Autism-Associated Genes and Neighboring lncRNAs Converge on Key Gene Regulatory Networks

Rebecca E. Andersen[1,2,3,4,*], Maya Talukdar[1,2,3,5,*], Tyler Sakamoto[1,6], Janet H.T. Song[1,2,4,7,8], Xuyu Qian[1,2,4,7,8], Seungil Lee[1,6], Ryan N. Delgado[9], Sijing Zhao[1,10], Gwenyth Eichfeld[1,11], Julia Harms[1,12], and Christopher A. Walsh[1,2,3,4,7,8]

Affiliations:

[1]Division of Genetics and Genomics and Manton Center for Orphan Diseases, Boston Children's Hospital, Boston, MA, USA; [2]Department of Pediatrics, Harvard Medical School, Boston, MA, USA; [3]Broad Institute of MIT and Harvard, Cambridge, MA, USA; [4]Allen Discovery Center for Human Brain Evolution, Boston, MA, USA; [5]Harvard-MIT MD/PhD Program, Program in Biomedical Informatics, Boston, MA, USA; [6]Harvard College, Cambridge, MA, USA; [7]Department of Neurology, Harvard Medical School, Boston, MA, USA; [8]Howard Hughes Medical Institute, Boston Children's Hospital, Boston, MA, USA; [9]Department of Genetics, Blavatnik Institute, Howard Hughes Medical Institute, Harvard Medical School, Boston, MA, USA; [10]Harvard BBS PhD Program, Boston, MA, USA; [11]Colgate University, Hamilton, NY, USA; [12]University of California Berkeley, Berkeley, CA, USA.

*These authors contributed equally to this work.

Corresponding author: Christopher A. Walsh, christopher.walsh@childrens.harvard.edu

ABSTRACT

The diversity of genes implicated in autism spectrum disorder (ASD) creates challenges for identifying core pathophysiological mechanisms. Aggregation of seven different classes of genetic variants implicated in ASD, in a database we call *Consensus-ASD*, reveals shared features across distinct types of ASD variants. Functional interrogation of 19 ASD genes and 9 neighboring long non-coding RNAs (lncRNAs) using CRISPR-Cas13 strikingly revealed differential gene expression profiles that were significantly enriched for other ASD genes. Furthermore, construction of a gene regulatory network (GRN) enabled the identification of central regulators that exhibit convergently altered activity upon ASD gene disruption. Thus, this study reveals how perturbing distinct ASD-associated genes can lead to shared, broad dysregulation of GRNs with critical relevance to ASD. This provides a crucial framework for understanding how diverse genes, including lncRNAs, can play convergent roles in key neurodevelopmental processes and ultimately contribute to ASD.

INTRODUCTION

Autism spectrum disorder (ASD) is a common, male-biased neurodevelopmental disorder characterized by deficits in social communication as well as restricted and/or repetitive behaviors. Although ASD is clinically heterogeneous, these central features can be used diagnostically to distinguish ASD from other neurodevelopmental disorders (DSM-5, 2013), suggesting that certain biological underpinnings of ASD are shared across individuals. Moreover, ASD has a significant genetic component (Bai et al. 2019), and several large-scale studies have identified genes implicated in ASD (ASD genes) (Grove et al. 2019; Satterstrom et al. 2020; Trost et al. 2022). Such studies have enabled the compilation of ASD genes into large databases such as the Simons Foundation Autism Research Initiative (SFARI) Gene database (Abrahams et al. 2013). While SFARI genes are predominantly implicated in ASD through rare copy number variants (CNVs) and single-gene mutations, other types of genetic variation also contribute to ASD (Dias and Walsh 2020). For instance, genome wide association studies (GWAS) have identified single nucleotide polymorphisms (SNPs) associated with ASD (Grove et al. 2019). Mosaic (Rodin et al. 2021), structural (Marshall et al. 2008), and inherited recessive variants (Doan et al. 2019) have also been implicated in ASD, though the extent to which these latter variant classes impact ASD is less understood. While hundreds of ASD genes have now been identified across these genetic variant classes, the specific biological processes through which disruptions in these genes lead to ASD remain largely unknown.

One possible explanation of the wide diversity of genes implicated in ASD is that disruptions of different ASD genes could lead to core ASD pathophysiology through convergence on shared downstream targets. Consistent with this, common patterns of differentially expressed genes (DEGs) have been identified in brain tissue from ASD cases versus neurotypical controls (Voineagu et al. 2011; Parikshak et al. 2016; Ramaswami et al. 2020; Gandal et al. 2022). Perturbation studies of distinct ASD genes have also identified certain overlapping changes in cell type composition and behavior, however evidence of shared effects on gene expression is

more limited (Paulsen et al. 2022; Weinschutz Mendes et al. 2023; Li et al. 2023), and thus the molecular details of how such convergence arises remain unclear (Liao et al. 2023).

While previous studies have primarily focused on assessing convergence of protein-coding genes (PCGs) implicated in ASD, there is emerging evidence that non-coding sequences can also play important roles in ASD. For instance, long non-coding RNAs (lncRNAs) are frequently dysregulated in ASD (Parikshak et al. 2016), and disruptions in multiple lncRNAs have now been implicated in ASD (Noor et al. 2010; Luo et al. 2018; Ang et al. 2019; Andersen et al. 2024), suggesting that mutations in lncRNAs can directly lead to ASD. Moreover, certain classes of lncRNAs have often been found to modulate the expression and/or function of their neighbors (Luo et al. 2016; Wang et al. 2020), raising the possibility that lncRNAs could play critical roles in ASD-relevant molecular pathways even if they are not directly disrupted.

