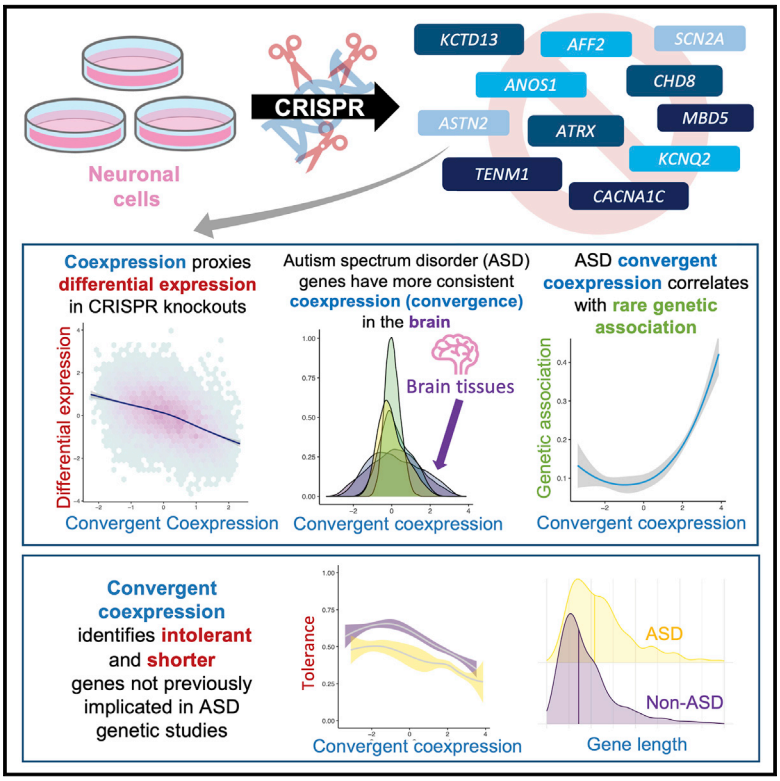
Cell Genomics

Article

Convergent coexpression of autism-associated genes suggests some novel risk genes may not be detectable in large-scale genetic studies

Graphical abstract



Highlights

- Coexpression from postmortem brain tissue proxies CRISPR perturbation in neurons
- ASD genes implicated from rare variant association are transcriptionally convergent
- Coexpression convergence tends to be tissue specific for disease
- Identified intolerant and shorter ASD risk genes that can be missed in genetic studies

Authors

Calwing Liao, Mariana Moyses-Oliveira, Celine E.F. De Esch, ..., Kristen J. Brennand, Michael E. Talkowski, Douglas M. Ruderfer

Correspondence

douglas.ruderfer@vanderbilt.edu

In brief

Liao et al. demonstrate that coexpression can proxy the transcriptional effects of CRISPR perturbation and highlight that autism-associated genes have convergent coexpression in the brain. This convergent coexpression tends to be tissue specific for disease and may help identify intolerant and shorter novel risk genes that can be missed in genetic studies.

Liao et al., 2023, Cell Genomics *3*, 100277
April 12, 2023
https://doi.org/10.1016/j.xgen.2023.100277

CellPress

Cell Genomics

CellPress
OPEN ACCESS

Article

Convergent coexpression of autism-associated genes suggests some novel risk genes may not be detectable in large-scale genetic studies

Calwing Liao,[1,2] Mariana Moyses-Oliveira,[3,4,5] Celine E.F. De Esch,[3,4,5] Riya Bhavsar,[3,4,5] Xander Nuttle,[3,4,5] Aiqun Li,[6,7,8,9,10] Alex Yu,[6,7,8] Nicholas D. Burt,[3,4,5] Serkan Erdin,[3,4] Jack M. Fu,[3,4,5] Minghui Wang,[6,7,8] Theodore Morley,[11] Lide Han,[11] CommonMind Consortium, Patrick A. Dion,[2,14] Guy A. Rouleau,[2,14] Bin Zhang,[6,7,8] Kristen J. Brennand,[6,7,8,9,10,12] Michael E. Talkowski,[3,4,5] and Douglas M. Ruderfer[11,13,15,*]

[1]Department of Human Genetics, McGill University, Montreal, QC, Canada
[2]Montreal Neurological Institute-Hospital, McGill University, Montreal, QC, Canada
[3]Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA 02114, USA
[4]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[5]Department of Neurology, Harvard Medical School, Boston, MA, 02114, USA
[6]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[7]Mount Sinai Center for Transformative Disease Modeling, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[8]Icahn Institute for Data Science and Genomic Technology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[9]Nash Family Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[10]Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[11]Division of Genetic Medicine, Department of Medicine, Vanderbilt Genetics Institute, Vanderbilt University Medical Center, 1211 Medical Center Dr., Nashville, TN 37232, USA
[12]Department of Psychiatry, Yale University New Haven, New Haven, CT 06511, USA
[13]Department of Biomedical Informatics and Department of Psychiatry and Behavioral Sciences, Vanderbilt University Medical Center, 1211 Medical Center Dr., Nashville, TN 37232, USA
[14]Department of Neurology and Neurosurgery, McGill University, Montreal, QC, Canada
[15]Lead contact
*Correspondence: douglas.ruderfer@vanderbilt.edu
https://doi.org/10.1016/j.xgen.2023.100277

SUMMARY

Autism spectrum disorder (ASD) is a heritable neurodevelopmental disorder characterized by deficits in social interactions and communication. Protein-altering variants in many genes have been shown to contribute to ASD; however, understanding the convergence across many genes remains a challenge. We demonstrate that coexpression patterns from 993 human postmortem brains are significantly correlated with the transcriptional consequences of CRISPR perturbations in human neurons. Across 71 ASD risk genes, there was significant tissue-specific convergence implicating synaptic pathways. Tissue-specific convergence was further demonstrated across schizophrenia and atrial fibrillation risk genes. The degree of ASD convergence was significantly correlated with ASD association from rare variation and differential expression in ASD brains. Positively convergent genes showed intolerance to functional mutations and had shorter coding lengths than known risk genes even after removing association with ASD. These results indicate that convergent coexpression can identify potentially novel genes that are unlikely to be discovered by sequencing studies.

INTRODUCTION

Autism spectrum disorder (ASD) is a highly heritable neuropsychiatric disorder with a population prevalence of ~1%.[1] Sequencing studies have implicated dozens of genes contributing to risk of ASD based on an excess of rare deleterious variations in cases compared with controls.[2–5] These findings have highlighted biological pathways implicated in ASD, including synaptic function, chromatin, and transcriptional regulation.[2,3] Many of these pathways have since been corroborated by transcriptomic studies.[6–8] However, our understanding of how these genes may interact and whether they converge on shared downstream pathways with potential to implicate novel risk genes remains incomplete.

Genetic perturbation studies involving induced pluripotent stem cells (iPSCs) and CRISPR offer insight into context-specific cellular and transcriptomic consequences of perturbing ASD-associated genes individually, highlighting downstream genes or pathways that might be relevant for disease risk.[9,10] Loss-of-function (LoF) models of *CHD8* and *FOXP1*, two transcriptional regulators with strong association with ASD, have shown dysregulation of multiple other ASD genes.[9,11] Although several

Please cite this article in press as: Liao et al., Convergent coexpression of autism-associated genes suggests some novel risk genes may not be detectable in large-scale genetic studies, Cell Genomics (2023), https://doi.org/10.1016/j.xgen.2023.100277

ASD risk genes have been shown to have a direct molecular relationships with each other (e.g., *CHD8* regulates and binds to other ASD risk genes like CTCF[9,12]), many altered genes lacked CHD8 and FOXP1 binding sites, suggesting that the effects of the perturbations can be propagated by downstream regulatory interactions.[9,11] In *Xenopus tropicalis*,[13] LoF CRISPR-mediated genome editing of 10 ASD-implicated genes resulted in an increased ratio of neural progenitor cells to neurons in the telencephalon, pointing to a convergent cellular outcome. Additional work using CRISPR-Cas9 to knock out ASD genes followed by single-cell RNA sequencing in the developing mouse brain identified recurrent glial and neuronal modules, again pointing to cellular convergence.[14] Another study used a multiplexed iPSC platform and introduced frameshift mutations in 27 ASD genes. The ASD mutations were classified into two subgroups based on alterations in prefrontal cortex neurogenesis. which then correlated with abnormal WNT signaling; one group inhibited and one group enhanced spontaneous cortical neurogenesis,[15] confirming the existence of convergent cellular and signaling phenotypes within this larger subset of ASD-associated genes.

Although these functional studies have shed some light on the interplay between multiple disease genes, determining convergent signatures across the large number of implicated ASD genes in human cell lines remains a technical and logistical challenge. These efforts are limited by the substantial time and cost required to generate human iPSCs, differentiate into a given cell type, edit the genes of interest, and investigate downstream consequences. Analysis of protein-protein interaction (PPI) networks and coexpression modules of ASD risk genes has shown that there is high connectivity, suggesting that there is more direct interaction than expected by chance.[16–20] Convergence in gene expression profiles has been demonstrated not only among ASD risk genes affected by rare protein truncation variants but also for genes associated with ASD by large copy number variants (CNVs)[21,22] and common variation.[23] Similarly, ASD risk genes have been shown to be highly coexpressed in the developing cortex.[18,24] Prior work has layered expression data onto to the genetic findings to predict novel ASD risk genes,[17] successfully identifying many genes that are now significant in larger genetic studies. The high degree of connectivity across ASD risk genes presents an opportunity to leverage coexpression to better understand molecular convergence in ASD. We hypothesized that coexpression from a relevant tissue would provide a meaningful proxy for the transcriptional effects of CRISPR perturbation, enabling an analysis of transcriptional convergence across many ASD risk genes to implicate novel shared downstream genes and pathways. Despite the incredible success in identifying genes for ASD, the power of genetic studies is dependent on observing variants frequently enough to ensure a statistically significant excess in cases. Therefore, power will be limited by shorter gene length or extreme genic intolerance. Implicating genes through transcriptional convergence presents one potential path to mitigating this limitation.

Here, we investigated the level of coexpression convergence—similarity in coexpression profiles—among ASD risk genes by leveraging large-scale postmortem brain tissue datasets. We first showed that differential expression from CRISPR-mediated knockdown or activation is significantly and directionally correlated with coexpression of the perturbed gene, suggesting that coexpression can proxy differential expression. Across a set of high-confidence ASD risk genes from sequencing studies, we meta-analyzed their coexpression patterns to show highly significant coexpression convergence in ASD-relevant tissue. The risk genes from schizophrenia and atrial fibrillation were also only convergent in disease-specific tissue (i.e., convergent coexpression of atrial fibrillation genes in only heart tissue). We identified genes that were significantly coexpressed with the 71 ASD genes that were enriched for (1) neurodevelopmental disorder (NDD) genes even after excluding previous ASD-associated genes, (2) differential expression in postmortem brains of ASD patients compared with controls, and (3) synaptic function. Finally, we were able to implicate genes as novel ASD risk candidates, including smaller genes and those highly intolerant to LoF variation, which precluded their identification from current genetic studies. Overall, *in silico* coexpression convergence approaches as used here could be important time- and cost-effective additions to our understanding of disease biology.

