmitted And *De novo* Association (TADA) model (He et al., 2013) to integrate protein-truncating and missense variants that are *de novo*, inherited, or from case-control populations and to stratify autosomal genes by FDR for association. Here, we update the TADA model to include pLI score as a continuous metric for PTVs, and MPC score as a two-tiered metric (≥2, 1-2) for missense variants (STAR Methods; Fig. S2; Fig. S3). From family data we include *de novo* PTVs as well as *de novo* missense variants, while for case-control we include only PTVs; we do not include inherited variants due to the limited liabilities observed (Fig. 1C). These modifications result in an enhanced TADA model with greater sensitivity and accuracy than the original model (Fig. 2A; STAR Methods).

Our refined TADA model identifies 102 ASD risk genes at FDR ≤ 0.1, of which 78 meet the more stringent threshold of FDR ≤ 0.05, with 26 significant after Bonferroni correction (Fig. 2B; Table S4). By simulation experiments (Supplemental Methods), we demonstrate the reliable performance of our model, in particular showing that FDR is properly calibrated (Fig. S2). Of the 102 ASD-associated genes, 60 were not discovered by our earlier analyses (De Rubeis et al., 2014; Sanders et al., 2015), including 31 that have not been implicated in autosomal dominant neurodevelopmental disorders and were not significantly enriched for *de novo* and/or rare variants in previous studies, and that can therefore be considered novel (Table S5). The patterns of liability seen for these 102 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.

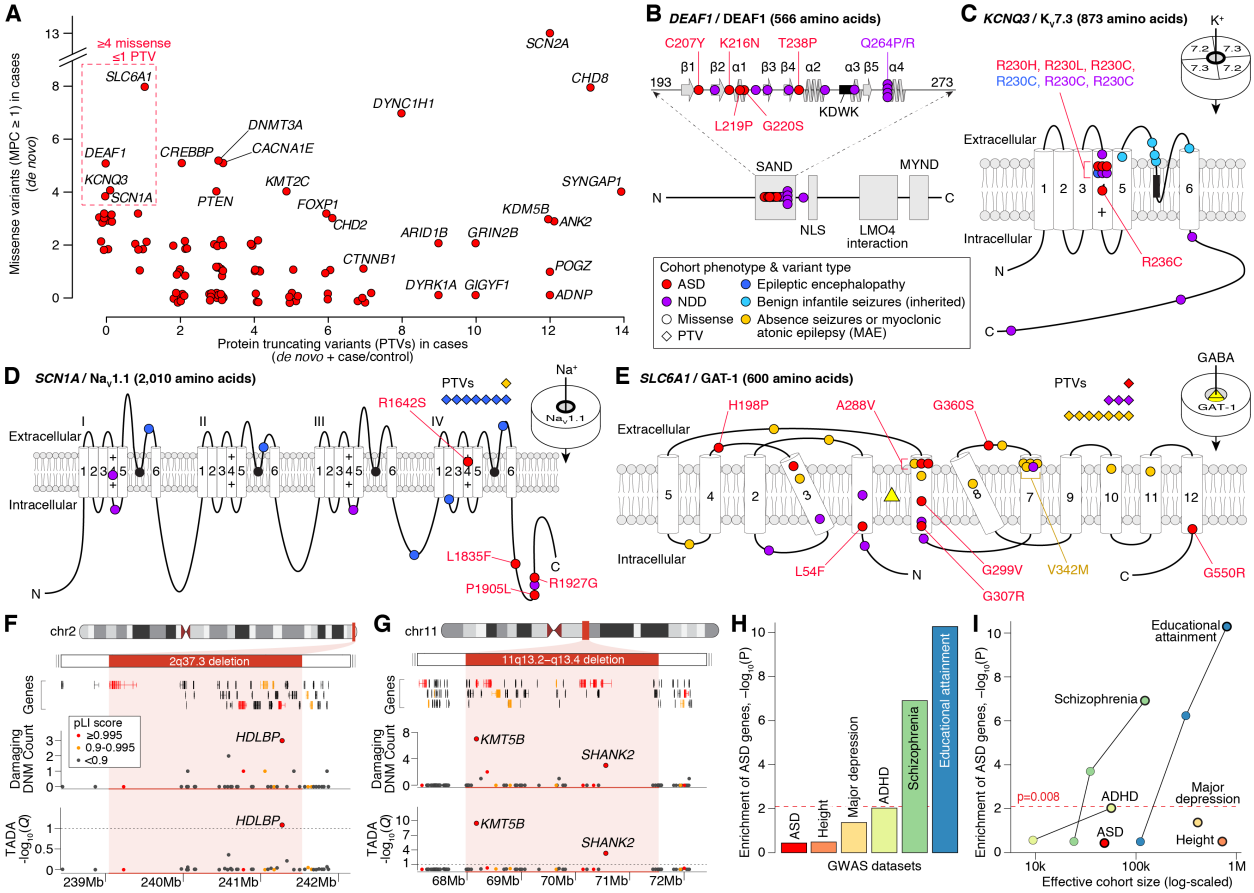


Figure 3. Genetic characterization of ASD genes. ***A****, Count of PTVs versus missense variants (MPC ≥ 1) in cases for each ASD-associated gene (red points, selected genes labeled). These counts reflect the data used by TADA for association analysis: de novo and case/control data for PTVs; de novo only for missense.* ***B****, Location of ASD de novo missense variants in DEAF1. The five ASD mutations (marked in red) are in the SAND DNA-binding domain (amino acids 193-273, spirals show alpha helices, arrows show beta sheets, KDWK is the DNA-binding motif) alongside ten variants observed in NDD, several of which have been shown to reduce DNA binding, including Q264P and Q264R (Chen et al., 2017; Heyne et al., 2018; Vulto-van Silfhout et al., 2014).* ***C****, Location of ASD missense variants in KCNQ3. 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 observed in NDD (Heyne et al., 2018). Five inherited variants observed in benign infantile seizures are shown in the pore loop (Landrum et al., 2014; Maljevic et al., 2016).* ***D****, Location of ASD missense variants in SCN1A, alongside 17 de novo variants in NDD and epilepsy (Heyne et al., 2018).* ***E****, Location of ASD missense variants in SLC6A1, alongside 31 de novo variants in NDD and epilepsy (Heyne et al., 2018; Johannesen et al., 2018).* ***F****, Subtelomeric 2q37 deletions are associated with facial dysmorphisms, brachydactyly, high BMI, neurodevelopmental delay, and ASD (Leroy et al., 2013). While three genes within the locus have a pLI score ≥ 0.995, only HDLBP is associated with ASD.* ***G****, Deletions at the 11q13.2−q13.4 locus have been observed in NDD, ASD, and otodental dysplasia (Coe et al., 2014; Cooper et al., 2011). Five genes within the locus have a pLI score ≥ 0.995, including two ASD genes: KMT5B and SHANK2.* ***H****, Assessment of gene-*

based enrichment, via MAGMA, of 102 ASD genes against genome-wide significant common variants from six GWAS. ***I****, Gene-based enrichment of 102 ASD 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.*

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, amongst which three genes have a significant excess of PTVs, relative to missense mutations, in the current dataset, based on gene mutability: *SYNGAP1*, *DYRK1A*, and *ARID1B* ($p < 0.0005$, binomial test). Because of the increased cohort size and availability of the MPC metric, we are also able for the first time to associate genes with ASD based primarily on *de novo* missense variation. We therefore examined four genes with four or more *de novo* missense variants (MPC ≥ 1) in ASD cases and one or no PTVs: *DEAF1*, *KCNQ3*, *SCN1A*, and *SLC6A1* (Fig. 3A; Table S6).