Here, we first present *Consensus-ASD*, a database comprised of genetic variants implicated in ASD across seven different genetic variant classes from 32 separate studies. Analysis of *Consensus-ASD* revealed shared genes and molecular pathways across different types of genetic variation, further illuminating the genetic architecture of ASD. Subsequently, functional interrogation of 19 ASD genes and 9 neighboring lncRNAs uncovered remarkably widespread convergence on shared differentially expressed genes (DEGs), with most perturbations resulting in DEGs enriched for *Consensus-ASD* genes. Key transcription factors (TFs) whose targets were consistently enriched within the DEGs included *ZFX*, an X-linked TF recently implicated in ASD (Shepherdson et al. 2024). *ZFX* mediates widespread sex-differential gene expression (San Roman et al. 2024) and thus could play an important role in the male-biased prevalence of ASD. Furthermore, reconstruction of a gene regulatory network (GRN) enabled the identification of 78 central regulators, including well known ASD genes such as *CHD8*, that are top candidates for driving molecular convergence in ASD. Disruption of several ASD genes and neighboring lncRNAs led to phenotypic convergence as well, affecting the critical process of neural progenitor cell (NPC) proliferation. Taken together, these results reveal how key GRNs

can become broadly dysregulated upon disruption of a single ASD gene, providing important insights into the molecular mechanisms of genetic convergence in ASD.

RESULTS

Consensus-ASD Reveals Shared Features Across Distinct Classes of Genetic Variation

Convergence in ASD has primarily been investigated in the context of highly constrained genes (Lek et al. 2016) in which rare *de novo* copy number variants (CNVs) or single nucleotide variants (SNVs) are implicated in ASD. However, ASD has a complex genetic architecture with contributions from multiple distinct genetic variant classes (Iakoucheva et al. 2019; Dias and Walsh 2020; Willsey et al. 2022) (**Fig. 1A**). Previous work suggests that these different variant classes can confer additive risk for ASD (Antaki et al. 2022), perhaps through jointly affecting similar pathways underlying ASD pathogenesis. To test this, we first systematically aggregated variants implicated in ASD across seven different genetic variant classes from 32 different studies: 1) germline SNVs and indels implicated in idiopathic ASD, 2) syndromic variants with ASD as a recurrent feature, 3) mosaic SNVs and indels, 4) germline CNVs, 5) mosaic CNVs, 6) inherited recessive variants, and 7) common variation (GWAS SNPs). We refer to this dataset as *Consensus-ASD* (**Fig. S1A**).

Germline SNVs or indels implicate 1,658 genes in ASD. Since ASD can be diagnosed independently or as a core feature of a larger clinical syndrome, 446 genes are additionally specified as having been implicated in syndromes involving ASD (category: syndromic). *Consensus-ASD* also includes 76 genes implicated in ASD by inherited recessive variants and 51 genes implicated by mosaic SNVs or indels. CNVs, including both germline and mosaic, have also strongly been implicated in ASD. As CNVs frequently overlap multiple genes, the CNVs are annotated in two ways: 1) with all genes that they physically overlap ("all genes" approach: 5,147 genes for germline CNVs and 12,181 genes for mosaic CNVs), and 2) with only the overlapped genes identified in at least one other variant class in *Consensus-ASD*

("known ASD genes" approach: 276 genes for germline CNVs and 534 genes for mosaic CNVs). Finally, common variation was assessed by including single nucleotide polymorphisms (SNPs) that have been implicated in ASD through genome-wide association studies (GWAS). As expected, the majority of significant GWAS SNPs exist in non-coding regions of the genome; thus, the SNP-to-gene linking method "combined SNP-to-gene" (cS2G) (Gazal et al. 2022) was used to determine the likely target gene of such SNPs, identifying 166 genes implicated in ASD (cS2G score > 0.5).

Consensus-ASD genes are generally intolerant to functional variation (e.g., constrained), as determined by pLOEUF score percentiles (Karczewski et al. 2020) (**Fig. 1B**, **Fig. S1B**, and **Table S1**). As expected, the syndromic category followed by the germline SNV or indel category had the highest median pLOEUF score percentiles of 0.894 (adj. p = 2.21e-128) and 0.808 (adj. p = 2.61e-212), respectively, consistent with previous reports that these genes are highly constrained (Karczewski et al. 2020). The recessive category had a median constraint percentile of 0.555 (adj. p = 0.0504), as expected for genes associated with disease via autosomal recessive modes of inheritance (Karczewski et al. 2020). In contrast to germline SNVs/indels, mosaic SNVs tended to be found in genes that were moderately constrained (median percentile = 0.680, adj. p = 4.49e-4), which may reflect competitive selection against cells with mutations that disrupt highly constrained genes in a mosaic context. Germline and mosaic CNVs tended to overlap only moderately constrained genes when considering all overlapped genes (median percentile = 0.573 and 0.515, adj. p = 8.93e-08 and 0.0064, respectively), although the median constraint of overlapped genes increased to 0.868 (adj. p = 1.81e-47) and 0.776 (adj. p = 4.78e-46) when implementing the "known ASD genes" approach. Finally, genes implicated in ASD through GWAS SNPs were moderately constrained (median percentile = 0.532, adj. p = 0.0396); however, these SNPs tended to be in highly constrained non-coding regions of the genome, as determined by the Gnocchi metric for calculating non-coding constraint (Chen et al. 2024) (median Gnocchi percentile = 0.85, adj. p = 1.58e-16).

Thus, despite the relatively moderate constraint of the putative target genes of these SNPs, non-coding variation in these regions is strongly depleted.