RESULTS

Characterization of CRISPR-mediated perturbations

RNA sequencing data from 17 CRISPR perturbation experiments in neurons comprising 15 unique genes were included from three independent sources (STAR Methods). In total, the CRISPR experiments included 12 CRISPR-mediated heterozygous or homozygous LoF mutational models (*AFF2*, *ANOS1*, *ASTN2*, *ATRX*, *CACNA1C*, *CHD8*, *KCTD13*, *KCNQ2*, *MBD5*, *SCN2A [2x]*, and *TENM1*),[10,25] four CRISPR activation models (*FURIN*, *SNAP91*, *TSNARE1*, and *CLCN3*),[26] and one CRISPR interference model (*SCN2A*). These genes were selected based on data availability and involvement or implication in neurodevelopmental disorders. We tested for differential expression between the CRISPR-edited cell lines and the unedited cell lines for each gene passing quality control, including observing the expected effect in the perturbed gene (STAR Methods). Each experiment was analyzed separately, and the differential expression effect sizes were converted to *Z* scores representing "CRISPR perturbation." The experiments had variable impact on global expression patterns, with the total number of significantly differentially expressed genes identified ranging from 164–4,857 and lambda inflation factors ranging from 0.29–4.13 (Table 1).

One gene (*SCN2A*) was perturbed in three independent experiments, allowing us to quantify the variability of global gene expression changes because of CRISPR perturbation and genetic background. We found significant but modest Pearson correlations of differential expression across experiments that varied from 0.202 between the two CRISPR-mediated gene editing experiments to 0.343 and 0.335 between the CRISPR interference (CRISPRi) experiment and the two gene editing experiments (Figure S1).

Gene coexpression correlates with downstream transcriptional consequences of CRISPR perturbation

We hypothesized that gene coexpression would proxy differential expression driven by CRISPR perturbations in the

Table 1. CRISPR modeling experimental design and numbers of differentially expressed genes per experiment

Gene	Type of CRISPR perturbation	Number of edited lines	Number of unedited lines	Number of significant negative genes ($p < 0.05$)	Number of significant positive genes ($p < 0.05$)	Number of significant genes ($p < 0.05$)	Lambda
AFF2[a]	hem gene edit	5	4	151	146	297	0.48
ANOS1[a]	hem gene edit	4	4	122	42	164	0.29
ASTN2[a]	hom gene edit	3	4	455	173	628	0.75
ATRX[a]	hem gene edit	4	4	991	966	1,957	1.39
CACNA1C[a]	hom gene edit	3	4	166	132	298	0.57
CHD8[a]	het gene edit	4	4	112	107	249	0.5
KCNQ2[a]	hom gene edit	4	4	381	329	710	1.29
KCTD13	het gene edit	6	5	2,167	2,690	4,857	4.13
MBD5	het gene edit	6	3	81	177	258	0.72
SCN2A	het gene edit	20	10	573	1,137	1,710	1.84
SCN2A[a]	hom gene edit	4	4	143	80	223	0.53
SCN2A	CRISPRi	2	2	2,062	2,525	4,587	2.13
TENM[a]	hem gene edit	5	4	1,114	1,280	2,394	2.33
CLCN3	CRISPRa	6	6	1,049	605	1,654	1.31
FURIN	CRISPRa	2	2	399	268	667	0.54
SNAP91	CRISPRa	2	2	434	764	1,198	1.11
TSNARE1	CRISPRa	2	2	626	825	1,451	1.07

hem, hemizygous; hom, homozygous; het, heterozygous.
[a]Dataset described in Deneault et al.[10]

experiments described above. We leveraged 993 postmortem brain samples from the dorsolateral prefrontal cortex (DLPFC) and calculated pairwise gene coexpression using Pearson correlation (STAR Methods). We identified consistent and significant negative correlation between differential expression because of CRISPR-mediated gene knockdown and the corresponding perturbed gene's normalized coexpression profile ranging from −0.089 to −0.45 (Figure 1A). The LOESS regression curve for each panel highlights that significantly differentially expressed genes tend to have a stronger negative correlation. The exception was the *ANOS1* LoF model, which had a nominally significant positive correlation ($R = 0.018$, $p = 0.046$). Overall, genes that are more significantly downregulated by a specific CRISPR perturbation are more highly and positively coexpressed with the perturbed gene. This holds true when CRISPR targets are downregulated via LoF mutation or CRISPRi as well as when they are upregulated using CRISPR activation (CRISPRa) (Figure S2).

With a variably sized but consistently significant negative relationship seen across single gene CRISPR perturbation and coexpression, we next asked whether meta-analyzing differential expression from multiple CRISPR perturbations could also be proxied by a meta-analysis of the coexpression profiles of the perturbed genes. In other words, we wanted to assess whether genes that were consistently differentially expressed in the same direction across multiple gene perturbations (convergent genes) could be inferred from a similar convergence of coexpression of the respective genes from postmortem brain tissue. When we separately meta-analyzed all CRISPR-mediated gene edits and corresponding coexpression profiles for the perturbed genes, we found that convergent CRISPR perturbation was significantly negatively correlated with convergent coexpression (Pearson's $R = -0.44$, $p < 1 \times 10^{-300}$; Figure 1B).

Given the significant correlation between CRISPR perturbation and coexpression, we next assessed how that compared with the correlation of differential expression from different CRISPR perturbations of the same gene (*SCN2A*) across three different experiments. Significant Pearson correlation was observed between *SCN2A* coexpression and differential expression in each CRISPR experiment of *SCN2A* ranging from −0.25 to −0.43 (Figure 2). The correlation statistics across CRISPR *SCN2A* experiments (0.2–0.34) were similar to those seen when compared with coexpression, suggesting that coexpression provided a similar proxy to transcriptional dysregulation as an independent CRISPR experiment. After meta-analyzing the differential expression profiles from the three *SCN2A* CRISPR experiments, the correlation with coexpression was stronger than each individual experiment alone (Pearson's $R = -0.45$, $p < 1 \times 10^{-300}$).

In silico functional convergence of ASD risk genes

We next sought to leverage coexpression to test convergence of the 71 risk genes implicated from the most recent ASD exome sequencing study.[5] After calculating meta-analysis effect sizes of Fisher's transformed Pearson's correlations, we assessed significance by running 10,000,000 permutations where the coexpression of 71 randomly selected genes (among those with coexpression) were meta-analyzed (STAR Methods). Significant

Please cite this article in press as: Liao et al., Convergent coexpression of autism-associated genes suggests some novel risk genes may not be detectable in large-scale genetic studies, Cell Genomics (2023), https://doi.org/10.1016/j.xgen.2023.100277

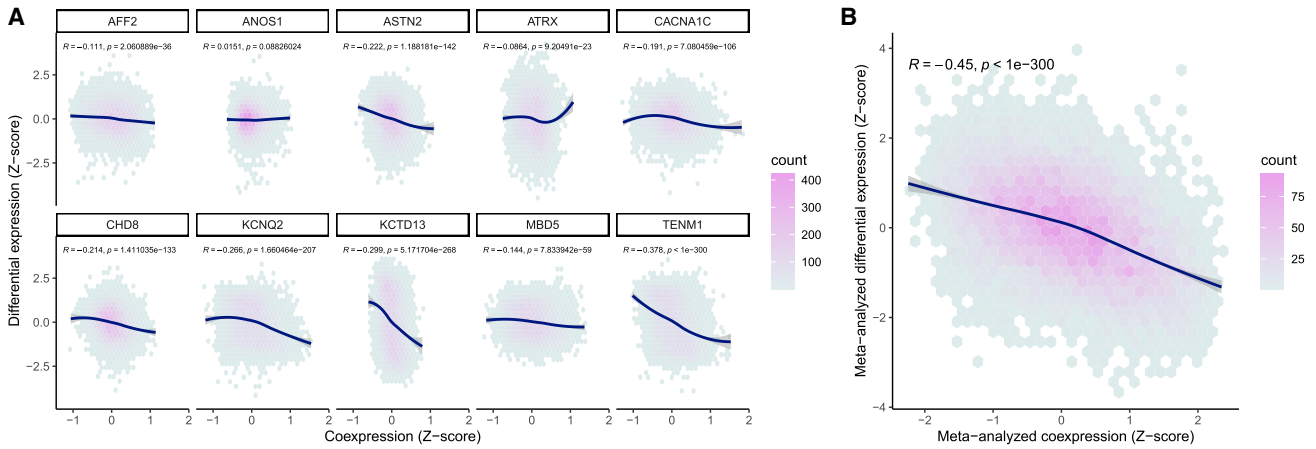


Figure 1. Coexpression proxies CRISPR-induced differential expression

(A) Correlation between differential expression profiles for 10 CRISPR knockdown models (y axis) with respective coexpression profiles (x axis) for each gene (points).

(B) Correlation of meta-analyzed CRISPR knockouts with respective meta-analyzed coexpression of perturbed genes. Coexpression is represented as a Fisher transformed Pearson's correlation *Z* score. A Pearson correlation was done to assess the correlation between coexpression and differential expression. The curve for each panel was fitted with a locally weighted smoothing (LOESS) regression.

overall convergence was identified by comparing the variance of the distribution of convergent coexpression of the ASD genes to the variance of the permuted random genes (ASD variance = 1.70, mean permuted null variance = 0.131, $p = 9.9 \times 10^{-7}$). A total of 3,553 genes were found to be significantly convergently coexpressed at a Bonferroni-adjusted empirical p value of less than 0.01. Nearly 60% of those (n = 2,157) were positively convergent, meaning that their expression was consistently in the same direction as the ASD genes. Comparison with convergence defined using the rank-based Spearman's correlation yielded consistent results and near-total correlation (Spearman's rank correlation, rho = 0.995, $p < 1 \times 10^{-300}$). Next, we relaxed the false discovery rate (FDR) threshold for ASD genes to an FDR of less than 0.05 and still found significant levels of convergence ($ASD_{185\ genes}$ variance = 1.23), with 2,842 genes with a Bonferroni-adjusted empirical p value of less than 0.01 and a positive *Z* score.