We observe five *de novo* missense variants and no PTVs in *DEAF1*, which encodes a self-dimerizing transcription factor involved in neuronal differentiation (Bottomley et al., 2001). All five missense variants reside in the SAND domain (Fig. 3B), which is critical for both dimerization and DNA binding (Bottomley et al., 2001; Jensik et al., 2004). A similar pattern of SAND domain missense enrichment is observed in individuals with intellectual disability, speech delay, and behavioral abnormalities (Chen et al., 2017; Heyne et al., 2018; Vulto-van Silfhout et al., 2014).

Four *de novo* missense variants and no PTVs are observed in *KCNQ3*, which encodes a subunit of a neuronal voltage-gated potassium channel (Fig. 3C). All four variants modify arginine

residues in the voltage-sensing fourth transmembrane domain, with three at a single residue previously characterized as gain-of-function in NDD (R230C, Fig. 3C) (Miceli et al., 2015). These data suggest gain-of-function mutations in *KCNQ3* also confer risk to ASD.

SCN1A encodes a voltage-gated sodium channel and has been associated, predominantly through PTVs, with Dravet syndrome (Claes et al., 2001), a form of progressive epileptic encephalopathy which often meets diagnostic criteria for ASD (Rosander and Hallbook, 2015). We observe four *de novo* missense variants and no PTVs in *SCN1A* (Fig. 3A; Table S6), with three located in the C-terminus (Fig. 3D), and all four cases are reported to have seizures.

The gene *SLC6A1* encodes a voltage-gated GABA transporter and has been associated with developmental delay and cognitive impairment (Deciphering Developmental Disorders, 2017; Heyne et al., 2018), as well as myoclonic atonic epilepsy and absence seizures (Johannesen et al., 2018). Here, we extend the phenotypic spectrum to include ASD, through the observation of eight *de novo* missense variants and one PTV, all in cases (Fig. 3E). Four of these variants reside in the sixth transmembrane domain, with one recurring in two independent cases (A288V). Five of the six cases with available information on history of seizure had seizures, and all four cases with available data on cognitive performance have intellectual disability.

ASD genes within recurrent copy number variants (CNVs)

Large CNVs represent another important source of risk for ASD (Sebat et al., 2007), but these genomic disorder (GD) segments can include dozens of genes, complicating the identification of driver gene(s) within these regions. We sought to determine whether the 102 ASD genes could

nominate driver 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 S7).

Within the 51 GDs, 12 GD loci encompassed 13 ASD-associated genes (Table S7), which is greater than expected by chance when simultaneously controlling for number of genes, PTV mutation rate, and brain expression levels per gene (2.3-fold increase; $p=2.3x10^{-3}$, permutation). These 12 GD loci divided into three groups: 1) the overlapping ASD gene matched the consensus driver gene, e.g., *SHANK3* for Phelan-McDermid syndrome (Soorya et al., 2013); 2) an ASD 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 in some a analyses (Williams et al., 2010); and 3) no previous driver gene had been established within the GD locus, such as *BCL11A* at 2p15-p16.1. One GD locus, 11q13.2-q13.4, had two of our 102 genes (*SHANK2* and *KMT5B*, Fig. 3G), highlighting that GDs can result from risk conferred by multiple genes, potentially including genes with small effect sizes that we are underpowered to detect.

Relationship of ASD genes with GWAS signal

Common variation plays an important role in ASD risk, and recent genome-wide association studies (GWAS) have revealed a handful of ASD-associated loci (Grove et al., 2019). Thus, we asked if common genetic variation within or near the 102 identified genes (within 10 Kb) influences ASD risk or other traits related to ASD risk. We note that, among the first five genome-wide significant ASD hits from the current largest GWAS (Grove et al., 2019), *KMT2E*

is a “double hit”—implicated by both the GWAS and the list of 102 FDR ≤ 0.1 genes described here.

We therefore ran a gene set enrichment analysis of our 102 ASD-associated genes against GWAS summary statistics using MAGMA (de Leeuw et al., 2015) 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 hyperactivity 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 human height as a negative control (Table S8) (Demontis et al., 2018; Grove et al., 2019; Lee et al., 2018; Neale et al., 2010; Okbay et al., 2016; Rietveld et al., 2013; Ripke et al., 2013a; Ripke et al., 2011; Ripke et al., 2013b; Schizophrenia Working Group of the Psychiatric Genomics, 2014; Wray et al., 2018; Yengo et al., 2018; Zheng et al., 2017). Correcting for six analyses, we observed significant enrichment only for SCZ and EA (Fig. 3H). Curiously, the ASD and ADHD GWAS signals were not enriched in the 102 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.