Transcriptomic analysis reveals that genes implicated in ASD through different classes of genetic variation exhibit distinct patterns of expression in the brain across development. Analysis of *BrainSpan* (https://www.brainspan.org/), a large bulk RNA-sequencing (RNA-seq) dataset of brain tissue encompassing diverse ages, revealed that genes implicated in ASD through germline SNVs are predominantly expressed in the fetal brain (**Fig. 1C**), highlighting the relevance of such ASD genes to neurodevelopment. This was also the case for syndromic and inherited recessive variants. In contrast, common variation (GWAS SNPs) more prominently implicated genes expressed postnatally, potentially reflecting a distinct biological role for common variation in ASD pathogenesis. Known ASD genes from mosaic CNVs were strongly expressed in early development, similar to germline SNVs or indels, while genes from germline CNVs predominantly exhibited postnatal expression, which may suggest that germline CNVs affecting genes with critical roles in early neurodevelopment tend to be lethal or result in more severe phenotypes than ASD. Mosaic coding SNVs also predominantly exhibited postnatal expression, which may similarly reflect competitive selection against cells with mutations disrupting key neurodevelopmental genes over the course of fetal development. Furthermore, analysis of single-cell RNA-seq (scRNA-seq) data from mid-gestation neocortex (Polioudakis et al. 2019) identified strong enrichment across nearly all variant classes in deep-layer excitatory neurons (**Fig. 1D**). While this has been suggested in previous reports (Willsey et al. 2013), *Consensus-ASD* expands this finding across diverse classes of genetic variation, including rare, mosaic, and common variants.

<u>Distinct Variant Classes Implicate Shared Molecular Pathways and Genes in ASD</u>

Key biological processes are implicated by multiple different classes of genetic variation in ASD (**Fig. 1E**). Analysis of the fifty Molecular Signature Database Hallmark Pathways (Liberzon

et al. 2015) uncovered that germline SNVs and indels are strongly enriched for certain signaling pathways, including the WNT/beta-catenin signaling pathway (adj. p = 2.7e-3) that has been previously implicated in ASD (Kalkman 2012). Additionally, they exhibited strong enrichment for pathways associated with cell division such as G2M checkpoint (adj. p = 1.4e-5) and mitotic spindle (adj. p = 3.9e-6), which was also the case for syndromic variants (adj. p = 2.0e-3 and 2.4e-3, respectively). In contrast, GWAS SNPs were primarily associated with pathways involved in immune signaling such as allograft rejection (adj. p = 2.8e-3) and interferon gamma response (adj. p = 8.9e-3), potentially suggesting a different functional role for common variation versus rare variation in ASD. Shared enrichment was also found between mosaic CNVs and both GWAS SNPs and germline SNVs/indels, further suggesting that these distinct variant classes can affect shared biological processes.

Furthermore, different types of genetic variation implicated the same specific genes in ASD (**Fig. 1F** and **Fig. S1C**). Strong cross-enrichment of the same genes was found across the syndromic, germline SNVs or indels, inherited recessive, and CNVs categories, even when using the “all genes” approach for CNVs. Cross-enrichment between GWAS SNPs and mosaic CNVs was also found, as had been observed through pathway analysis (**Fig. 1E**). Taken together, this suggests that there is functional and genic overlap amongst rare variant classes in ASD, with some evidence of overlap with and between common and mosaic variation.

<u>Depletion of ASD Genes and Adjacent lncRNAs Similarly Affects Gene Expression</u>

Disruption of individual ASD genes or their adjacent lncRNAs resulted in unexpectedly high degrees of transcriptional convergence on shared downstream targets. Functional studies were performed for a set of 19 pairs of highly constrained SFARI genes and their neighboring lncRNAs, as closely adjacent lncRNAs have often been found to regulate the expression and/or function of their neighbors (Villegas and Zaphiropoulos 2015; Luo et al. 2016; Wang et al. 2020). This included the SFARI-adjacent lncRNA *NR2F1-AS1*, which has itself been implicated

in ASD (Ang et al. 2019). The lncRNA *SOX2-OT*, a disruption in which has recently been reported in an individual with ASD (Andersen et al. 2024), was also included along with its overlapped protein-coding gene (PCG) *SOX2*. For each of these genes (**Table S2**), CRISPR-Cas13 (Konermann et al. 2018; Wei et al. 2023) and an array of three guide RNAs (gRNAs) for each gene (**Table S3**) were used to knock down (KD) the target gene (**Fig. 1A**) in human NPC cultures (**Fig. 1B**). RNA-seq was performed after 24 hours to determine the acute effects of these perturbations.

Most (28/38, 73.7%) of the target genes exhibited statistically significant KD (**Fig. 2C** and **Table S4**), including 19/20 (95%) of the PCGs and 9/18 (50%) of the lncRNAs. Differences in KD efficiency between the PCGs and lncRNAs may reflect reduced accuracy of lncRNA transcript annotations or secondary structures within the lncRNAs that are difficult for Cas13 to target. Subsequent analyses were limited to the 28 conditions that demonstrated significant KD.

While no lncRNA (0/19) was significantly differentially expressed upon KD of its neighboring PCG, most (5/9, 55.6%) successful lncRNA KDs led to significant differential expression of their neighboring PCG (**Fig. 2D** and **Fig. S2A-B**). Of these, most (4/5, 80%) of the lncRNA KDs resulted in decreased expression of the PCG neighbor, consistent with previous reports that closely adjacent lncRNAs often positively regulate the expression of their nearby neighbors (Villegas and Zaphiropoulos 2015; Luo et al. 2016; Wang et al. 2020). Transcriptome-wide analysis, limited to the 25 samples with at least 50 differentially expressed genes (DEGs), revealed a strong and significant correlation in fold change of DEGs between each neighboring lncRNA-PCG pair ($p < 2.2e-16$ in each comparison; testing beta $\neq 0$) (**Fig. S2C**). Linear modeling determined that lncRNA-KD fold changes were smaller than PCG-KD fold changes on a per-gene basis ($p < 2.2e-16$ in each comparison; testing beta < 0) (**Fig. S2C**), suggesting that perturbation of a lncRNA and its neighboring PCG may lead to similar downstream transcriptional changes but that the magnitude of the effect tends to be stronger upon KD of the PCG.