The analysis that identified the 71 ASD genes directly incorporated constraint against LoF variation (LoF observed/expected upper bound fraction [LOEUF][27]) into estimates of prior relative risk, which could inflate coexpression convergence among intolerant genes. To account for this potential confounding, we reran the permutation procedure, randomly selecting 71 genes with matching LOEUF scores to the ASD genes. The p values between intolerance-matched permutation and not were strongly correlated (Spearman's rank correlation, rho = 0.93, $p < 1 \times 10^{-300}$), suggesting that the models of ASD risk gene discovery did not significantly inflate discovery of convergent genes (Figure S3). We similarly adjusted with coding sequence length and found strong correlation as well (Spearman's rank correlation, rho = 0.925, $p < 1 \times 10^{-300}$). Further, an analysis based on a Fisher's combined test using only *de novo* and missense variants without incorporating intolerance identified 49 ASD genes (FDR < 0.001) that showed no significant difference in intolerance scores compared with the 71 ASD genes (two-sided Wilcoxon test, W = 1825.5, p = 0.648). Given that intolerance did not seem to confound our analyses, we meta-analyzed the initial 71 genes for downstream analyses.

Convergent genes were overrepresented in gene sets related to neuronal and synaptic function with the most significant pathways across categories, including the synapse ($p = 1.79 \times 10^{-21}$), synaptic signaling ($p = 6.23 \times 10^{-19}$), neuronal system ($p = 1.74 \times 10^{-13}$), and abnormal CNS synaptic transmission ($p = 3.03 \times 10^{-13}$) (Table S1). Convergent genes also represented 127 of the 410 genes implicated in neurodevelopmental disorders from DisGeNet[28] ($p = 2.96 \times 10^{-7}$) and 71 of the 193 genes in the cation channel complex ($p = 1.45 \times 10^{-7}$). Next, we partitioned the significantly convergent genes by direction of effect and found that positively convergent genes were largely enriched in synaptic pathways (Table S2), whereas negatively convergent genes had a much less clear pattern of enrichment across pathways (Table S3).

Convergence captures ASD signals from exomes and postmortem brain studies

Given the strong enrichment of synaptic functions observed across convergent genes and the neurodevelopmental deficits observed in ASD, we reasoned that convergent genes contributing to ASD etiology would display disease-relevant tissue specificity. We tested this hypothesis using GTEx[29] data and found no meaningful convergence in liver, left ventricle, or muscle coexpression, but we saw convergence in frontal cortex tissue (Figure 3A). The mean absolute convergence *Z* score for the liver, heart, and muscle were 0.318, 0.655, and 0.591 respectively, while the mean absolute *Z* score from the DLPFC was 1.05 and was significantly different than the non-brain tissues (two-sided Wilcoxon test, $p < 2.2 \times 10^{-16}$), suggesting tissue specificity of ASD convergence. The mean absolute convergence *Z* score for the frontal cortex was 1.10, which was similar to the DLPFC. Furthermore, we did not find any consistent relationships between CRISPR differential expression in neurons

Please cite this article in press as: Liao et al., Convergent coexpression of autism-associated genes suggests some novel risk genes may not be detectable in large-scale genetic studies, Cell Genomics (2023), https://doi.org/10.1016/j.xgen.2023.100277

CellPress
OPEN ACCESS

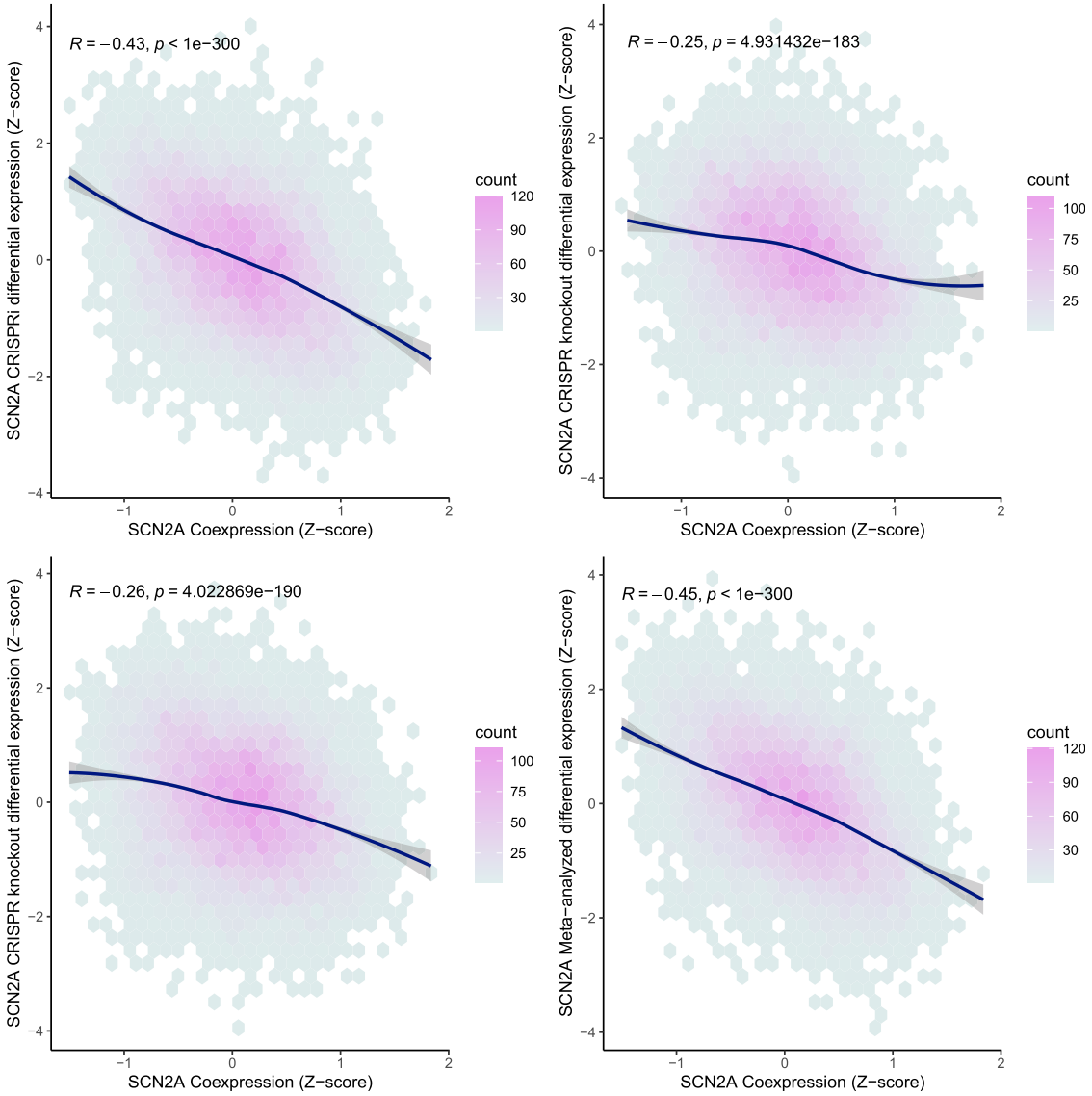


Figure 2. Coexpression consistently proxies differential expression for *SCN2A* across three independent perturbations

The top left panel shows CRISPRi differential expression in neurons of *SCN2A* correlated with coexpression.

The top right panel shows CRISPR heterozygous knockout differential expression in neurons of *SCN2A* correlated with coexpression.

The bottom left panel shows a second CRISPR heterozygous knockout differential expression in neurons of *SCN2A* correlated with coexpression. The bottom right panel shows the meta-analysis of all three *SCN2A* CRISPR experiments correlated with coexpression. Coexpression is represented as a Fisher transformed Pearson's correlation *Z* score (x axis). A Pearson correlation was calculated to assess the correlation between coexpression and differential expression (y axis) for each gene (points). The curve for each panel was fitted with a locally weighted smoothing (LOESS) regression.

and coexpression from the liver, heart tissue, and muscle and the correlation coefficients were significantly different from those seen in the brain (two-tailed Fisher's *Z* test, $p < 1 \times 10^{-300}$) (Figures S4–S6). To determine whether the enrichment of convergent genes associated with synaptic biology was driven by ASD risk genes associated with synaptic functions, we assigned 60 of the 71 ASD genes to relevant pathway terms associated with synaptic, transcription factor, and chromatin biology. The convergence permutations were rerun for each set separately, and we found that convergence was strongly driven by synaptic genes, whereas the chromatin genes and transcription factors displayed weaker convergence (synaptic variance = 2.57, chromatin variance = 0.39, transcription variance = 0.18) (Figure S7).

To assess the generalizability of this approach, we quantified convergence amongst 10 schizophrenia (SCZ) risk genes identified by the Schizophrenia Exome Meta-analysis consortium (SCHEMA).[30] We found significant convergence in the DLPFC (absolute mean = 0.74, variance = 0.81) but not in the unrelated tissues, such as the left ventricle, muscle, and liver. Next, we asked whether convergent genes overlapped between SCZ and ASD, given the known overlap in disease biology. Among

Please cite this article in press as: Liao et al., Convergent coexpression of autism-associated genes suggests some novel risk genes may not be detectable in large-scale genetic studies, Cell Genomics (2023), https://doi.org/10.1016/j.xgen.2023.100277

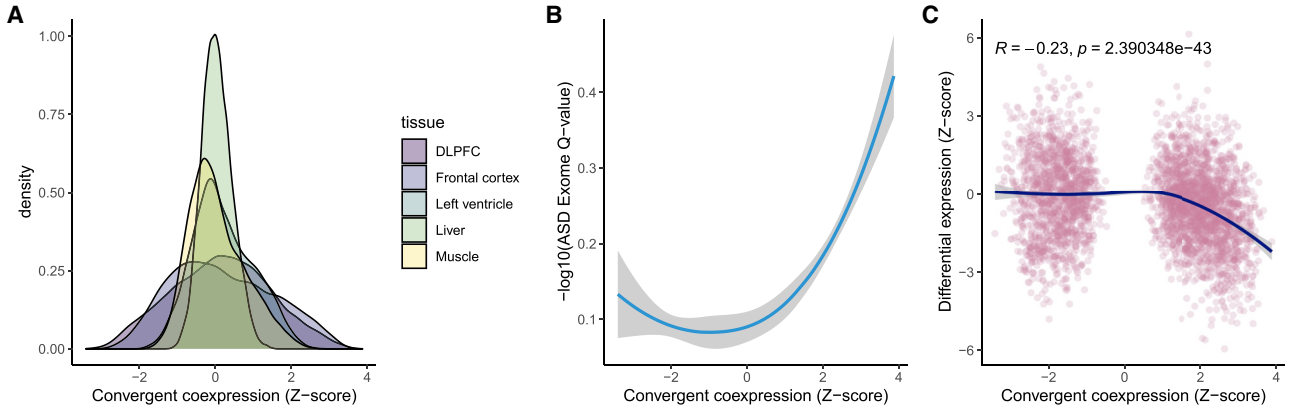


Figure 3. Convergence is tissue specific and associated with ASD

(A) Distributions of transcriptional convergence of 71 high-confidence ASD genes in different tissues.