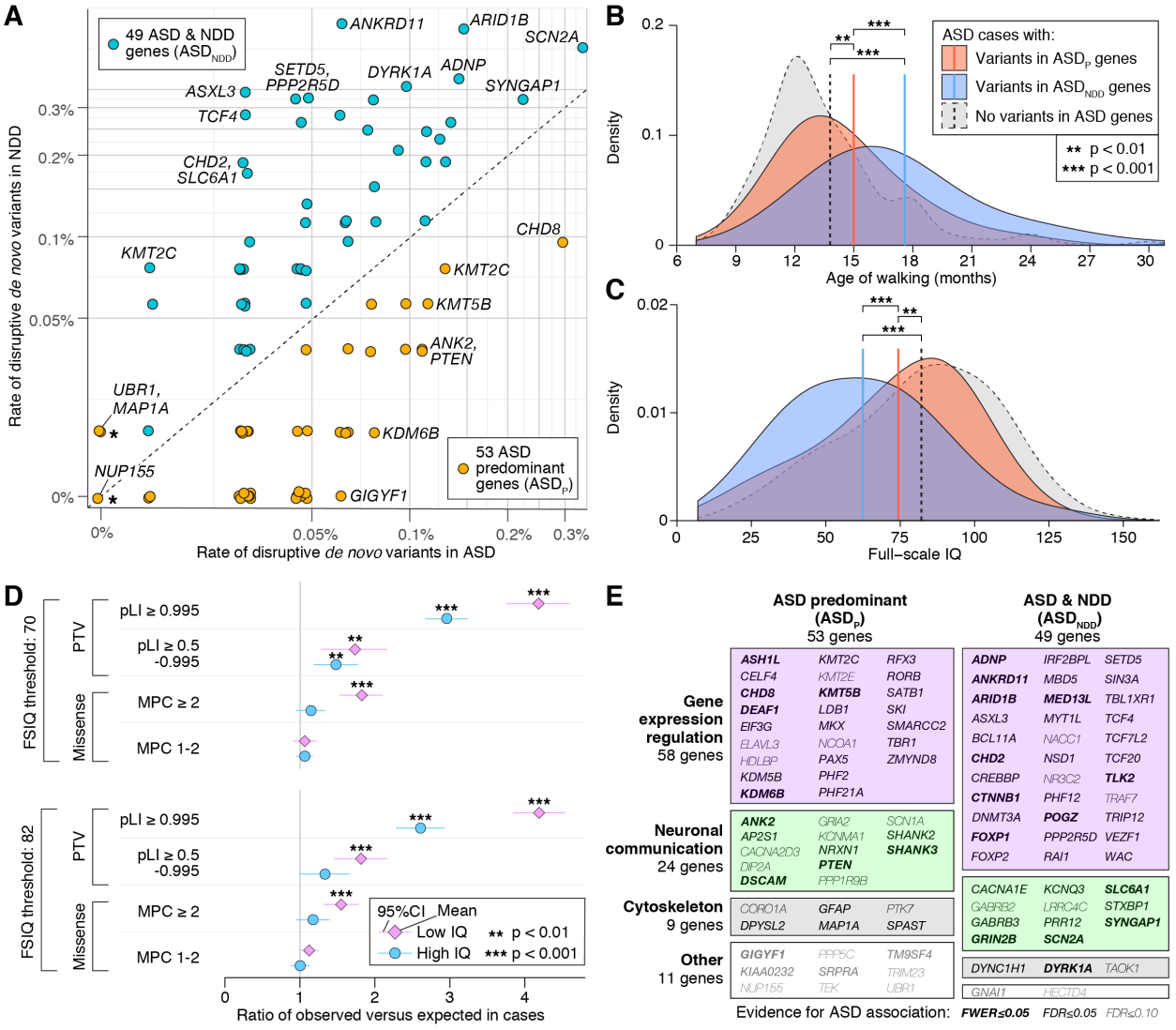


Figure 4. Phenotypic and functional categories of ASD-associated genes. A*, The frequency of disruptive de novo variants (e.g. PTVs or missense variants with MPC ≥ 1) in ASD-ascertained and NDD-ascertained cohorts (Table S4) is shown for the 102 ASD-associated genes (selected genes labeled). Fifty genes with a higher frequency in ASD are designated ASD-predominant (ASD_P), while the 49 genes more frequently mutated in NDD are designated as ASD_{NDD}. 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 mean (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 102 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.

Relationship between ASD and other neurodevelopmental disorders

Family studies yield high heritability estimates in ASD (Yip et al., 2018), but comparable estimates of heritability in severe NDD are lower (Reichenberg et al., 2016). Consistent with these observations, exome studies identify a higher frequency of disruptive *de novo* variants in severe NDD than in ASD (Deciphering Developmental Disorders, 2017). Because of the 30% co-morbidity between ASD subjects and intellectual disability/NDD, it is unsurprising that many genes are associated with both disorders (Pinto et al., 2010). Distinguishing genes that, when disrupted, lead to ASD more frequently than NDD may shed new light on how atypical neurodevelopment maps onto the core deficits of ASD.

To partition the 102 ASD genes in this manner, we compiled data from 5,264 trios ascertained for severe NDD (Table S9) and compared the relative frequency, R, of disruptive *de novo* variants (which we define as PTVs or missense variants with MPC $\geq$ 1) in ASD- or NDD-ascertained trios. Genes with $R > 1$ were classified as ASD-predominant (ASD_P, 50 genes), while those with $R < 1$ were classified as ASD with NDD (ASD_{NDD}, 49 genes). An additional three genes were assigned to the ASD_P group on the basis of case-control data, totaling 53 ASD_P genes (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, 44 such PTVs were transmitted and 18 were not (p=0.001, transmission disequilibrium test [TDT]), whereas, for ASD_{NDD} genes, 14 were transmitted and 8 were not (p=0.29; TDT). The frequency of PTVs in parents is significantly greater in ASD_P genes (1.17 per gene) than in ASD_{NDD} genes (0.45 per gene;

1 $p=6.6x10^{-6}$, binomial test), while the frequency of *de novo* PTVs in probands is not markedly
2 different between the two groups (95 in ASD_P genes, 121 in ASD_{NDD} genes, p=0.07, binomial
3 test with probability of success = 0.503 [PTV in ASD_P gene]). The paucity of inherited PTVs in
4 ASD_{NDD} genes is consistent with greater selective pressure acting against disruptive variants in
5 these genes.
6
7 Consistent with this partition, ASD subjects who carry disruptive *de novo* variants in ASD_{NDD}
8 genes walk 2.6 ± 1.2 months later (Fig. 4B; $p=2.3x10^{-5}$, t-test, df=251) and have an IQ 11.9 ± 6.0
9 points lower (Fig. 4C; $p=1.1x10^{-4}$, two-sided t-test, df=278), on average, than ASD subjects with
10 disruptive *de novo* variants in ASD_P genes (Table S10). Both sets of subjects differ significantly
11 from the rest of the cohort with respect to IQ and age of walking (Fig. 4B, 4C; Fig. S4; Table
12 S10). Thus, the data support some overall distinction between the genes identified in ASD and
13 NDD *en masse*, although our current analyses are not powered for variant-level or gene-level
14 resolution.
15
16 ***Burden of mutations in ASD as a function of IQ***
17 Of the 6,430 ASD probands, 3,010 had a detected *de novo* variant and either a recorded full-scale
18 IQ or a clinical assessment of ID. We partitioned these subjects into those with IQ ≥ 70 (69.4%)
19 versus those with IQ < 70 (30.6%), then characterized the burden of *de novo* variants within
20 these groups. ASD subjects in the lower IQ group carry a greater burden of *de novo* variants,
21 relative to both expectation and the high IQ group, in the two top tiers of PTVs and the top tier of
22 missense variants (Fig. 4D). Excess burden relative to expectation is also observed in the two top
23 PTV tiers for the high IQ group (Fig. 4D). Similar patterns were observed partitioning the

sample at IQ ≥ 82 (53.7%) versus IQ < 82 (46.3%), which was the mean IQ after removing subjects who carry disruptive variants in the 102 ASD genes (Fig. 4C). Finally, we observe excess burden in the high IQ group when considering the 102 ASD genes only, as documented by model-driven simulations accounting for selection bias due to an FDR threshold (STAR Methods). Thus, excess burden is not limited to low IQ cases, supporting the idea that *de novo* variants do not solely impair cognition (Robinson et al., 2014).

Functional dissection of ASD genes

Past WES analyses have identified two major functional groups of ASD genes: those involved in gene expression regulation (GER), including chromatin regulators and transcription factors, and those involved in neuronal communication (NC), including synaptic function (De Rubeis et al., 2014). A simple gene ontology enrichment analysis with the new list of 102 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}). We used a combination of gene ontology and primary literature research to assign additional genes to the GER (58 genes) and NC (24 genes) categories for further analyses (STAR Methods; Table S11; Fig. 4E). We also see the emergence of a new functional group of nine genes implicated in category GO:0007010 "cytoskeleton organization". The remaining 11 genes are described as "Other" (Table S11 and Fig. 4E), many of which have roles in signaling cascades and/or ubiquitination.