Linear modeling analysis of certain subsets of DEGs further reveals shared and unique effects of the lncRNA and PCG neighbors. When linear modeling analysis is limited to genes that are differentially expressed in both the lncRNA and PCG KDs, the fold changes are much more similar (**Fig. S2D**), suggesting that a subset of target genes are similarly affected by disruption of either the lncRNA or the neighboring PCG. However, limiting this analysis to DEGs from the PCG KDs (irrespective of whether they are DEGs upon lncRNA KD) once again demonstrates a strong correlation but larger fold changes in the PCG KDs (**Fig. S2E**), similar to the observations from all expressed genes (**Fig. S2C**). This may reflect that the PCG KDs resulted in a greater number of DEGs than the lncRNA KDs for most (7/8, 87.5%) of the lncRNA-PCG pairs, leading to greater similarity with the analysis of all expressed genes. In contrast, limiting the analysis to the DEGs from lncRNA KDs revealed negative correlations between three of the lncRNA-PCG pairs (**Fig. S2F**), indicating that KD of the lncRNA from these pairs leads to unique effects on a subset of genes that are affected in the opposite direction by KD of the neighboring PCG. Taken together, this suggests that lncRNA and neighboring PCG KDs generally lead to similar changes in gene expression, but that the lncRNAs tend to have more subtle effects and can also uniquely affect a subset of targets.

<u>Individual ASD-Related Perturbations Result in Widespread Dysregulation of ASD Genes</u>

Further differential expression analysis revealed remarkable convergence on shared ASD DEGs. Grouping all 25 ASD-relevant perturbations together and comparing to the set of 3 non-targeting control (NTC) samples identified 1,507 DEGs (excluding the directly targeted genes) (**Fig. 2E**). These DEGs were strongly enriched for genes in *Consensus-ASD* (p = 4.41e-9), and this enrichment persisted when considering only high confidence SFARI (HC-SFARI) genes (p = 2.78e-7), the SFARI genes in the highest evidentiary tier for association with ASD. Genes implicated in intellectual disability (SysID) (Kochinke et al. 2016) exhibited a significant but weaker enrichment (p = 0.0003), particularly when considering genes specifically implicated in

ID but not ASD (p = 0.01). This suggests that perturbation of known ASD genes and neighboring lncRNAs leads to broad disruption of other ASD genes.

Nearly every set of DEGs from the individual KDs also exhibited enrichment for genes in *Consensus-ASD* (21/25, 84%) (**Fig. 3A**) as well as SFARI genes (19/25, 76%) (**Fig. S3A**). This enrichment was uniformly weaker when considering genes implicated in intellectual disability (SysID) (**Fig. S3A**), further suggesting that perturbation of ASD genes predominantly affects other ASD genes and not neural genes more broadly. Additionally, there was a striking pattern of convergence on the same individual DEGs across the different perturbations, particularly when focusing on HC-SFARI genes (**Fig. 3B**). Not only were the same specific genes shared across multiple perturbations, but they were also largely affected in the same direction by the different KDs. Moreover, HC-SFARI genes tended to shift in a concordant direction as a result of the different perturbations even when they did not rise to the level of a statistically significant DEG within a particular KD sample (**Fig. 3C** and **S3B**).

A similar degree of convergence was found upon reanalysis of previously published RNA-seq data from CRISPR-Cas9 knockouts (KOs) of ten ASD genes – none of which were targeted in our study – in human induced pluripotent stem cells (iPSCs) and excitatory neurons (ExNs) (Deneault et al. 2018). Despite the differences in type of genetic perturbation, target genes, and cell types as compared to our study, similar levels of cross-enrichment were found within but not between distinct cell types (**Fig. S3C**). As these analyses were restricted to only genes that are expressed in all three cell types, this is not attributable to cell-type-specific expression of DEGs but instead suggests that distinct genes are affected by ASD gene perturbations in different cell types. Nevertheless, the convergence of ASD genes on recurrent DEGs is a shared feature across cell types.

Gene modules known to be dysregulated in tissue from ASD individuals (Parikshak et al. 2016; Gandal et al. 2022; Zhang et al. 2023) or organoid models of ASD (Paulsen et al. 2022) were also enriched in the DEGs from many of the KD samples (**Fig. 3D**), although this was

primarily seen in perturbations of PCGs as opposed to lncRNAs. Notably, the two lncRNAs in our study that have previously been implicated in ASD, *NR2F1-AS1* (Ang et al. 2019) and *SOX2-OT* (Andersen et al. 2024), did exhibit enrichment for several of these modules.