(B) Smoothed relationship with confidence interval between ASD exome significance and significantly convergent coexpression effect sizes ($p_{bonf} < 0.01$).

(C) Correlation between convergently coexpressed genes ($p_{bonf} < 0.01$) and differential expression of ASD postmortem DLPFC compared with controls from Gandal et al.[8] The curve was fitted with a locally weighted smoothing (LOESS) regression.

the 135 significant SCZ convergent genes, 124 were also significant ASD convergent genes (Fisher's exact test, odds ratio [OR]: 60.80 [29.9–144.0], $p = 2.43 \times 10^{-74}$). Then, we selected an unrelated non-brain phenotype in atrial fibrillation and found no convergence in the brain tissues but 3,040 significant convergent genes ($p_{bonf} < 0.01$) in the left ventricle (Figure S8). Given the overlap between ASD and developmental disorders (DDs), we sought to compare the overlap of significantly convergent genes between ASD-predominant and DD-predominant genes. We found that there was a significant overlap between the 541 convergent ASD-specific genes and 91 convergent DD-specific genes (OR: 19.5 [12.20–30.60], $p = 7.11 \times 10^{-28}$). Interestingly, there was a strong overlap between ASD-predominant and SCZ (OR: 13.10 [8.64–19.5], $p = 4.99 \times 10^{-26}$), but the overlap between DD-predominant and SCZ was not significant (OR: 1.40, [0.035–8.16], p = 0.52).

Finally, we asked whether the ASD convergent genes could inform on the genetic etiology of ASD. Excluding the 71 ASD genes used to calculate convergence, there was a significant correlation between convergence and evidence for ASD risk based on the significance of each of the remaining genes to ASD (Spearman's rank correlation, rho = −0.316, $p = 3.03 \times 10^{-65}$) (Figure 3B). We find a similar correlation between convergence in frontal cortex tissue from GTEx 53 (Spearman's rank correlation, rho = −0.213, $p = 6.02 \times 10^{-25}$). There was also a significant enrichment of ASD genes implicated through rare variant analyses ($q < 0.05$) among the positively convergent genes ($Z > 2$) (OR: 4.63 [2.97–7.10], $p = 1.02 \times 10^{-10}$). This effect size was larger when only including Bonferroni-significant ($p < 0.01$) positively convergent genes (OR: 7.41 [3.10–20.40], $p = 1.85 \times 10^{-7}$). That is, genes most positively convergent were also most likely to be associated with ASD through rare variant analyses. This finding was also seen among SCZ convergence and the SCHEMA association results (Spearman's rank correlation, rho = −0.276, p = 0.005).

ASD risk genes are more likely to be intolerant to loss of function variation. Convergent genes also are significantly more likely to be intolerant (mean LOEUF for convergent genes = 0.73, mean LOEUF for other genes = 0.83, Wilcoxon $p = 2.85 \times 10^{-31}$). However, after splitting genes based on intolerance, we found a significant correlation between convergence and ASD association for tolerant (LOEUF > 0.35, Spearman's rank correlation, rho = 0.18, $p = 1.91 \times 10^{-15}$) and intolerant genes (LOEUF < 0.35, Spearman rank correlation, rho = 0.14, $p = 8.57 \times 10^{-5}$) (Figure S9). A significant relationship also existed between convergence and differential expression of ASD postmortem brain tissue compared with controls, with the positive convergence being correlated with downregulation in ASD (Spearman's rank correlation = −0.23, $p = 2.39 \times 10^{-43}$; Figure 3C). We find a similarly negative correlation when using frontal cortex from GTEx 53 (Spearman's rank correlation = −0.369, $p = 1.61 \times 10^{-76}$).

Potential to identify novel ASD genes not identified in current genetic analyses

Genes with shorter coding sequences will have reduced power in genetics analyses that require multiple deleterious variants among cases to quantify risk. Similar issues could exist for genes where deleterious variation is inviable but less deleterious modulation could contribute to ASD risk. Given the significant relationship between convergence and ASD association, we asked whether our set of convergent genes included potential ASD risk genes with properties that current genetics studies might be underpowered to identify. Among the most significantly associated 71 ASD risk genes, there was a highly significant skew toward a longer coding sequence (median = 3,642 bp) compared with 1,293 bp among all other genes. After splitting our positive convergent genes into those with even a weak association to ASD (Bayes factor [BF] > 2) or not, we identified significantly increased median coding sequence lengths in the ASD associated genes compared with the rest (median ASD-associated coding sequence length = 2,723 bp, median non-ASD -associated coding sequence length = 1,809 bp, $p = 5.76 \times 10^{-25}$, Wilcoxon signed-rank Test; Figure 4A). The most positively convergent genes not associated with ASD had intolerance scores

Please cite this article in press as: Liao et al., Convergent coexpression of autism-associated genes suggests some novel risk genes may not be detectable in large-scale genetic studies, Cell Genomics (2023), https://doi.org/10.1016/j.xgen.2023.100277

CellPress
OPEN ACCESS

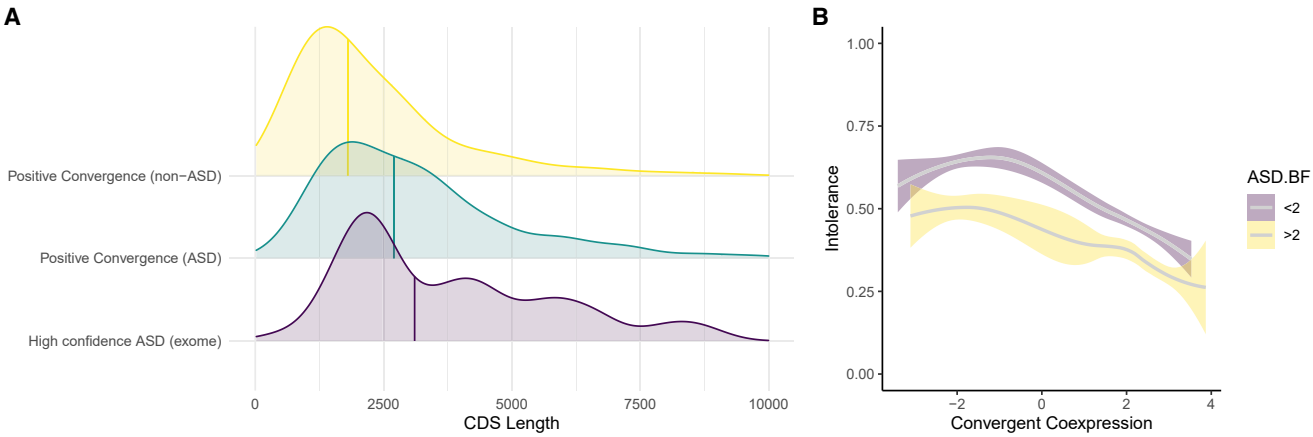


Figure 4. Convergence captures novel intolerant and shorter genes implicated in ASD risk

(A) Distributions and medians (colored vertical lines) of coding sequence lengths for significant positively convergent genes stratified by their association with ASD. High-confidence ASD exome genes (purple) consist of the 71 genes with a q-value of less than 0.001. A Bayes factor (BF) greater than 2 in the ASD exome data is considered ASD associated. The positive convergent genes are split by those with a BF greater than 2 (green) and those with a BF of less than 2 (yellow). (B) Relationship between convergent coexpression and intolerance (LOEUF scores) stratified by association with ASD among significantly convergent genes. Purple represents ASD non-associated genes (BF < 2), and yellow indicates ASD-associated genes (BF > 2). Significant convergence is defined as $p_{bonf} < 0.01$.

similar to those that were linked to ASD (Figure 4B). Consistent results were identified among convergent SCZ genes (Figure S10). We show that positive convergence is more significantly correlated with the ASD effect size driven by missense than protein-truncating variation (Spearman's rank correlation, missense type A, missense variants with missense badness, Polyphen-2, and constraint [MPC][31] scores between 1 and 2: rho = 0.117, $p = 5.36 \times 10^{-5}$; missense type B, missense variants with MPC scores greater than or equal to 2: rho = 0.103, $p = 3.96 \times 10^{-4}$; protein truncating: rho = 0.04, p = 0.039).

Finally, we sought to assess whether convergence may increase the power to detect genes that have some level of evidence linking them to ASD risk but have not been implicated in exome consortium analyses. To address this question, we leveraged the manually curated database of genes implicated in ASD susceptibility by the Simons Foundation Autism Research Initiative (SFARI; https://gene.sfari.org/). Genes in this database come from large-scale sequencing studies but also from functional studies, clinical reports, and genetic studies of common variation. After excluding all genes associated with ASD using the same criteria as before (BF > 2), there was a significant enrichment of the remaining positively convergent genes among this broader ASD gene set (Fisher's exact test, OR: 3.00 [2.07–4.37], $p = 1.00 \times 10^{-9}$) (Table S4). We additionally find similar synaptic pathways enriched among the non-ASD convergent genes (Table S5), which was replicated in the frontal cortex (Table S6). This finding suggests that convergence may provide a useful supplement to sequencing studies when searching for additional ASD risk genes.

DISCUSSION

Understanding the context-specific functional consequences of perturbing genes will be important in elucidating the molecular underpinnings of disease. Because *in vitro* experiments remain costly and challenging to scale, *in silico* approaches provide immediate opportunities to improve understanding. Here, we show that coexpression can proxy the regulatory consequences of CRISPR perturbation across shared contexts with similar correlation to replicate CRISPR experiments involving the same gene. Postmortem brain coexpression meta-analyzed across 71 genes implicated in ASD was then used to demonstrate highly significant regulatory convergence among synaptic genes. Finally, these convergent genes were found to be enriched for genes with support linking them to ASD but without definitive statistical evidence from large-scale rare genetics studies, which are underpowered to evaluate genes with short coding sequence length and high intolerance.

Our hypothesis was that gene-gene coexpression would capture some proportion of the downstream transcriptional consequences of perturbing a gene through CRISPR in a shared context, providing an opportunity to "proxy" these effects *in silico* and assess functional effects shared across many disease genes ("convergence"). Regardless of whether a CRISPR target gene was upregulated or downregulated, genes that were positively coexpressed with the target gene experienced transcriptional dysregulation in the same direction as the target. For CRISPRi gene knockdown experiments, genes positively coexpressed with the perturbed gene showed decreased expression, and for CRISPRa, such genes showed increased expression. Importantly, we show that the correlation of coexpression and CRISPR perturbation is similar to the correlation of different CRISPR experiments modulating the same gene. Furthermore, the meta-analysis incorporating multiple CRISPR experiments is even more highly correlated with coexpression, highlighting the variability in perturbation across CRISPR experiments and potential benefits of repeated experiments with different variants and backgrounds. Additionally, we found that the convergent differential expression from multiple CRISPR perturbations could be inferred by convergent coexpression of the same perturbed

genes. These results suggest that coexpression could be used to assess transcriptional convergence of disease-relevant genes.