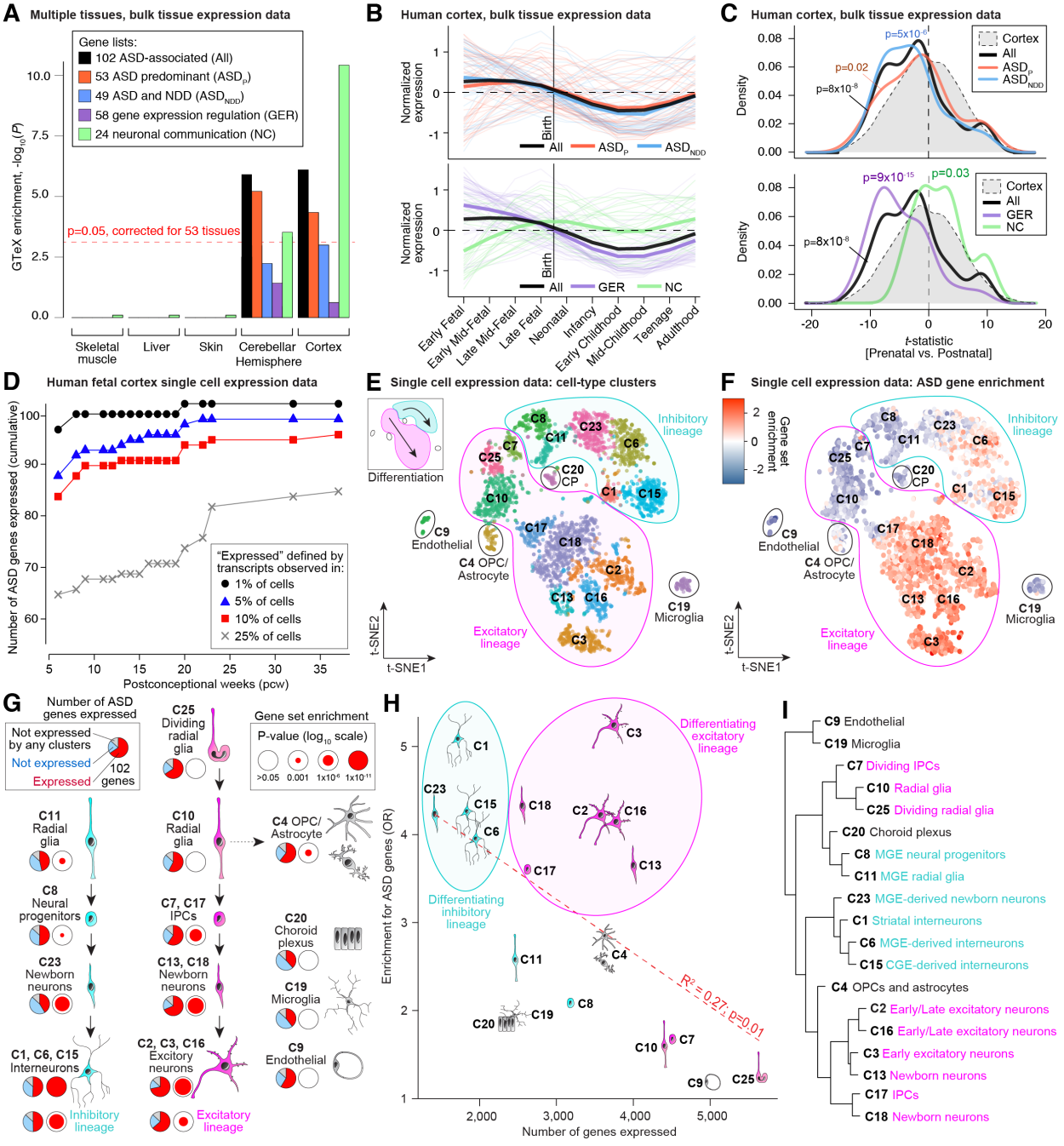


Figure 5. Analysis of 102 ASD-associated genes in the context of gene expression data. A, *GTEx bulk RNA-seq data from 53 tissues was processed to identify genes enriched in specific tissues. Gene set enrichment was performed for the 102 ASD genes and four subsets (ASD$_P$, ASD$_{NDD}$, GER, NC) for each tissue. Five representative tissues are shown here, including cortex, which has the greatest degree of enrichment (OR=3.7; p=2.6×10^{-6}).* ***B****, BrainSpan bulk RNA-seq data across 10 developmental stages was used to plot the normalized expression of the 101 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 101 ASD-associated genes (excluding PAX5, which is not expressed in the cortex)*

shows a prenatal bias ($p=8x10^{-8}$) for GER genes ($p=9\times10^{-15}$), while NC genes are postnatally biased (p=0.03). ***D****, The cumulative number of ASD-associated genes expressed in RNA-seq data for 4,261 cells collected from human forebrain across prenatal development.* ***E****, t-SNE analysis identifies 19 clusters with unambiguous cell type in this single-cell expression data.* ***F****, The enrichment of the 102 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 are 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 102 ASD-associated genes observed in at least 25% of cells within the cluster is shown by the pie chart, while the log-transformed Bonferroni corrected 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, H, I, Fisher's Exact Test.*

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

The 102 ASD genes can thus be subdivided by functional role (58 GER genes, 24 NC genes) and phenotypic impact (53 ASD_P genes, 49 ASD_{NDD} genes) to give five gene sets (including the set of all 102). We first evaluated enrichment of these five gene sets in the 53 tissues with bulk RNA-seq data in the Genotype-Tissue Expression (GTEx) resource (GTEx-Consortium, 2017). To enhance tissue-specific resolution, we selected genes that were expressed in one tissue at a significantly higher level than the remaining 52 tissues, specifically log fold-change > 0.5 and FDR < 0.05 (*t*-test). 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 multiple-testing threshold of $p \leq 9x10^{-4}$, reflecting 53 tissues, enrichment was observed in 11 of the 13 brain regions, with the strongest enrichment in cortex (∩=30 genes; $p=3\times10^{-6}$; OR=3.7; Fig. 5A) and cerebellar hemisphere (∩=48 genes; $p=3\times10^{-6}$; OR=2.9; Fig. 5A). Of the four gene subsets, NC genes were the most highly enriched in cortex (∩=17/23

genes; $p=3\times10^{-11}$; OR=25; Fig. 5A), while GER genes were the least enriched (∩=10/58 genes; p=0.36; OR=1.8; Fig. 5A).

We next leveraged BrainSpan human neocortex bulk RNA-seq data (Li et al., 2018) to assess enrichment of ASD genes across development (Fig. 5B, 5C). Of the 17,484 autosomal protein-coding genes assessed for association, 13,735 (78.5%) were expressed in the cortex (RPKM ≥ 0.5 in 80% of samples of at least one cortical region and developmental period). Of the 102 ASD genes, only the cerebellar transcription factor *PAX5* (FDR=0.005, TADA) was not expressed in the cortex (78 expected; $p=1x10^{-9}$, binomial test). Compared to other genes expressed in the cortex, the remaining 101 ASD genes are expressed at higher levels during prenatal 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 expressed genes. Using this metric, the 101 cortically-expressed ASD genes showed enrichment in the prenatal cortex ($p=8\times10^{-8}$, 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 ($p=5\times10^{-6}$, Wilcoxon test; Fig. 5C). The GER genes reach their highest levels during early to late fetal development (Fig. 5B) with a marked prenatal bias ($p=9\times10^{-15}$, Wilcoxon test; Fig. 5C), while the NC genes are highest between late midfetal development and infancy (Fig. 5B) and show a trend towards postnatal bias (p=0.03, Wilcoxon test; Fig. 5C). Thus, in keeping with prior analyses (Chang et al., 2014; Parikshak et al., 2013; Willsey et al., 2013; Xu et al., 2014), the ASD genes are expressed at high levels in human cortex 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 inhibitory and excitatory neurons

Prior analyses have implicated excitatory glutamatergic neurons in the cortex and medium spiny neurons in the striatum in ASD (Chang et al., 2014; Parikshak et al., 2013; Willsey et al., 2013; Xu et al., 2014). Here, we perform a more direct assessment, examining expression of the 102 ASD-associated genes in an existing single-cell RNA-seq dataset of 4,261 cells from the prenatal human forebrain (Nowakowski et al., 2017), ranging from 6 to 37 post-conception weeks (pcw) with an average of 16.3 pcw (Table S12).