Key Transcription Factors Target Recurrently Differentially Expressed Genes

The DEGs of many perturbations are consistently enriched with target genes of key transcription factors (TFs), suggesting that these TFs may mediate convergence in ASD. Enrichment analysis using the bioinformatics tool MAGIC (Roopra 2020) identified 316 TFs that were significant regulators in at least one perturbation, 26 of which were significant in at least 20/25 (80%) of the perturbations (recurrent MAGIC TFs) (**Fig. 3E**). Many of these significant TFs were involved in pathways or complexes previously linked to ASD, such as SWI/SNF (Wenderski et al. 2020; Valencia et al. 2023), WNT (Kalkman 2012), and CTCF (Price et al. 2023), or had themselves been implicated in ASD. Of the recurrent MAGIC TFs that had not been directly implicated in ASD, several were highly constrained. Mutations in these genes may cause severe phenotypes, preventing their discovery through genetic studies of ASD; however, co-expression with other known ASD risk genes can be used to implicate such highly constrained genes in ASD (Liao et al. 2023). Consistently, nearly all recurrent MAGIC TFs exhibited significant co-expression with SFARI ASD genes in *BrainSpan* (21/26, 80.8%) and/or in our NPC RNA-seq data (21/26, 80.8%) (**Fig. 3E**), and this was true even for several TFs that had not previously been associated with ASD. Thus, these TFs may represent an important class of genes relevant to ASD genetic networks that have been “hidden” from sequencing studies of ASD individuals, perhaps by resulting in other phenotypes when disrupted.

One noteworthy TF is RE1 silencing transcription factor (REST), a master negative regulator of neurogenesis that was a significant MAGIC TF in 24/25 (96%) of our perturbations. REST is a highly constrained gene (pLOEUF = 0.125; 78th percentile) not previously directly associated with ASD. Re-analysis of REST binding sites in both human and mouse embryonic stem cells

(ESCs) (Rockowitz and Zheng 2015) revealed strong enrichment of SFARI and *Consensus-ASD* genes within REST target genes from both human (adj. p = 2.7e-47 and 1.0e-43, respectively) and mouse (adj. p = 4.2e-40 and 4.2e-40, respectively) (**Fig. S3D**). Genes implicated in intellectual disability (SysID) were also enriched (adj. p = 7.6e-13 and 2.3e-10, respectively), although this enrichment was notably weaker than that for ASD genes, suggesting that REST may be especially important in ASD pathogenesis.

Another recurrent MAGIC TF of particular interest is ZFX, an X-linked transcription factor that is expressed from both the active and "inactive" X chromosome in females. Re-analysis of RNA-sequencing data from ZFX KD in XX fibroblasts (San Roman et al. 2024) revealed significant enrichment of ASD genes within the DEGs (SFARI genes: adj. p = 5.56e-05; HC-SFARI genes: adj. p = 1.76e-03; *Consensus-ASD* genes: adj. p = 6.92e-15), suggesting ZFX may regulate the expression of genes implicated in ASD. Moreover, a set of genes that exhibits female-biased expression in neurotypical individuals as well as downregulation in autistic individuals, called $asdM12_V$ (Werling et al. 2016), was also significantly enriched (adj. p = 5.08e-03) (**Fig. 3F**), supporting a potential role of ZFX in mediating the sex-biased expression of genes that are dysregulated in ASD.

<u>Gene Regulatory Network Analysis Identifies Top Candidates for Driving Convergence in ASD</u>

Reconstruction of a gene regulatory network (GRN) enabled the characterization of important regulatory relationships in NPCs. The bioinformatic tool ARACNe (Margolin et al. 2006; Lachmann et al. 2016) was used to build a GRN linking transcription factors (Lambert et al. 2018) and epigenetic regulators (Boukas et al. 2019) to their downstream targets (regulons), using the data processing inequality applied to the inferred mutual information between individual genes to distinguish direct and indirect regulatory relationships. Separately, a GRN built from data excluding the 7 samples in which the KD target was a potential regulator (*BAZ2B*, *EHMT1*, *KMT2E*, *NR2F1*, *SETD1B*, *SETD2*, and *SETD5*) enabled evaluation of the

predictive ability of the ARACNe-derived GRN through comparison with the held-out RNA-seq data. Strong and significant cross-enrichment was found between 2nd degree ARACNe regulons (i.e., direct targets as well as their direct targets) and downregulated DEGs for all 7/7 (100%) of the KD targets ($p < 2.2e-16$, Fisher's combined probability test) (**Fig. 4A** and **S4**). Thus, ARACNe is able to effectively capture true regulatory relationships between genes of interest, and so we then built an ARACNe GRN from our full set of RNA-seq data (**Fig. S5**). In total, this identified 101,918 putative regulatory relationships between 1,098 known transcription factors or epigenetic regulators and 9,464 target genes, including many highly interconnected relationships between known ASD genes (**Fig. 4B**).

Analysis using the PageRank algorithm (Page et al. 1999) demonstrated that genes implicated in ASD are central regulators of the NPC GRN. PageRank orders nodes (genes) based on their importance to the overall network given the number and strength of their interactions. Gene set enrichment analysis (GSEA) determined that genes in *Consensus-ASD* were strongly enriched for high PageRank scores (adj. $p = 1.3e-9$) (**Fig. S6**), suggesting that they are central regulators in the NPC GRN. This enrichment was also significant for individual genetic variant classes, such as genes in the germline SNVs or indels (adj. $p = 2.2e-9$) and syndromic (adj. $p = 6.7e-7$) categories, as well as for SFARI genes (adj. $p = 6.7e-4$), while SysID genes showed much weaker enrichment (adj. $p = 0.03$). Analysis of Hallmark Molecular Signature Databases pathways (Subramanian et al. 2005) further identified strong enrichment for genes associated with cell division and proliferation such as G2M checkpoint (adj. $p = 1.8e-3$) and mitotic spindle (adj. $p = 1.1e-4$) within the top PageRank genes, while analysis of the STRING database of protein-protein interactions (PPIs) (Szklarczyk et al. 2023) showed that genes with SET domains were enriched (adj. $p = 6.2e-4$). The SET domain is associated with methyltransferase activity and is found within several high confidence ASD genes; epigenetic modifiers of methylation may thus be particularly important in modulating GRNs in NPCs.