We explored this possibility by assessing the transcriptional convergence of 71 ASD risk genes and found that there was highly significant convergence. The degree of convergence was context specific and absent in coexpression data from liver, left ventricle, or muscle but with a strong signature in both brain tissues. We found that context-specific convergence existed in other brain and non-brain diseases, such as SCZ and atrial fibrillation, demonstrating the generalizability of the approach. Significant convergence was also correlated with evidence of association with ASD from sequencing studies and with differential expression between ASD cases and controls in postmortem brain tissue. For SCZ, we found previously implicated genes such as *FURIN* to be strongly convergent, but there were fewer significant genes, which may reflect lower power[36]. Moreover, there are multiple pathways relevant to NDDs significantly enriched for the convergent genes, including pathways involving the synapse and ion channels,[32–34] reinforcing the link to ASD. Intriguingly, the relationship between convergence and ASD is predominantly a product of positive convergence, or genes positively coexpressed with many of the ASD risk genes. In general, there are more strongly positively convergent genes than negative. However, despite there being many significant negative convergent genes, these genes as a class lack existing evidence of ASD association, and compared with the positive convergent genes, they do not show enrichment within previously implicated ASD pathways. These results could point to a bias of coexpression, reflect ascertainment limitations arising from genes harboring of *de novo* LoF variants, or signal a biological phenomenon where only downstream functional effects in the same direction as dysregulation of risk genes contribute to risk.

Large-scale genetics studies have contributed dramatically to our current knowledge regarding the biological basis of ASD and spearheaded the identification of the risk genes used here to quantify convergence. These studies depend on observing a statistically significant excess of deleterious variation among cases. Factors that reduce the likelihood of finding variants such as a short coding sequence or mutational inviability diminish the power of these studies, potentially precluding genuine ASD risk genes from discovery via this approach. We were interested in whether our convergence metric could identify such putative "hidden" ASD risk genes. We show that our convergent genes are substantially shorter but similarly intolerant compared with genes previously associated with ASD and SCZ. Finally, we show that convergent genes not associated with ASD from sequencing studies are still enriched for genes implicated in ASD from clinical diagnostic studies, functional studies, or analyses of common variation, nominating our convergent genes described here as potential novel ASD risk genes. For example, *LMTK2* was the most convergently coexpressed gene with no association to ASD. It is highly intolerant to LoF mutation (LOEUF = 0.24) and highly expressed in brain tissue.[27,29] Interestingly, disruption of this gene contributes to infertility phenotypes in male mice.[35] The gene has also been linked to Alzheimer's disease and has been suggested to play an important role in axonal transport.[36,37] Despite being strongly coexpressed with high-confidence ASD genes and having characteristics similar to other ASD genes, it has not been implicated in ASD. Given the infertility phenotypes, this gene may have been missed as an ASD contributor simply because of having too few LoF *de novo* variants and thus insufficient power for association.

Limitations of the study

The ability to proxy CRISPR perturbation with coexpression enables quick *in silico* analyses to better understand transcriptional consequences of disruptive mutations and functional convergence. However, there are several limitations to this strategy. First, most transcriptional data are derived from bulk tissue. This can obscure relevant coexpression patterns, given the cellular heterogeneity and different transcriptional backgrounds among different cell types.[38] With increasingly larger single-cell datasets, there is a path to overcome this issue in the near future. Second, our primary ASD convergence analysis assumes a single underlying convergent pathway, while multiple pathways likely contribute, especially given the heterogeneity in presentation across individuals diagnosed with ASD. The analytical approach described here can be extended to search for multiple convergent pathways, and as the genotype-phenotype association becomes more granular, we may observe differing degrees of convergence for genes contributing to distinct phenotypic components of ASD. Finally, the use of postmortem samples cannot fully capture convergence during early development. Transcriptional consequences that may affect neurobiology prenatally cannot be easily captured using postnatal tissue. Future investigations assessing how transcriptional convergence differs across a developmental time span will prove critical to assess the relevance of using stage-specific biospecimens to answer specific biological questions.

In conclusion, coexpression provides an imperfect but simple proxy for context-specific transcriptional consequences of CRISPR perturbation and enables assessing convergence across many risk genes to provide insight into the biology of disease. Most notably, this approach may facilitate the identification of novel risk genes not captured by even the best-powered sequencing studies to date.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- METHOD DETAILS
  - CRISPR perturbation functional models
  - RNA sequencing of CRISPR perturbations
  - Differential expression analysis for CRISPR perturbations
  - Postmortem brain cohorts
  - RNA sequencing of postmortem samples
  - Generating pairwise coexpression of genes in the DLPFC

- Assessing relationship between CRISPR perturbation and coexpression
- ASD convergent coexpression meta-analysis
- Dissecting relationship between convergence, intolerance and ASD association
- Classification of 71 ASD associated genes

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xgen.2023.100277.

ACKNOWLEDGMENTS

C.L. was supported by a Vanier Graduate Scholarship from the Canadian Institutes of Health Research (CIHR). M.M.-O. and J.M.F. were supported by Autism Science Foundation Postdoctoral Fellowships (11815 and 11852, respectively). X.N. was supported by F32MH115614 and K99MH121577. G.A.R. was supported by a CIHR Foundation Scheme grant (332971). This work was also supported by R01MH123155 (to D.M.R., K.J.B., and M.E.T.), R01HD096326 (to M.E.T.), R01NS093200 (to M.E.T.), R01MH111776 (to D.M.R.), and the Simons Foundation for Autism Research (573206 to M.E.T.). Data were generated as part of the CommonMind Consortium, supported by funding from Takeda Pharmaceuticals Company Limited; F. Hoffman-La Roche Ltd.; NIH grants R01MH085542, R01MH093725, P50MH066392, P50MH080405, R01MH097276, RO1-MH- 075916, P50M0 96891, P50MH084053S1, R37MH057881, AG02219, AG05138, MH06692, R01MH110921, R01MH109677, R01MH109897, U01MH103392, and U01MH 116442; project ZIC MH002903; and contract HHSN271201300031C through IRP NIMH. Brain tissue for the study was obtained from the following brain bank collections: The Mount Sinai/James J. Peters VA Medical Center NIH Brain and Tissue Repository, the University of Pennsylvania Alzheimer's Disease Core Center, the University of Pittsburgh Brain Tissue Donation Program, and the NIMH Human Brain Collection Core. CMC Leadership: Panos Roussos, Joseph Buxbaum, Andrew Chess, Schahram Akbarian, Vahram Haroutunian (Icahn School of Medicine at Mount Sinai), Bernie Devlin, David Lewis (University of Pittsburgh), Raquel Gur (University of Pennsylvania), Chang-Gyu Hahn (Thomas Jefferson University), Enrico Domenici (University of Trento), Mette A. Peters, Solveig Sieberts (Sage Bionetworks), Stefano Marenco, Barbara K. Lipska, and Francis J. McMahon (NIMH). All ASD gene lists were provided by the ASC-SSC Genomics Consortium.

AUTHOR CONTRIBUTIONS

C.L. contributed to analysis, design of the study, and writing and editing of the manuscript. M.M.–O., C.D.E., R.B., X.N., A.L., A.Y., N.D.B., M.W., and B.Z., contributed to generation of data and writing the manuscript. S.E., J.M.F., T.M., and L.H. contributed to data generation and processing. G.A.R. and P.A.D. contributed to design and writing of the manuscript. K.J.B., M.E.T., and D.M.R. contributed to design, analysis of the study, and writing and editing of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: August 5, 2022
Revised: December 1, 2022
Accepted: February 10, 2023
Published: March 9, 2023

REFERENCES

1. 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. Nat. Genet. *51*, 431–444. https://doi.org/10.1038/s41588-019-0344-8.
2. De Rubeis, S., He, X., Goldberg, A.P., Poultney, C.S., Samocha, K., Cicek, A.E., Kou, Y., Liu, L., Fromer, M., Walker, S., et al. (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. Nature *515*, 209–215. https://doi.org/10.1038/nature13772.
3. Satterstrom, F.K., Kosmicki, J.A., Wang, J., Breen, M.S., De Rubeis, S., An, J.-Y., Peng, M., Collins, R., Grove, J., Klei, L., et al. (2020). Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. Cell *180*, 568–584.e23. https://doi.org/10.1016/j.cell.2019.12.036.
4. 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. https://doi.org/10.1016/j.neuron.2015.09.016.
5. Fu, J.M., Satterstrom, F.K., Peng, M., Brand, H., Collins, R.L., Dong, S., Klei, L., Stevens, C.R., Cusick, C., Babadi, M., et al. (2021). Rare coding variation illuminates the allelic architecture, risk genes, cellular expression patterns, and phenotypic context of autism. Nat. Genet. *10*. https://doi.org/10.1038/s41588-022-01104-0.
6. 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. https://doi.org/10.1038/nature10110.
7. Gupta, I., Collier, P.G., Haase, B., Mahfouz, A., Joglekar, A., Floyd, T., Koopmans, F., Barres, B., Smit, A.B., Sloan, S.A., et al. (2018). Single-cell isoform RNA sequencing characterizes isoforms in thousands of cerebellar cells. Nat. Biotechnol. *36*, 1197–1202. https://doi.org/10.1038/nbt.4259.
8. Gandal, M.J., Zhang, P., Hadjimichael, E., Walker, R.L., Chen, C., Liu, S., Won, H., Van Bakel, H., Varghese, M., Wang, Y., et al. (2018). Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. Science *362*, eaat8127. https://doi.org/10.1126/science.aat8127.
9. Sugathan, A., Biagioli, M., Golzio, C., Erdin, S., Blumenthal, I., Manavalan, P., Ragavendran, A., Brand, H., Lucente, D., Miles, J., et al. (2014). CHD8 regulates neurodevelopmental pathways associated with autism spectrum disorder in neural progenitors. Proc. Natl. Acad. Sci. USA *111*, E4468–E4477. https://doi.org/10.1073/pnas.1405266111.
10. Deneault, E., White, S.H., Rodrigues, D.C., Ross, P.J., Faheem, M., Zaslavsky, K., Wang, Z., Alexandrova, R., Pellecchia, G., Wei, W., et al. (2018). Complete disruption of autism-susceptibility genes by gene editing predominantly reduces functional connectivity of isogenic human neurons. Stem Cell Rep. *11*, 1211–1225. https://doi.org/10.1016/j.stemcr.2018.10.003.
11. Araujo, D.J., Anderson, A.G., Berto, S., Runnels, W., Harper, M., Ammanuel, S., Rieger, M.A., Huang, H.-C., Rajkovich, K., Loerwald, K.W., et al. (2015). FoxP1 orchestration of ASD-relevant signaling pathways in the striatum. Genes Dev. *29*, 2081–2096. https://doi.org/10.1101/gad.267989.115.
12. Cotney, J., Muhle, R.A., Sanders, S.J., Liu, L., Willsey, A.J., Niu, W., Liu, W., Klei, L., Lei, J., Yin, J., et al. (2015). The autism-associated chromatin modifier CHD8 regulates other autism risk genes during human neurodevelopment. Nat. Commun. *6*, 6404. https://doi.org/10.1038/ncomms7404.
13. Willsey, H.R., Exner, C.R.T., Xu, Y., Everitt, A., Sun, N., Wang, B., Dea, J., Schmunk, G., Zaltsman, Y., Teerikorpi, N., et al. (2021). Parallel in vivo analysis of large-effect autism genes implicates cortical neurogenesis and estrogen in risk and resilience. Neuron *109*, 788–804.e8. https://doi.org/10.1016/j.neuron.2021.01.002.
14. Jin, X., Simmons, S.K., Guo, A., Shetty, A.S., Ko, M., Nguyen, L., Jokhi, V., Robinson, E., Oyler, P., Curry, N., et al. (2020). In vivo Perturb-Seq reveals neuronal and glial abnormalities associated with autism risk genes. Science *370*, eaaz6063, eaaz6063. https://doi.org/10.1126/science.aaz6063.