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 genes were expressed at the earliest developmental stages (e.g. 68 of 102 at 13 pcw), the most dramatic increase in the number of genes expressed occurred during midfetal development (70 by 19 pcw, rising to 81 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 genes expressed (Fig. 5D), but the patterns of expression were similar across definitions and when considering gene function or cell type (Fig. S5).

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 (Nowakowski et al., 2017) (Fig. 5E, Table S12), 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. By contrasting one cell type to the others, we observed enrichment for the 102 ASD genes in maturing and mature neurons of the excitatory and inhibitory lineages (Fig. 5F, 5G) but not in non-neuronal lineages. Early excitatory neurons (C3) expressed the most ASD genes ($\cap$=72 genes, OR=5.0, $p < 1\times10^{-10}$, Fisher's exact test [FET]), while choroid plexus (C20) and microglia (C19) expressed the fewest ASD genes ($\cap$=39 genes, p=0.09 and 0.137, respectively, FET); 14 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 ($\cap$=72 and $\cap$=51 genes, $p < 1\times10^{-10}$, FET; Fig. 5F, 5G; Table S12). 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, paralleling 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; $\cap$=62 genes, OR=2.8, $p=8\times10^{-5}$, FET). Of the 60 ASD genes expressed in OPCs, 58 overlapped with radial glia, which may reflect shared developmental origins rather than an independent enrichment signal. In contrast to post-mortem findings in adult ASD brains (Gandal et al., 2018; Voineagu et al., 2011), no enrichment was observed in microglia. To validate 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).

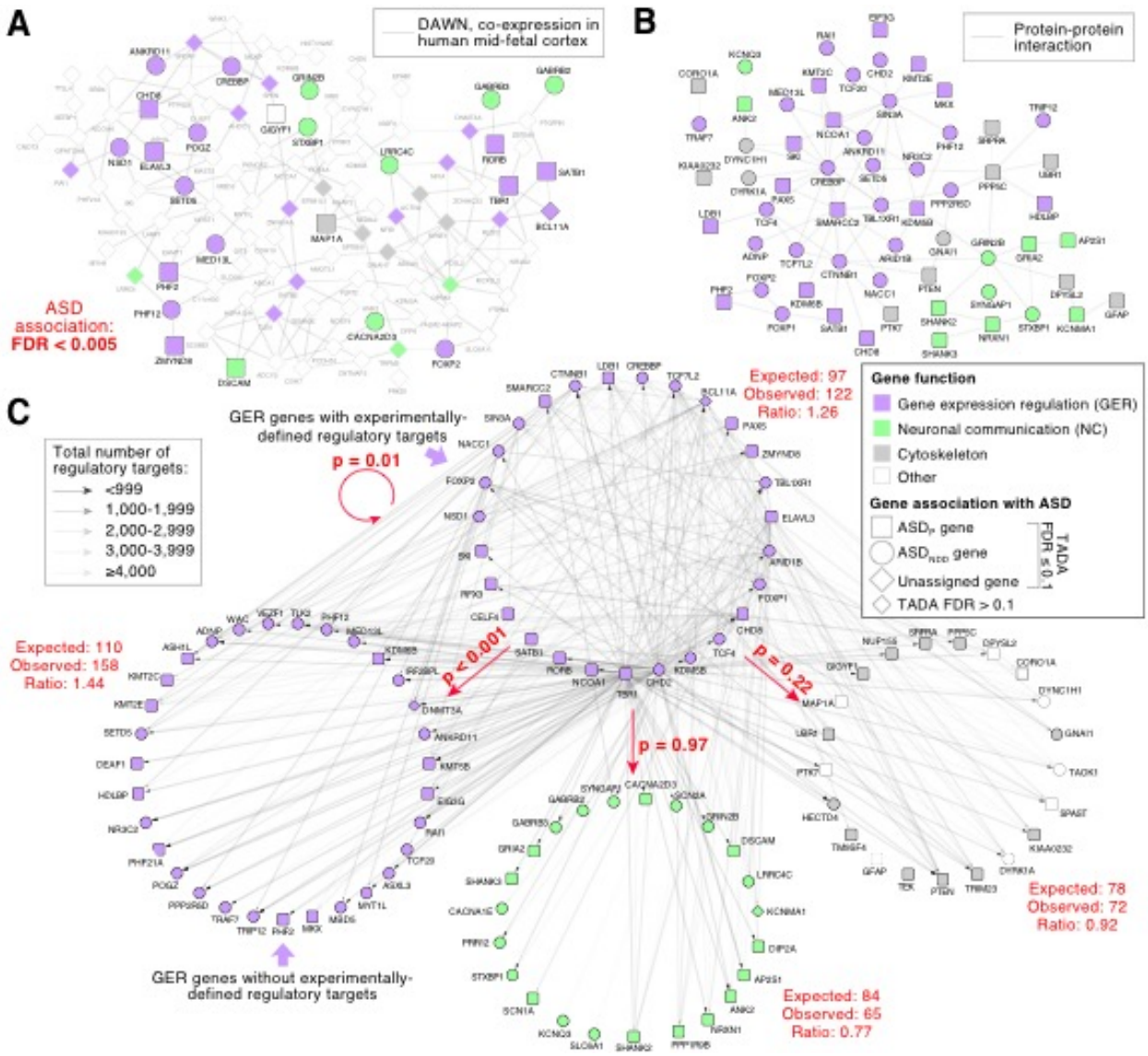


Figure 6. Functional relationships of ASD risk genes. A*, ASD association data from TADA (Table S4) is integrated with co-expression data from the midfetal human brain to implicate additional genes in ASD using DAWN (Discovering Association With Networks). The top 138 genes that share edges are shown (FDR ≤ 0.005).* ***B****, ASD-associated genes form a single protein-protein interaction network with more edges than expected by chance (p=0.01).* ***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.

Functional relationships among ASD genes and prediction of novel risk genes

The ASD genes show convergent functional roles (Fig. 4E) and expression patterns in the 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. We have previously developed the Discovering Association With Networks (DAWN) approach to integrate TADA scores of genetic association and gene co-expression data in order to uncover more risk genes. By using the TADA results and BrainSpan gene co-expression data from the midfetal human cortex, DAWN yields 138 genes (FDR $\leq$ 0.005), including 83 genes that are not captured by TADA, with 69 of these 83 correlated with many other genes (Fig. 6A; Table S13). Of the 83, 19 are implicated in neurodevelopmental disorders, and seven of these have autosomal recessive inheritance (Table S5, S13). Of the 138 genes, 38 are expressed in excitatory cell types ($p < 1.6\text{x}10^{-4}$, FET), 25 are also expressed in inhibitory cell types ($p < 7.9\times10^{-4}$, FET), and many play a role in GER or NC (Fig. 6A).