The bioinformatic tool VIPER (Alvarez et al. 2016) identified 145/1098 (13.2%) of the regulators that exhibited altered activity upon KD of ASD genes. Of these, 78 top candidates for driving convergence in ASD were prioritized based on having regulons enriched for genes with high PageRank scores and for genes in *Consensus-ASD* (**Fig. S7**). Over half of these top regulators are *Consensus-ASD* genes (41/78, 52.6%, adj. p = 1.5e-19), and over one quarter are HC-SFARI genes (21/78, 26.9%, adj. p = 2.e3-20). These top regulators exhibited remarkably strong cross-regulation (p = 0), as determined by mean normalized degree (**Fig. S8A**). Analysis of STRING (Szklarczyk et al. 2023) further confirmed that these top regulators are highly interconnected (p < 1e-16) (**Fig. 4C** and **Fig. S8B**). Taken together, this suggests that these top candidate regulators not only modulate many ASD genes but also cross-regulate each other, driving convergence in ASD.

Iterative walktrap clustering identified 5 modules of the top candidate regulators that were especially strongly interconnected (**Fig. 4D** and **Fig. S9**). Of particular interest, the largest of these (**Fig. S9C**) demonstrates a high degree of cross interaction throughout the module and contains several notable ASD genes including *CHD8*, one of the strongest known risk genes for ASD (Haddad Derafshi et al. 2022). The inferred *CHD8* regulon was strongly enriched (p = 2e-8) for genes that were previously identified as direct CHD8 targets through ChIP-seq from NPCs (Sugathan et al. 2014). The CHD8 ChIP-seq targets were also highly enriched for genes in *Consensus-ASD* (p = 3.38e-20), with a much weaker enrichment of SysID genes (p = 1.8e-5). The DEGs from KD of *KMT2E*, a regulator we perturbed that is in the same module as *CHD8*, also demonstrated enrichment for CHD8 ChIP-seq targets (p = 1.33e-4). Interestingly, *CHD8* is not differentially expressed upon *KMT2E* KD; similarly, *KMT2E* was not identified as a direct target of CHD8 by ChIP-seq (Sugathan et al. 2014). Consistent with this, ARACNe analysis did not identify a direct interaction between *CHD8* and *KMT2E*. This suggests that even genes that do not directly regulate each other may jointly regulate the same downstream genes, leading to convergent changes upon their disruption.

ASD Genes and Adjacent lncRNAs Alter NPC Proliferation

Convergent changes in gene expression also led to shared effects on NPC proliferation. Analysis using the heritable CytoTrack dye (**Fig. 4E**) revealed that 8/16 (50%) of the SFARI gene KDs led to a statistically significant decrease in proliferation (**Fig. 4F**). This was also the case for the lncRNA *SOX2-OT*, which has previously been implicated in ASD (Andersen et al. 2024), as well as its overlapped PCG *SOX2*, a TF known to regulate NPC proliferation (Graham et al. 2003; Lee et al. 2014). In contrast, 1/16 (6.25%) SFARI KDs and 3/8 (37.5%) lncRNA KDs led to increased NPC proliferation (**Fig. 4F**), while the remaining 11 perturbations led to insignificant and/or inconsistent effects (**Fig. S10A**). The RNA-seq data from these perturbations did not cluster based on proliferation phenotype (**Fig. S10B**), suggesting that phenotypic effects on processes such as proliferation may be difficult to predict from RNA-seq data alone, and underscoring the value of experimental analysis of such phenotypes. Together, these results reveal subsets of ASD genes and neighboring lncRNAs whose disruption leads to shared effects on gene expression as well as consistent phenotypes related to NPC proliferation.

DISCUSSION

Here, we first present *Consensus-ASD*, a compilation of ASD genes from seven different variant classes, which uncovered that distinct types of genetic variation implicate shared genes and molecular pathways in ASD. Furthermore, functional studies of ASD genes and neighboring lncRNAs in human NPCs demonstrated strong convergence on the widespread dysregulation of ASD genes. Recurrently differentially expressed genes implicated key transcription factors, including *REST* and *ZFX*, that may mediate these shared changes in gene expression. Moreover, GRN analysis revealed regulatory relationships that highlighted a set of central regulators, including *CHD8*, as critical drivers of convergence. These regulators were highly

interconnected, with many demonstrating mutual regulation of one another, revealing how disruption of a single regulator can lead to changes that propagate throughout the network. Several ASD gene or neighboring lncRNA perturbations also altered NPC proliferation, demonstrating that these overlapping changes in gene expression can result in shared phenotypic effects. Thus, this work revealed critical insights into how diverse genes, including lncRNAs, can converge on core transcriptomic changes and neurodevelopmental phenotypes, ultimately resulting in ASD.

To reconstruct a detailed GRN, this work focused on functional studies of ASD genes and neighboring lncRNAs in highly homogeneous human NPC cultures. This enabled bulk RNA-seq of each perturbation, providing the deep transcriptomic data that was essential for robust GRN reconstruction. While some of the specific regulatory relationships uncovered here may be unique to NPCs, this work further demonstrated that convergence of ASD genes on shared downstream targets is exhibited by diverse cell types, as re-analysis of RNA-seq data from ASD gene KO studies in iPSCs and excitatory neurons (Deneault et al. 2018) similarly revealed highly overlapping DEGs within each cell type. Thus, future studies to reconstruct GRNs from additional cell types such as excitatory neurons will be critical for understanding how regulatory relationships are reorganized in different contexts. Additionally, bioinformatic pipelines for reconstructing GRNs from scRNA-seq data are actively being refined (Vlahos et al. 2021; Obradovic et al. 2021), although these approaches are currently hampered by limited RNA capture inherent to the scRNA-seq techniques. As these continue to improve, it will be feasible to reconstruct robust GRNs from heterogeneous models, including organoids, as well as patient tissue samples, which will further refine our understanding of gene regulation in complex biological systems.