15. Cederquist, G.Y., Tchieu, J., Callahan, S.J., Ramnarine, K., Ryan, S., Zhang, C., Rittenhouse, C., Zeltner, N., Chung, S.Y., Zhou, T., et al. (2020). A multiplex human pluripotent stem cell platform defines molecular and functional subclasses of autism-related genes. Cell Stem Cell *27*, 35–49.e6. https://doi.org/10.1016/j.stem.2020.06.004.

16. Iakoucheva, L.M., Muotri, A.R., and Sebat, J. (2019). Getting to the cores of autism. Cell *178*, 1287–1298. https://doi.org/10.1016/j.cell.2019.07.037.

17. Liu, L., Lei, J., Sanders, S.J., Willsey, A.J., Kou, Y., Cicek, A.E., Klei, L., Lu, C., He, X., Li, M., et al. (2014). DAWN: a framework to identify autism genes and subnetworks using gene expression and genetics. Mol. Autism. *5*, 22. https://doi.org/10.1186/2040-2392-5-22.

18. 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. https://doi.org/10.1016/j.cell.2013.10.031.

19. Krishnan, A., Zhang, R., Yao, V., Theesfeld, C.L., Wong, A.K., Tadych, A., Volfovsky, N., Packer, A., Lash, A., and Troyanskaya, O.G. (2016). Genome-wide prediction and functional characterization of the genetic basis of autism spectrum disorder. Nat. Neurosci. *19*, 1454–1462. https://doi.org/10.1038/nn.4353.

20. 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*, eaat7615. https://doi.org/10.1126/science.aat7615.

21. Corominas, R., Yang, X., Lin, G.N., Kang, S., Shen, Y., Ghamsari, L., Broly, M., Rodriguez, M., Tam, S., Trigg, S.A., et al. (2014). Protein interaction network of alternatively spliced isoforms from brain links genetic risk factors for autism. Nat. Commun. *5*, 3650. https://doi.org/10.1038/ncomms4650.

22. Pinto, D., Delaby, E., Merico, D., Barbosa, M., Merikangas, A., Klei, L., Thiruvahindrapuram, B., Xu, X., Ziman, R., Wang, Z., et al. (2014). Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. Am. J. Hum. Genet. *94*, 677–694. https://doi.org/10.1016/j.ajhg.2014.03.018.

23. Walker, R.L., Ramaswami, G., Hartl, C., Mancuso, N., Gandal, M.J., de la Torre-Ubieta, L., Pasaniuc, B., Stein, J.L., and Geschwind, D.H. (2019). Genetic control of expression and splicing in developing human brain informs disease mechanisms. Cell *179*, 750–771.e22. https://doi.org/10.1016/j.cell.2019.09.021.

24. 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. https://doi.org/10.1016/j.cell.2013.10.020.

25. Seabra, C.M., Aneichyk, T., Erdin, S., Tai, D.J.C., De Esch, C.E.F., Razaz, P., An, Y., Manavalan, P., Ragavendran, A., Stortchevoi, A., et al. (2020). Transcriptional consequences of MBD5 disruption in mouse brain and CRISPR-derived neurons. Mol. Autism. *11*, 45. https://doi.org/10.1186/s13229-020-00354-1.

26. Schrode, N., Ho, S.-M., Yamamuro, K., Dobbyn, A., Huckins, L., Matos, M.R., Cheng, E., Deans, P.J.M., Flaherty, E., Barretto, N., et al. (2019). Synergistic effects of common schizophrenia risk variants. Nat. Genet. *51*, 1475–1485. https://doi.org/10.1038/s41588-019-0497-5.

27. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nature *581*, 434–443. https://doi.org/10.1038/s41586-020-2308-7.

28. Piñero, J., Bravo, À., Queralt-Rosinach, N., Gutiérrez-Sacristán, A., Deu-Pons, J., Centeno, E., García-García, J., Sanz, F., and Furlong, L.I. (2017). DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. Nucleic Acids Res. *45*, D833–D839. https://doi.org/10.1093/nar/gkw943.

29. GTEx Consortium, Laboratory, Data Analysis &Coordinating Center LDACC—Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx eGTEx groups, NIH Common Fund, NIH/NCI, NIH/NHGRI, NIH/NIMH, NIH/NIDA, Biospecimen Collection Source Site—NDRI (2017). Genetic effects on gene expression across human tissues. Nature *550*, 204–213. https://doi.org/10.1038/nature24277.

30. Singh, T., Poterba, T., Curtis, D., Akil, H., Al Eissa, M., Barchas, J.D., Bass, N., Bigdeli, T.B., Breen, G., Bromet, E.J., et al. (2022). Rare coding variants in ten genes confer substantial risk for schizophrenia. Nature *604*, 509–516. https://doi.org/10.1038/s41586-022-04556-w.

31. 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. Preprint at bioRxiv, 148353. https://doi.org/10.1101/148353.

32. Ebrahimi-Fakhari, D., and Sahin, M. (2015). Autism and the synapse: emerging mechanisms and mechanism-based therapies. Curr. Opin. Neurol. *28*, 91–102. https://doi.org/10.1097/WCO.0000000000000186.

33. Daghsni, M., Rima, M., Fajloun, Z., Ronjat, M., Brusés, J.L., M'rad, R., and De Waard, M. (2018). Autism throughout genetics: perusal of the implication of ion channels. Brain Behav. *8*, e00978. https://doi.org/10.1002/brb3.978.

34. Bourgeron, T. (2015). From the genetic architecture to synaptic plasticity in autism spectrum disorder. Nat. Rev. Neurosci. *16*, 551–563. https://doi.org/10.1038/nrn3992.

35. Kawa, S., Ito, C., Toyama, Y., Maekawa, M., Tezuka, T., Nakamura, T., Nakazawa, T., Yokoyama, K., Yoshida, N., Toshimori, K., and Yamamoto, T. (2006). Azoospermia in mice with targeted disruption of the Brek/Lmtk2 (brain-enriched kinase/lemur tyrosine kinase 2) gene. Proc. Natl. Acad. Sci. USA *103*, 19344–19349. https://doi.org/10.1073/pnas.0603603103.

36. Bencze, J., Mórotz, G.M., Seo, W., Bencs, V., Kálmán, J., Miller, C.C.J., and Hortobágyi, T. (2018). Biological function of Lemur tyrosine kinase 2 (LMTK2): implications in neurodegeneration. Mol. Brain *11*, 20. https://doi.org/10.1186/s13041-018-0363-x.

37. Bencze, J., Szarka, M., Bencs, V., Szabó, R.N., Smajda, M., Aarsland, D., and Hortobágyi, T. (2019). Neuropathological characterization of Lemur tyrosine kinase 2 (LMTK2) in Alzheimer's disease and neocortical Lewy body disease. Sci. Rep. *9*, 17222. https://doi.org/10.1038/s41598-019-53638-9.

38. Farahbod, M., and Pavlidis, P. (2020). Untangling the effects of cellular composition on coexpression analysis. Genome Res. *30*, 849–859. https://doi.org/10.1101/gr.256735.119.

39. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

40. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

41. DeLuca, D.S., Levin, J.Z., Sivachenko, A., Fennell, T., Nazaire, M.-D., Williams, C., Reich, M., Winckler, W., and Getz, G. (2012). RNA-SeQC: RNA-seq metrics for quality control and process optimization. Bioinformatics *28*, 1530–1532. https://doi.org/10.1093/bioinformatics/bts196.

42. Wang, L., Wang, S., and Li, W. (2012). RSeQC: quality control of RNA-seq experiments. Bioinformatics *28*, 2184–2185. https://doi.org/10.1093/bioinformatics/bts356.

43. Anders, S., Reyes, A., and Huber, W. (2012). Detecting differential usage of exons from RNA-seq data. Genome Res. *22*, 2008–2017. https://doi.org/10.1101/gr.133744.111.

44. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26. https://doi.org/10.1038/nbt.1754.

45. Ho, S.-M., Topol, A., and Brennand, K.J. (2015). From "directed differentiation" to "neuronal induction": modeling neuropsychiatric disease. Biomark. Insights *10*, 31–41. https://doi.org/10.4137/BMI.S20066.

46. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

47. Golzio, C., Willer, J., Talkowski, M.E., Oh, E.C., Taniguchi, Y., Jacquemont, S., Reymond, A., Sun, M., Sawa, A., Gusella, J.F., et al. (2012). KCTD13 is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant. Nature *485*, 363–367. https://doi.org/10.1038/nature11091.

48. Tai, D.J.C., Razaz, P., Erdin, S., Gao, D., Wang, J., Nuttle, X., Esch, C.E. de, Collins, R.L., Currall, B.B., O'Keefe, K., et al. (2022). Tissue and cell-type specific molecular and functional signatures of 16p11.2 reciprocal genomic disorder across mouse brain and human neuronal models. AJHG. https://doi.org/10.1101/2022.05.12.491670.

49. Fromer, M., Roussos, P., Sieberts, S.K., Johnson, J.S., Kavanagh, D.H., Perumal, T.M., Ruderfer, D.M., Oh, E.C., Topol, A., Shah, H.R., et al. (2016). Gene expression elucidates functional impact of polygenic risk for schizophrenia. Nat. Neurosci. *19*, 1442–1453. https://doi.org/10.1038/nn.4399.