To interpret gene co-expression and enrichment across a broader range of early developmental samples, we used Weighted Gene Coexpression Network Analysis (WGCNA) to assess spatiotemporal co-expression from 177 high-quality BrainSpan samples aged 8 pcw to 1 year. WGCNA yielded 27 early developmental co-expression modules, two of which show significant over-representation of 102 ASD genes after correction for multiple testing (Fig. S6, Table S14):

M4 for the NC gene set (∩=5 genes, OR=13.7, p=0.002, FET); and M25 for all 102 ASD genes (∩=17 genes, OR=12.1, $p=3x10^{-11}$, FET), although driven by the GER gene set (∩=17 genes, OR=26.2, $p=9x10^{-16}$, FET). With regard to single-cell gene expression, genes in the NC-specific M4 showed greatest enrichment in maturing 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). The DAWN associated genes are enriched in M3 (∩=10, OR=10.1, $p=5x10^{-6}$, FET) and M25 (∩=7, OR=5.9, p=0.004, FET) but not M4, although expression of genes in M4 are highly correlated with those of M3 during early development, and both are highly expressed prenatally. Comparing our gene modules to previously published candidate ASD gene networks obtained using WGCNA (Parikshak et al., 2013) shows that our M4 strongly overlaps with previously identified M16 ($p=5.5x10^{-69}$, FET), and our M24 overlaps with previously identified M2 ($p=1.3x10^{-239}$, FET) (note, however that both studies make use of BrainSpan data so the overlap is not unexpected but helps to relate the results from the two studies).

To explore whether GER and NC gene sets interact more than would be expected by chance, we analyzed protein-protein interaction (PPI) networks (Fig. 6B; Table S15) and found they do not: there was a significant excess of interactions among all ASD genes (∩=82 genes, *p*=0.02, FET), GER genes (∩=49 genes, *p*=0.006, FET), and NC genes (∩=12 genes, *p*=0.03, FET), but not among GER and NC genes (∩=2 genes, *p*=1.00, FET). Nor do GER genes regulate the NC genes, according to our analyses, although GER-GER regulation was enriched (Table S16, Fig. 6C). Even *CHD8*, the most prominent and well characterized ASD GER gene, did not regulate NC genes more than expected by chance (Fig. S7).

Discussion

By characterizing rare *de novo* and inherited coding variation from 35,584 individuals, including 11,986 ASD cases, we implicate 102 genes in risk for ASD at FDR ≤ 0.1 (Fig. 2), of which 31 are novel risk genes. Notably, analyses of this set of risk genes lead to novel genetic, phenotypic, and functional findings. Evidence for several of the 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). Further, we strengthen evidence for driver genes in genomic disorder loci and we propose a new driver gene (*BCL11A*) for the recurrent CNV at 2p15-p16.1. By evaluating GWAS results for ASD and related phenotypes and asking whether their common variant association signals overlap significantly with the 102 risk genes, we find substantial enrichment of GWAS signal for two traits genetically correlated with ASD—schizophrenia and educational attainment. For ASD itself, however, this enrichment is not significant, likely due to the limited power of the ASD GWAS. Despite this, *KMT2E* is significantly associated with ASD by both common and rare risk variation.

We perform a genetic partition between genes predominantly conferring liability for ASD (ASD_P) and genes imparting risk to both ASD and NDD (ASD_{NDD}). Two 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 (Fig. 4B, 4C); second, we find that inherited variation plays a lesser role in ASD_{NDD} than in ASD_P genes. 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 strong negative selection (e.g. *ANKRD11*, *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*, *ANK2*). This distinction has important ramifications for clinicians, geneticists, and neuroscientists, because 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. Larger cohorts will be required to reliably identify specific genes as being enriched in ASD compared to NDD.

Single-cell gene expression data from the developing human cortex implicate mid-to-late fetal development and maturing and mature neurons in both excitatory and inhibitory lineages (Fig. 5). Expression of GER genes shows a prenatal bias, while expression of NC genes does not. Placing these results in the context of multiple non-exclusive hypotheses around the origins of ASD, it is intriguing to speculate that the NC ASD genes provide compelling support for E/I imbalance in ASD (Rubenstein and Merzenich, 2003) through direct impact on neurotransmission. However, as there was no support for a regulatory role for GER ASD genes on either NC or cytoskeletal ASD genes, additional mechanisms, having to do with cell migration and neurodevelopment, also appear to be at play. This might suggest that GER ASD genes impact E/I balance by altering the numbers of excitatory and inhibitory neurons in given regions of the brain. ASD must arise by phenotypic convergence amongst these diverse neurobiological trajectories, and further dissecting the nature of this convergence, especially in

the genes that we have identified herein, 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 NIMH (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.)), NHGRI (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 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

Baio, J., Wiggins, L., Christensen, D.L., Maenner, M.J., Daniels, J., Warren, Z., Kurzius-Spencer, M., Zahorodny, W., Robinson Rosenberg, C., White, T.*, et al.* (2018). Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years - Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014. MMWR Surveill Summ *67*, 1-23.

Ben-Shalom, R., Keeshen, C.M., Berrios, K.N., An, J.Y., Sanders, S.J., and Bender, K.J. (2017). 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.

Bernier, R., Golzio, C., Xiong, B., Stessman, H.A., Coe, B.P., Penn, O., Witherspoon, K., Gerdts, J., and Baker, C.a. (2014). Disruptive CHD8 mutations define a subtype of autism early in development. Cell *158*, 263-276.

Bottomley, M.J., Collard, M.W., Huggenvik, J.I., Liu, Z., Gibson, T.J., and Sattler, M. (2001). The SAND domain structure defines a novel DNA-binding fold in transcriptional regulation. Nature structural biology *8*, 626-633.

Chang, J., Gilman, S.R., Chiang, A.H., Sanders, S.J., and Vitkup, D. (2014). Genotype to phenotype relationships in autism spectrum disorders. Nature Neuroscience *18*, 191-198.

Chen, L., Jensik, P.J., Alaimo, J.T., Walkiewicz, M., Berger, S., Roeder, E., Faqeih, E.A., Bernstein, J.A., Smith, A.C.M., Mullegama, S.V.*, et al.* (2017). Functional analysis of novel DEAF1 variants identified through clinical exome sequencing expands DEAF1-associated neurodevelopmental disorder (DAND) phenotype. Human mutation *38*, 1774-1785.

Christensen, D.L., Baio, J., Braun, K.V.N., Bilder, D., Charles, J., Constantino, J.N., Daniels, J., Durkin, M.S., Fitzgerald, R.T., Kurzius-Spencer, M.*, et al.* (2016). 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.

Claes, L., Del-Favero, J., Ceulemans, B., Lagae, L., Van Broeckhoven, C., and De Jonghe, P. (2001). De novo mutations in the sodium-channel gene SCN1A cause severe myoclonic epilepsy of infancy. American journal of human genetics *68*, 1327-1332.

Coe, B.P., Witherspoon, K., Rosenfeld, J.A., van Bon, B.W., Vulto-van Silfhout, A.T., Bosco, P., Friend, K.L., Baker, C., Buono, S., Vissers, L.E.*, et al.* (2014). Refining analyses of copy

number variation identifies specific genes associated with developmental delay. Nature genetics *46*, 1063-1071.