This study also identified several lncRNAs that affect the expression of their ASD gene neighbor or other ASD genes more broadly. Intriguingly, the transcriptomic changes resulting from the lncRNA KDs were highly correlated with those from the KD of their ASD gene

neighbors, though the lncRNAs had milder effects, consistent with previous findings that lncRNAs tend to be more subtle modulators of gene expression (Gao et al. 2020). This was also the case for *NR2F1-AS1* (Ang et al. 2019) and *SOX2-OT* (Andersen et al. 2024), lncRNAs that have been directly implicated in ASD. Thus, these lncRNAs may play important neurodevelopmental roles through their effects on ASD genes, providing context to previous findings that many lncRNAs are dysregulated in ASD (Parikshak et al. 2016). Moreover, lncRNAs are often missed or overlooked in genetic studies (Mattick et al. 2023), and these findings highlight the value of especially careful consideration of lncRNAs that closely neighbor known critical genes.

Taken together, these results have broad implications for our understanding of neurodevelopment and disorders such as ASD. In particular, this work identified several regulators that are top candidates for driving convergent changes in gene expression, including the transcription factor *REST*. While *REST* has not been directly implicated in ASD, individuals with ASD exhibit decreased expression of *REST* targets (Katayama et al. 2016), and treatment with a *REST* inhibitor can ameliorate deficits in social interaction in a valproic acid mouse model of ASD (Kawase et al. 2019). Our findings further support that *REST* is a critical regulator of ASD-associated gene expression and suggest a general paradigm for identifying regulators of transcriptomic changes in ASD that may ultimately suggest new therapeutic targets. More broadly, while this work focused on ASD, the approaches used here are widely applicable to different cell types and conditions. Thus, this study serves as a framework for uncovering critical regulators across a wide diversity of contexts, from development through disease.

ACKNOWLEDGEMENTS

We thank members of the Walsh Lab for helpful discussions. We are also grateful to Dr. August Yue Huang for technical feedback regarding statistical analyses, and to Drs. Adrianna San Roman, Marla Tharp, and David Page for their insights regarding the role of *ZFX*. This work was facilitated by the BCH Flow Cytometry Research Core, the BCH IDDRC Molecular Genetics Core Facility, the BCH Cellular Imaging Core, and Lingsheng Dong of the HMS Research Computing Core. This project was supported by Autism Speaks Postdoctoral Fellowship 13008 and NIGMS T32 GM007748 (R.E.A.); T32 GM007753 and T32 GM144273 (M.T.); the Harvard Program for Research in Science and Engineering and the Harvard College Research Program (T.S.); a Howard Hughes Medical Institute Fellowship from the Helen Hay Whitney Foundation and Grant K99MH136290 from NIMH/NIH (J.H.T.S); K99 NS135123 from the NINDS/NIH (X.Q.); the Howard Hughes Medical Institute and Grant K99EY034603 from NEI/NIH (R.N.D); the Harvard-Amgen Scholars Program (G.E.); and grants from the Simons Foundation and Autism BrainNet (953759), NIMH (U01MH106883), the Hock E. Tan and K. Lisa Yang Center for Autism Research at Harvard University, and the Allen Discovery Center program, a Paul G. Allen Frontiers Group advised program of the Paul G. Allen Family Foundation (C.A.W). C.A.W. is an investigator of the Howard Hughes Medical Institute.

FIGURE LEGENDS

Figure 1

A) Schematic of the major classes of genetic variants within *Consensus-ASD*. **B)** Genic constraint scores for each *Consensus-ASD* category (see also **Table S1**). * adj. $p < 0.05$; ** adj. $p < 0.01$; *** adj. $p < 0.001$; *** adj. $p < 0.0001$. **C)** Expression of *Consensus-ASD* genes within BrainSpan. **D)** Expression of *Consensus-ASD* genes within different cell types during mid-gestation neocortical development (data from Polioudakis et al., 2019). **E)** Enrichment of genetic variant classes for fifty Molecular Signature Database Hallmark pathways. **F)** Cross-enrichment of *Consensus-ASD* genes across variant classes. Throughout the figure, CNVs are mapped to "known ASD genes".

Figure 2

A) Experimental design schematic. **B)** Immunocytochemistry of iPSC-derived NPC cultures. **C)** Relative expression of the targeted gene within each KD sample compared to non-targeting control (NTC) samples. Only targets with statistically significant KD are shown (see **Table S4**). Error bars: 95% confidence interval. **D)** Proportion of lncRNAs significantly differentially expressed upon KD of their neighboring PCG, and proportion of PCGs significantly differentially expressed upon KD of their neighboring lncRNA. **E)** Volcano plot depicting differential expression from grouping all successful KD samples together and comparing to NTC samples.

Figure 3

A) Bar plot depicting enrichment of *Consensus-ASD* genes within the DEGs of each KD condition. **B)** Circos plot depicting HC-SFARI DEGs shared between different KD conditions. **C)** Heatmap showing the relative expression of HC-SFARI genes that are DEGs in at least 5 KD conditions. Statistically significant DEGs are indicated with *. **D)** Enrichment of gene modules

previously identified in other ASD studies within the DEGs of each KD condition. **E)** MAGIC analysis identifying transcriptional regulators whose targets are enriched within the DEGs from each KD. **F)** Volcano plot depicting differential expression upon KD of ZFX in fibroblasts (data from San Roman et al., 2024). Genes from the $asdM12_V$ module, which is enriched for genes that are downregulated in ASD as well as genes that exhibit female-biased expression (Werling et al., 2016), are highlighted in purple.