50. Han, L., Zhao, X., Benton, M.L., Perumal, T., Collins, R.L., Hoffman, G.E., Johnson, J.S., Sloofman, L., Wang, H.Z., Stone, M.R., et al. (2020). Functional annotation of rare structural variation in the human brain. Nat. Commun. *11*, 2990. https://doi.org/10.1038/s41467-020-16736-1.

51. Fu, J.M., Satterstrom, F.K., Peng, M., Brand, H., Collins, R.L., Dong, S., Wamsley, B., Klei, L., Wang, L., Hao, S.P., et al. (2022). Rare coding variation provides insight into the genetic architecture and phenotypic context of autism. Nat. Genet. *54*, 1320–1331. https://doi.org/10.1038/s41588-022-01104-0.

52. Koopmans, F., van Nierop, P., Andres-Alonso, M., Byrnes, A., Cijsouw, T., Coba, M.P., Cornelisse, L.N., Farrell, R.J., Goldschmidt, H.L., Howrigan, D.P., et al. (2019). SynGO: an evidence-based, expert-curated knowledgebase for the synapse. Neuron *103*, 217–234.e4. https://doi.org/10.1016/j.neuron.2019.05.002.

Please cite this article in press as: Liao et al., Convergent coexpression of autism-associated genes suggests some novel risk genes may not be detectable in large-scale genetic studies, Cell Genomics (2023), https://doi.org/10.1016/j.xgen.2023.100277

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
Doxycycline hyclate	Millipore Sigma	24390-14-5
Puromycin Dihydrochloride	Thermo Fisher Scientific	A1113802
Recombinant BDNF Protein	Peprotech	450–02
Recombinant GDNF Protein	Peprotech	450–10
Y-27632 dihydrochloride (Rock Inhibitor)	MedChemExpress	HY-10583
Critical commercial assays		
Human Stem Cell Nucleofector Kit 1	Lonza	VPH-5012
TruSeq stranded mRNA Sample Prep Kit	Illumina	RS-122-2102
Deposited data		
RNA-seq from CRISPR-mediated LoF gene editing neuronal models (*ANOS1, ASTN2, ATRX, CACNA1C, CHD8, DLGAP2, AFF2/FMR2, KCNQ2, SCN2A, TENM1*)	Deneault et al.[10]	GSE107878
RNA-seq from CRISPR-mediated LoF gene editing neuronal models (*MBD5*)	Seabra et al.[25]	GSE144279
RNA-seq from CRISPR neuronal models (*TSNARE1, SNAP91, CLCN3, and FURIN*)	Schrode et al.[26]	syn20502314
RNA-seq from CRISPR interference (CRISPRi) in neuronal models (*SCN2A*)	This paper	syn26970716
RNA-seq from CRISPR neuronal models (*SCN2A*)	This paper	GSE222259
Experimental models: Cell lines		
Human pluripotent stem cell line:	Massachusetts General Hospital	GM08330
iPSC derived glutamatergic neurons	In house differentiation	N/A
Oligonucleotides		
SCN2A crRNA sequence: 5-tatcgtagggggaccaacc-3′	This paper	N/A
SCN2A crRNA sequence: 5′-gcgtggatctagtgaactt-3′	This paper	N/A
KCTD13 crRNA sequence: 5′- taaaaaggatggatgtaggc-3′	This paper	N/A
KCTD13 crRNA sequence: 5′-tgcctgtgttaggaggtatc-3′	This paper	N/A
Recombinant DNA		
pSpCas9(BB)-2A-Puro (PX459) Amp	Addgene	48139
Lentivirus TetO-Ngn2-Puro	Addgene	52047
Lentivirus FUW-M2rtTA	Addgene	20342
Software and algorithms		
ImageJ	Schneider et al.[7]	https://imagej.nih.gov/ij/
Scripts used for QC, analyses and relevant bioinformatic analyses in the study	This study	https://doi.org/10.5281/zenodo.7603485
Bowtie2	Langmead and Salzberg[8]	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
Samtools	Li et al.[9]	http://samtools.sourceforge.net/
FastQC		http://www.bioinformatics.babraham.ac.uk/projects/fastqc
Trimmomatic	Bolger[39]	https://github.com/timflutre/trimmomatic
STAR	Dobin[40]	https://github.com/alexdobin/STAR
Picard Tools		https://broadinstitute.github.io/picard/
RNASeQC	DeLuca[41]	https://github.com/getzlab/rnaseqc

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
RSeQC	Wang[42]	https://rseqc.sourceforge.net/
DEXseq	Anders[43]	https://bioconductor.org/packages/release/bioc/html/DEXSeq.html
Integrative Genomics Viewer (IGV)	Robinson[44]	https://software.broadinstitute.org/software/igv/
Other		
mTeSR medium	STEMCELL Technologies	85850
Essential 8 Medium	Thermo Fisher Scientific	A1517001

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Douglas M. Ruderfer (douglas.ruderfer@vanderbilt.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

All relevant code and scripts are available on Zenodo listed in the key resources table. Any new generated data has been deposited in the Gene Expression Omnibus (GEO) with the accession number listed in the key resources table.

METHOD DETAILS

CRISPR perturbation functional models

We collected data from 17 newly generated and previously published CRISPR experiments targeting 15 different ASD-associated genes. The transcriptomic data from all these CRISPR experiments were generated from isogenic iPSC-derived glutamatergic neurons induced by Neurogenin 2 (Ngn2) overexpression. Data of previously published cellular models generated by CRISPR perturbation were ascertained from the NCBI's Gene Expression Omnibus (GEO) and Synapse. For the CRISPR-mediated LoF gene editing models described in Deneault et al., gene counts and covariate information were ascertained from GEO (Accession # GSE107878). These consisted of 10 ASD-relevant genes (*ANOS1*, *ASTN2*, *ATRX*, *CACNA1C*, *CHD8*, *DLGAP2*, *AFF2/FMR2*, *KCNQ2*, *SCN2A*, *TENM1*). Gene counts and covariate information for *MBD5* CRISPR-mediated LoF gene editing model were ascertained from GEO (Accession #: GSE144279). For the CRISPR activation (CRISPRa) targets (*TSNARE1*, *SNAP91*, *CLCN3*, and *FURIN*), gene counts and covariate information were ascertained from Synapse (syn20502314).[18] CRISPR interference (CRISPRi) for *SCN2A* was ascertained from Synapse (syn26970716). From those datasets, only neuronal lines were included, and experiments were removed if there was any evidence of ineffective perturbation such as discordant RNA and protein levels of the perturbed gene.

Additionally, we used unpublished transcriptomic data from *SCN2A* and *KCTD13* CRISPR-mediated LoF gene editing models which leveraged a dual-gRNA strategy to promote gene deletions. *SCN2A* sgRNAs targeted intron 4 and intron 11 (NM_001040142) to generate a 7.5kb partial gene deletion, and the crRNA sequences used were: 5′-TATCGTAGGGGGACCAACC-3' and 5′-GCGT GGATCTAGTGAACTT-3'. *KCTD13* sgRNAs targeted 5′ AND-3′ regions from *KCTD13* (NM_178863) to generate a 23.3kb full gene deletion, and the crRNA sequences used were: 5′-taaaaaggatggatgtaggc-3′ and 5′-tgcctgtgttaggaggtatc-3′.

The deletion lines were generated in the male control human iPSC line GM08330-8 using the Human Stem Cell Nucleofector Kit 1 (Lonza), transfecting 1μg CRISPR/Cas9 PX459 plasmid and 1μg of each gRNA using the Amaxa Nucleofection II device (Lonza), according to the manufacturer's instructions. Cells were subsequently plated on Matrigel plates in mTeSR1 or Essential 8 medium supplemented with ROCK inhibitor for 24 h. For clonal isolation of *SCN2A* models, puromycin selection was started 24 h after transfection and resistant colonies were picked and expanded 48 h after selection. For clonal isolation of *KCTD13* models, cells were separated by fluorescence-activated cell sorting (FACS) 48h after transfection. Genotyping of the resultant colonies for *SCN2A* and *KCTD13* was performed by Sanger sequencing of the deletion-specific region and ddPCR assays for copy number. A total of 4–6 successfully edited clones with heterozygous deletions plus 2–6 unedited (i.e. WT-Cas9 exposed) clonal colonies were expanded per target. Prior to neuronal differentiation, iPSC clones were split into multiple replicates, and each was manipulated in parallel during subsequent experiments.

For differentiation of *SCN2A* and *KCTD13* human iPSC models into glutamatergic neurons, Ngn2-neuronal induction was performed as previously described.[45] Briefly, iPSCs were seeded at a density of 10^6 cells/mL and transduced with a lentivirus expressing TetO-Ngn2-GFP-Puro or TetO-Ngn2-Puro along with rtTA. Twenty-four hours after transduction, doxycycline was added to initiate

Ngn2 expression, and then 24h later the cells were selected with puromycin. Ngn2-glutamatergic iPSC-derived neurons were cultured in neuronal maintenance medium supplemented with BDNF and GDNF growth factors for an additional 22 days. Subsequent experiments were performed with 24-day-old Ngn2-glutamatergic neurons, using 6–34 total replicates per genotype (i.e. WT, heterozygous deletion) per target gene.

RNA sequencing of CRISPR perturbations

SCN2A RNAseq libraries were prepared from 200 ng of total RNA using a TruSeq stranded mRNA Sample Prep Kit (Illumina cat# RS-122-2102). Libraries were multiplexed, pooled and sequenced on multiple lanes of the Illumina NovaSeq platform, generating an average of 30.7M paired-end 150 bp-cycle reads for 30 samples (20 *SCN2A*$^{+/-}$, 10 *SCN2A*$^{+/+}$). RNAseq data was processed using a standard workflow, which includes quality assessment of fastq reads using FastQc (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Raw sequence reads were trimmed against Illumina adapters using Trimmomatic[39] (v. 0.36) with parameters ILLUMINACLIP:adapter.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:75. Sequence reads were aligned to the human reference genome (GRCh37, Ensembl build 75) using STAR[40] (v. 2.5.2.a) with parameters '–outSAMunmapped Within –outFilterMultimapNmax 1 –outFilterMismatchNoverLmax 0.1 –alignIntronMin 21 – alignIntronMax 0 –alignEndsType Local –quantMode GeneCounts –twopassMode Basic'. STAR aligner also generated gene level counts for all libraries relying on the human genome annotation provided for Ensembl GRCh37, build 75. Quality checking of alignments was assessed by custom scripts utilizing PicardTools (https://broadinstitute.github.io/picard/), RNASeQC,[41] RSeQC[42] and SamTools.[46] The deletion introduced by CRISPR on *SCN2A* loci was validated generating exon-level coverage using DEXseq[43] and visually investigating target loci using IGV.[44] *KCTD13* was included in gene edited as a potential contributor to 16p11.2 genomic disorder,[47] and these lines are described elsewhere.[48] In brief, RNAseq libraries were prepared using a TruSeq stranded mRNA library kit (Illumina) and were multiplexed, pooled and sequenced on multiple lanes of Illumina HiSeq 2500 platform, generating an average of 46.5M paired-end reads of 75bp for 11 samples (5 KCTD13$^{+/-}$, 6 KCTD13$^{+/+}$). The same RNAseq data processing pipeline without trimming step as above was applied to *KCTD13* RNAseq libraries. Processing of *MBD5* RNAseq libraries were previously described elsewhere.[25]

For the CRISPR-mediated LoF gene editing models described in Deneault et al., gene counts and covariate information were ascertained from NCBI's Gene Expression Omnibus (GEO accession: GSE107878). The processing of samples followed standard RNA sequencing pipelines as previous described in Deneault et al. (2018).[10]

Differential expression analysis for CRISPR perturbations

Gene counts were input into DESeq2 for differential expression analysis. Genes with normalized counts <10 in at least half of samples were excluded. Covariates including RNA integrity number (RIN), batch, and processing date were included when available. A Wald's test was used for differential expression and Z-scores were calculated by dividing the fold change by the SE. Each CRISPR perturbation was analyzed separately.