Cooper, G.M., Coe, B.P., Girirajan, S., Rosenfeld, J.A., Vu, T.H., Baker, C., Williams, C., Stalker, H., Hamid, R., Hannig, V.*, et al.* (2011). A copy number variation morbidity map of developmental delay. Nature genetics *43*, 838-846.

de Leeuw, C.A., Mooij, J.M., Heskes, T., and Posthuma, D. (2015). MAGMA: generalized gene-set analysis of GWAS data. PLoS computational biology *11*, e1004219.

De Rubeis, S., He, X., Goldberg, A.P., Poultney, C.S., Samocha, K., Cicek, A.E., Kou, Y., Liu, L., Fromer, M., and Walker, S. (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. Nature *515*, 209-215.

Deciphering Developmental Disorders, S. (2017). Prevalence and architecture of de novo mutations in developmental disorders. Nature *542*, 433-438.

Demontis, D., Walters, R.K., Martin, J., Mattheisen, M., Als, T.D., Agerbo, E., Baldursson, G., Belliveau, R., Bybjerg-Grauholm, J., Bækvad-Hansen, M.*, et al.* (2018). Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nature genetics.

Gandal, M.J., Haney, J.R., Parikshak, N.N., Leppa, V., Ramaswami, G., Hartl, C., Schork, A.J., Appadurai, V., Buil, A., Werge, T.M.*, et al.* (2018). Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. Science (New York, NY) *359*, 693-697.

Gaugler, T., Klei, L., Sanders, S.J., Bodea, C.A., Goldberg, A.P., Lee, A.B., Mahajan, M., Manaa, D., Pawitan, Y., Reichert, J.*, et al.* (2014). Most genetic risk for autism resides with common variation. Nature genetics *46*, 881-885.

Grove, J., Ripke, S., Als, T.D., Mattheisen, M., Walters, R.K., Won, H., Pallesen, J., Agerbo, E., Andreassen, O.A., Anney, R.*, et al.* (2019). Identification of common genetic risk variants for autism spectrum disorder. Nature genetics.

GTEx-Consortium (2017). Genetic effects on gene expression across human tissues. Nature *550*, 204-213.

He, X., Sanders, S.J., Liu, L., De Rubeis, S., Lim, E.T., Sutcliffe, J.S., Schellenberg, G.D., Gibbs, R.A., Daly, M.J., Buxbaum, J.D.*, et al.* (2013). Integrated model of de novo and inherited genetic variants yields greater power to identify risk genes. PLoS genetics *9*, e1003671.

Heyne, H.O., Singh, T., Stamberger, H., Abou Jamra, R., Caglayan, H., Craiu, D., De Jonghe, P., Guerrini, R., Helbig, K.L., Koeleman, B.P.C., *et al.* (2018). De novo variants in neurodevelopmental disorders with epilepsy. Nature genetics *50*, 1048-1053.

Iossifov, I., O'Roak, B.J., Sanders, S.J., Ronemus, M., Krumm, N., Levy, D., Stessman, H.A., Witherspoon, K.T., Vives, L., Patterson, K.E., *et al.* (2014). The contribution of de novo coding mutations to autism spectrum disorder. Nature *515*, 216-221.

Jensik, P.J., Huggenvik, J.I., and Collard, M.W. (2004). 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.

Johannesen, K.M., Gardella, E., Linnankivi, T., Courage, C., Saint Martin, A., Lehesjoki, A.-E., Mignot, C., Afenjar, A., Lesca, G., Abi-Warde, M.-T., *et al.* (2018). Defining the phenotypic spectrum of SLC6A1 mutations. Epilepsia *59*, 389-402.

Kosmicki, J.A., Samocha, K.E., Howrigan, D.P., Sanders, S.J., Slowikowski, K., Lek, M., Karczewski, K.J., Cutler, D.J., Devlin, B., Roeder, K., *et al.* (2017). Refining the role of de novo protein-truncating variants in neurodevelopmental disorders by using population reference samples. Nature genetics *49*, 504-510.

Landrum, M.J., Lee, J.M., Riley, G.R., Jang, W., Rubinstein, W.S., Church, D.M., and Maglott, D.R. (2014). ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic acids research *42*, D980-D985.

Lee, J.J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., Nguyen-Viet, T.A., Bowers, P., Sidorenko, J., Karlsson Linnér, R., *et al.* (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nature genetics *50*, 1112-1121.

Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., *et al.* (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285-291.

Leroy, C., Landais, E., Briault, S., David, A., Tassy, O., Gruchy, N., Delobel, B., Gregoire, M.J., Leheup, B., Taine, L., *et al.* (2013). The 2q37-deletion syndrome: an update of the clinical spectrum including overweight, brachydactyly and behavioural features in 14 new patients. European journal of human genetics : EJHG *21*, 602-612.

Li, M., Santpere, G., Imamura Kawasawa, Y., Evgrafov, O.V., Gulden, F.O., Pochareddy, S., Sunkin, S.M., Li, Z., Shin, Y., Zhu, Y., *et al.* (2018). Integrative functional genomic analysis of human brain development and neuropsychiatric risks. Science *362*.

Maljevic, S., Vejzovic, S., Bernhard, M.K., Bertsche, A., Weise, S., Dcker, M., Lerche, H., Lemke, J.R., Merkenschlager, A., and Syrbe, S. (2016). Novel KCNQ3 mutation in a large family with benign familial neonatal epilepsy: A rare cause of neonatal seizures. Molecular Syndromology *7*, 189-196.

Miceli, F., Soldovieri, M.V., Ambrosino, P., De Maria, M., Migliore, M., Migliore, R., and Taglialatela, M. (2015). 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.

Neale, B.M., Medland, S.E., Ripke, S., Asherson, P., Franke, B., Lesch, K.P., Faraone, S.V., Nguyen, T.T., Schafer, H., Holmans, P., *et al.* (2010). Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. Journal of the American Academy of Child and Adolescent Psychiatry *49*, 884-897.

Nowakowski, T.J., Bhaduri, A., Pollen, A.A., Alvarado, B., Mostajo-Radji, M.A., Di Lullo, E., Haeussler, M., Sandoval-Espinosa, C., Liu, S.J., Velmeshev, D., *et al.* (2017). Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. Science *358*, 1318-1323.

Okbay, A., Beauchamp, J.P., Fontana, M.A., Lee, J.J., Pers, T.H., Rietveld, C.A., Turley, P., Chen, G.-B., Emilsson, V., Meddens, S.F.W., *et al.* (2016). Genome-wide association study identifies 74 loci associated with educational attainment. Nature *533*, 539.

Parikshak, N.N., Luo, R., Zhang, A., Won, H., Lowe, J.K., Chandran, V., Horvath, S., and Geschwind, D.H. (2013). Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. Cell *155*, 1008-1021.

Pinto, D., Pagnamenta, A.T., Klei, L., Anney, R., Merico, D., Regan, R., Conroy, J., Magalhaes, T.R., Correia, C., Abrahams, B.S., *et al.* (2010). Functional impact of global rare copy number variation in autism spectrum disorders. Nature *466*, 368-372.

Power, R.A., Kyaga, S., Uher, R., MacCabe, J.H., Lngstrm, N., Landen, M., McGuffin, P., Lewis, C.M., Lichtenstein, P., and Svensson, A.C. (2013). Fecundity of patients with

schizophrenia, autism, bipolar disorder, depression, anorexia nervosa, or substance abuse vs their unaffected siblings. JAMA psychiatry *70*, 22-30.

Reichenberg, A., Cederlöf, M., McMillan, A., Trzaskowski, M., Kapara, O., Fruchter, E., Ginat, K., Davidson, M., Weiser, M., Larsson, H.*, et al.* (2016). Discontinuity in the genetic and environmental causes of the intellectual disability spectrum. Proceedings of the National Academy of Sciences *113*, 1098-1103.

Rietveld, C.A., Medland, S.E., Derringer, J., Yang, J., Esko, T., Martin, N.W., Westra, H.J., Shakhbazov, K., Abdellaoui, A., Agrawal, A.*, et al.* (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. Science *340*, 1467-1471.

Ripke, S., O'Dushlaine, C., Chambert, K., Moran, J.L., Kahler, A.K., Akterin, S., Bergen, S.E., Collins, A.L., Crowley, J.J., Fromer, M.*, et al.* (2013a). Genome-wide association analysis identifies 13 new risk loci for schizophrenia. Nature genetics *45*, 1150-1159.

Ripke, S., Sanders, A.R., Kendler, K.S., Levinson, D.F., Sklar, P., Holmans, P.A., Lin, D.-Y.Y., Duan, J., Ophoff, R.A., Andreassen, O.A.*, et al.* (2011). Genome-wide association study identifies five new schizophrenia loci. Nature genetics *43*, 969--976.

Ripke, S., Wray, N.R., Lewis, C.M., Hamilton, S.P., Weissman, M.M., Breen, G., Byrne, E.M., Blackwood, D.H., Boomsma, D.I., Cichon, S.*, et al.* (2013b). A mega-analysis of genome-wide association studies for major depressive disorder. Molecular psychiatry *18*, 497-511.

Robinson, E.B., Samocha, K.E., Kosmicki, J.A., McGrath, L., Neale, B.M., Perlis, R.H., and Daly, M.J. (2014). Autism spectrum disorder severity reflects the average contribution of de novo and familial influences. Proc Natl Acad Sci U S A *111*, 15161-15165.

Rosander, C., and Hallbook, T. (2015). Dravet syndrome in Sweden: a population-based study. Developmental medicine and child neurology *57*, 628-633.

Rubenstein, J.L., and Merzenich, M.M. (2003). Model of autism: increased ratio of excitation/inhibition in key neural systems. Genes, brain, and behavior *2*, 255-267.

Ruzzo, E.K., Perez-Cano, L., Jung, J.-Y., Wang, L.-k., Kashef-Haghighi, D., Hartl, C., Hoekstra, J., Leventhal, O., Gandal, M.J., Paskov, K.*, et al.* (2018). Whole genome sequencing in multiplex families reveals novel inherited and de novo genetic risk in autism. bioRxiv.

Samocha, K.E., Kosmicki, J.A., Karczewski, K.J., O'Donnell-Luria, A.H., Pierce-Hoffman, E., MacArthur, D.G., Neale, B.M., and Daly, M.J. (2017). Regional missense constraint improves variant deleteriousness prediction. bioRxiv.

Sanders, S.J., He, X., Willsey, A.J., Ercan-Sencicek, A.G., Samocha, K.E., Cicek, A.E., Murtha, M.T., Bal, V.H., Bishop, S.L., Dong, S.*, et al.* (2015). Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. Neuron *87*, 1215-1233.

Satterstrom, F.K., Walters, R.K., Singh, T., Wigdor, E.M., Lescai, F., Demontis, D., Kosmicki, J.A., Grove, J., Stevens, C., Bybjerg-Grauholm, J.*, et al.* (2018). ASD and ADHD have a similar burden of rare protein-truncating variants. bioRxiv.

Schizophrenia Working Group of the Psychiatric Genomics, C. (2014). Biological insights from 108 schizophrenia-associated genetic loci. Nature *511*, 421-427.

Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., Yamrom, B., Yoon, S., Krasnitz, A., Kendall, J.*, et al.* (2007). Strong association of de novo copy number mutations with autism. Science (New York, NY) *316*, 445-449.

Soorya, L., Kolevzon, A., Zweifach, J., Lim, T., Dobry, Y., Schwartz, L., Frank, Y., Wang, A.T., Cai, G., Parkhomenko, E.*, et al.* (2013). Prospective investigation of autism and genotype-phenotype correlations in 22q13 deletion syndrome and SHANK3 deficiency. Mol Autism *4*, 18.

Voineagu, I., Wang, X., Johnston, P., Lowe, J.K., Tian, Y., Horvath, S., Mill, J., Cantor, R.M., Blencowe, B.J., and Geschwind, D.H. (2011). Transcriptomic analysis of autistic brain reveals convergent molecular pathology. Nature *474*, 380-384.

Vulto-van Silfhout, A.T., Rajamanickam, S., Jensik, P.J., Vergult, S., de Rocker, N., Newhall, K.J., Raghavan, R., Reardon, S.N., Jarrett, K., McIntyre, T.*, et al.* (2014). Mutations affecting the SAND domain of DEAF1 cause intellectual disability with severe speech impairment and behavioral problems. American journal of human genetics *94*, 649-661.

Werling, D.M. (2016). The role of sex-differential biology in risk for autism spectrum disorder. Biology of Sex Differences *7*, 1-18.

Williams, S.R., Aldred, M.A., Der Kaloustian, V.M., Halal, F., Gowans, G., McLeod, D.R., Zondag, S., Toriello, H.V., Magenis, R.E., and Elsea, S.H. (2010). Haploinsufficiency of HDAC4 causes brachydactyly mental retardation syndrome, with brachydactyly type E, developmental delays, and behavioral problems. American journal of human genetics *87*, 219-228.

Willsey, A.J., Sanders, S.J., Li, M., Dong, S., Tebbenkamp, A.T., Muhle, R.A., Reilly, S.K., Lin, L., Fertuzinhos, S., Miller, J.A.*, et al.* (2013). Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. Cell *155*, 997-1007.

Wray, N.R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E.M., Abdellaoui, A., Adams, M.J., Agerbo, E., Air, T.M., Andlauer, T.M.F., *et al.* (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nature genetics *50*, 668-681.

Xu, X., Wells, A.B., O'Brien, D.R., Nehorai, A., and Dougherty, J.D. (2014). 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.

Yengo, L., Sidorenko, J., Kemper, K.E., Zheng, Z., Wood, A.R., Weedon, M.N., Frayling, T.M., Hirschhorn, J., Yang, J., Visscher, P.M., *et al.* (2018). 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.

Yip, B.H.K., Bai, D., Mahjani, B., Klei, L., Pawitan, Y., Hultman, C.M., Grice, D.E., Roeder, K., Buxbaum, J.D., Devlin, B., *et al.* (2018). Heritable variation, with little or no maternal effect, accounts for recurrence risk to autism spectrum disorder in Sweden. Biological psychiatry *83*, 589-597.

Zheng, J., Erzurumluoglu, A.M., Elsworth, B.L., Kemp, J.P., Howe, L., Haycock, P.C., Hemani, G., Tansey, K., Laurin, C., Pourcain, B.S., *et al.* (2017). LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. Bioinformatics *33*, 272-279.