Postmortem brain cohorts

Two separate cohorts of postmortem brain samples of the dorsolateral prefrontal cortex (DLPFC) were used. The CommonMind Consortium (CMC) included tissue samples from Mount Sinai NIH Brain Bank and Tissue Repository, The University of Pittsburgh NIH NeuroBioBank Brain and Tissue Repository, and the University of Pennsylvania Brain Bank of Psychiatric Illnesses and Alzheimer's Disease Core Center. The DLPFC was dissected at each bank and sent to a centralized center, the Icahn School of Medicine at Mount Sinai (ISMMS) for RNA extraction. Tissues from bipolar disorder (BD) or schizophrenia cases were included if they met the DSM-IV diagnostic criteria for schizoaffective disorder or schizophrenia, or for BD, which were determined in consensus conferences after reviewing of medical records, direct clinical assessments, and care provider interviews. Samples were excluded if donors had a history of Alzheimer's disease, Parkinson disease, were on ventilators near time of death, or had acute neurological insults (anoxia, strokes and/or traumatic brain injury) before death.

The Human Brain Core Collection (HBCC) cohort consisted of DLPFC samples from the NIMH HBCC. The samples were analyzed clinically, neuropathologically, and toxicologically. The DSM-IV clinical diagnosis was determined through review of medical records by two psychiatrists and family interviews. Non-psychiatric controls did not have a history of substance use disorder or psychiatric conditions. Across all 933 samples, there were 345 females and 588 males. The self-reported ethnicities were 637 Europeans, 249 African, 33 Hispanic, 13 Asian and 1 other. Nearly fifty percent of samples were non-psychiatric controls (N = 462) and the remainder had a psychiatric diagnosis (113 BD, 350 schizophrenia, and 8 affective disorder).[49] The research abided by ethical regulations and was approved by the Vanderbilt University Medical Center Review Board (IRB: 220287).

RNA sequencing of postmortem samples

Approximately 50 mg of homogenized tissue from the DLPFC was used to isolate RNA. The two cohorts were processed separately. Samples with age <18 were excluded prior to analysis. The RNAseq processing is identical to that described in Han et al. (2020)[50] except for not using surrogate variable analysis (SVA) here to avoid removing *trans*-regulatory effects. Briefly, STAR was used to align RNA sequencing reads to GRCh37. FeatureCounts (v1.5.2) was used to count uniquely mapped reads that overlapped genes using the Ensembl v75 annotations. Fixed/mixed effects modeling were used for library normalization and covariate adjustments. Genes that were expressed at levels >1 counts per million (CPM) in at least half of the samples in each study were retained for analysis.

Please cite this article in press as: Liao et al., Convergent coexpression of autism-associated genes suggests some novel risk genes may not be detectable in large-scale genetic studies, Cell Genomics (2023), https://doi.org/10.1016/j.xgen.2023.100277

Conditional quantile normalization was done to account for variation in GC-content and gene length. A weighted-linear model using voom-limma was used to assess the sampling abundance confidence. Normalized $\log_2$(CPM) values were used for hierarchical clustering and principal component analysis to detect outlier samples. Samples were removed if deemed outliers using either method. For the CMC cohort, covariates were identified using a stepwise fixed/mixed effect regression model to identify covariates significantly associated with gene expression. The covariates were added if there was an association with principal components explaining greater than 1% of expression residual variance. For the HBCC cohort, model selection was determined through Bayesian information criteria (BIC). BIC were used to find fixed effect covariates that improved the model for most genes. Covariate adjustment was done using with a fixed/mixed effect linear regression variant, choosing mixed effect models when several samples were available per donor: gene expression $\sim$ covariates + sex + diagnosis + (1|Donor). Observation weights were calculated using voom-limma to adjust for the mean-variance relationship. The covariate-adjusted expression was generated after adding back the diagnostic component. This was done for both HBCC and CMC cohorts.

Generating pairwise coexpression of genes in the DLPFC

For each of the two cohorts, the covariate-adjusted expression was used to calculate coexpression values across all 16,992 genes. A pairwise Pearson's correlation was calculated for each pair of genes and the correlation coefficient was subsequently transformed into a *Z* score using a Fisher transformation. The coexpression Z-scores for each pair of genes were subsequently meta-analyzed across the two cohorts using Stouffer's weighted *Z* score method.

Assessing relationship between CRISPR perturbation and coexpression

For each experiment, we calculated Pearson correlation between the rank normalized differential expression and the perturbed gene's coexpression profiles from the postmortem brain tissue. ASD CRISPR perturbations were meta-analyzed using Stouffer's weighted *Z* score and compared to the meta-analyzed coexpression of the same genes. The relationship was subsequently assessed with a Pearson's correlation.

ASD convergent coexpression meta-analysis

We included 71 genes implicated in risk of ASD from a cross-consortia exome sequencing study that combined *de novo* and inherited single-nucleotide variant (SNV), indel, and CNV analyses (FDR <0.001, approximating exome-wide Bonferroni correction).[5] After filtering out genes due to low expression level, we performed a meta-analysis of the coexpression profiles of the 71 genes. Coexpression Z-scores were meta-analyzed using Stouffer's weighted *Z* score method to generate meta-analyzed coexpression effect sizes, which represent the convergent coexpression effect. To assess whether convergence of ASD risk genes is tissue specific, we used Genotype Tissue Expression Consortium (GTEx) RNAseq counts of the frontal cortex (BA9) (N = 209), left ventricle (N = 432), skeletal muscle (N = 803) and liver (N = 226) as negative controls. We selected these as they are not derived from the ectoderm or previously implicated in ASD. Convergence was calculated using the same methods as described previously. To assess the null distribution for ASD convergence, we conducted 10,000,000 permutations where for each permutation a meta-analysis was performed using 71 randomly-selected genes, excluding the 71 ASD genes. At the time these analyses were done, there were only 71 high-confidence ASD genes based on the preprint but there are now 72 implicated ASD genes. Additionally, we conducted convergence leveraging all nominally significant 185 ASD genes (FDR <0.05) and ASD-specific or development disorder-specific genes.[51] The empirical p values were calculated, as shown here: (# of absolute convergent Z-scores greater than or equal to the absolute convergent ASD *Z* score +1)/(Total # of permutations +1). A Bonferroni-correction was then applied to the empirical p values and a threshold of $p < 0.01$ after correction was used to increase stringency.

To account for the potential effects of confounding from LoF observed/expected upper bound fraction (LOEUF), the permutation was repeated by matching on LOEUF scores within $-/+$ 0.05 of each gene. Genes that were Bonferroni-significant and large effect (absolute *Z* score >2) were input into ToppGene for pathway enrichment, with the background gene set being all unique genes with coexpression values. Afterward, we sought to determine whether the convergent coexpression relates to differentially expressed ASD genes. Summary statistics from the ASD vs control postmortem brain PsychENCODE dataset were ascertained and correlated against the convergent coexpression.

Dissecting relationship between convergence, intolerance and ASD association

Since ASD risk genes are strongly LoF intolerant, and intolerant genes are more coexpressed with each other, we wanted to assess whether among the significant convergent genes were also tolerant genes associated with ASD. To define intolerance, we divided the genes into two sets, tolerant (LOEUF >0.35) and intolerant (LOEUF <0.35). The correlation was assessed between transcriptional convergence and the $-\log_{10}(p)$ significance of the exome data. Next, we sought to identify whether transcriptional convergence can identify novel ASD genes that have not been previously implicated due to limitations of genetic studies. We defined ASD association as having a total Bayes factor (BF) > 2 for the exome data which is includes the presence of even a weak association. A BF of greater than 3 is typically used to represent meaningful significance. We assessed the correlation between LOEUF and transcriptional convergence for both associated and non-associated ASD genes.

Classification of 71 ASD associated genes

We categorized 71 ASD associated genes into three main functional groups including chromatin, transcription and synaptic by their association with chromatin function related GO terms ("GOMF_CHROMATIN_BINDING", "GOMF_CHROMATIN_DNA_BINDING", "GOBP_REGULATION_OF_CHROMATIN_ASSEMBLY_OR_DISASSEMBLY","GOBP_CHROMATIN_ORGANIZATION","GOBP_CHROMATIN_REMODELING","GOBP_REGULATION_OF_CHROMATIN_BINDING","GOBP_CHROMATIN_MEDIATED_MAINTENANCE_OF_TRANSCRIPTION","GOBP_CHROMATIN_MAINTENANCE","GOBP_REGULATION_OF_CHROMATIN_ORGANIZATION") and transcription related GO terms ("GOBP_mRNA_TRANSCRIPTION", "GOBP_REGULATION_OF_TRANSCRIPTION_REGULATORY_REGION_DNA_BINDING", "GOBP_mRNA_TRANSCRIPTION_BY_RNA_POLYMERASE_II","GOBP_CHROMATIN_ORGANIZATION_INVOLVED_IN_REGULATION_OF_TRANSCRIPTION","GOMF_RNA_POLYMERASE_II_TRANSCRIPTION_FACTOR_BINDING","GOMF_TRANSCRIPTION_COREGULATOR_BINDING","GOMF_DNA_BINDING_TRANSCRIPTION_FACTOR_ACTIVITY","GOMF_TRANSCRIPTION_FACTOR_BINDING","GOMF_TRANSCRIPTION_REGULATOR_ACTIVITY") and synaptic function based on manual curation reported in SynGO v1.1 database.[52] The above listed GO terms and their associated genes were retrieved from MSigDB (v7.4) database. Genes that were not identified in any of these categories were manually further classified based on https://www.genecards.org/.