Figure 4

A) Volcano plot depicting differential expression upon KD of SETD5. Enrichment of 2nd degree target genes within downregulated DEGs is indicated with the adjusted p-value above the plot. **B)** Depiction of the NPC GRN constructed using ARACNE. All regulators are shown along with first degree target genes that are in *Consensus-ASD* or the HC-SFARI gene set. **C)** Protein-protein interactions between the top candidate regulators (data from the STRING database). See **Fig. S8B** for gene labels. **D)** Depiction of modules of top regulators identified through iterative walktrap clustering. **E)** Schematic of experimental design to analyze proliferation phenotypes. **F)** Relative CytoTrack intensity over time for each KD condition that exhibited statistically significant proliferation phenotypes that were consistent over time. Error bars indicate standard error of the mean. Statistical significance was determined using 2-way ANOVA. See **Table S5** for full statistical results.

Figure S1

A) Upset plot of *Consensus-ASD* genes. CNVs are mapped to “known ASD genes”. **B)** Constraint scores for each *Consensus-ASD* category (see also **Table S1**). CNVs are mapped to all overlapping genes. NC: non-coding. * adj. $p < 0.05$; ** adj. $p < 0.01$; *** adj. $p < 0.001$; *** adj. $p < 0.0001$. **C)** Cross-enrichment of *Consensus-ASD* genes across variant classes. CNVs are mapped to all overlapping genes.

Figure S2

A) Dot plot of relative expression of neighboring PCG and lncRNA pairs in successful KDs of PCGs. Significance of KD of the neighboring gene is given by the color of the dot. **B)** Dot plot of relative expression of neighboring PCG and lncRNA pairs in successful KDs of lncRNAs. Significance of KD of the neighboring gene is given by the color of the dot. **C-F)** Log_2 fold change upon KD of a PCG versus KD of its neighboring lncRNA. The identity line is shown as a dashed black line. Individual plots depict: (C) all expressed genes, (D) genes that are differentially expressed in both the PCG KD and the lncRNA KD irrespective of direction, (E) genes that are differentially expressed upon PCG KD, and (F) genes that are differentially expressed upon lncRNA KD.

Figure S3

A) Bar plot depicting enrichment of *Consensus-ASD* genes or SysID genes within the DEGs of each KD condition. **B)** Heatmap showing the relative expression of HC-SFARI genes that are DEGs in at least 1 KD condition. Statistically significant DEGs are indicated with *. **C)** Cross-enrichment of DEGs between different ASD gene perturbations in different cell types, indicated by color. NPC DEG are from this study, while iPSC and neuron DEGs are from Deneault et al., 2018. Only genes expressed in all three cell types were analyzed. Statistical significance was determined using a hypergeometric test with a Benjamini-Hochberg correction for multiple hypothesis testing. **D)** Bar plot depicting enrichment of REST targets within SFARI, *Consensus-ASD*, or SysID genes in human and mouse embryonic stem cells. REST target data from Rockowitz et al., 2015.

Figure S4

A-F) Volcano plots depicting differential expression upon KD of the gene indicated above the plot. Enrichment of 2nd degree target genes within downregulated DEGs is indicated with the

adjusted p-value above the plot.

Figure S5

Depiction of the NPC GRN constructed using ARACNE. Only the top 20 edges are shown.

Figure S6

Gene set enrichment analysis (GSEA) demonstrating enrichment of genes ordered by PageRank score within certain gene sets.

Figure S7

Circular network visualization depicting the 78 regulators that meet the following three criteria: 1) identified as a significantly differentially active regulator between control (NTC) samples and samples in which genes known to be associated with ASD had been perturbed; 2) has a regulon that is enriched for genes with high PageRank scores; and 3) has a regulon that is enriched for genes in *Consensus-ASD*. Genes from *Consensus-ASD* are highlighted in yellow, and HC-SFARI genes are depicted with a dark border.

Figure S8

A) Bar graph of mean normalized degree, a measure of centrality, across regulators from different gene sets. **B)** Depiction of protein-protein interactions between the top candidate regulators (data from the STRING database).

Figure S9

A-E) Depiction of protein-protein interactions between the top candidate regulators of each module (data from the STRING database). **A)** Module 1 (red): histone demethylase. **B)** Module

2 (yellow): regulation of transcription. **C)** Module 3 (green): histone methylation, especially H3-K36. **D)** Module 4 (blue): transcriptional regulation. **E)** Module 5 (purple): SWI/SNF.

Figure S10

A-B) Analysis of the effects on proliferation resulting from KD of target genes in NPCs, related to **Fig. 4E**. **A)** Relative CytoTrack intensity ($-\log_2$) over time for each KD condition that did not exhibit statistically significant proliferation phenotypes or exhibited phenotypes that were inconsistent over time. Error bars indicate standard error of the mean. Statistical significance was determined using 2-way ANOVA. See **Table S5** for full statistical results. **B)** PCA plot of RNA-seq from each KD, colored by the effect of that KD on proliferation.

REFERENCES

Abrahams BS, Arking DE, Campbell DB, et al (2013) SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). Mol Autism 4:36. https://doi.org/10.1186/2040-2392-4-36

Alvarez MJ, Shen Y, Giorgi FM, et al (2016) Functional characterization of somatic mutations in cancer using network-based inference of protein activity. Nat Genet 48:838–47. https://doi.org/10.1038/ng.3593

Andersen RE, Alkuraya IF, Ajeesh A, et al (2024) Chromosomal structural rearrangements implicate long non-coding RNAs in rare germline disorders. Hum Genet 143:921–938. https://doi.org/10.1007/s00439-024-02693-y

Ang CE, Ma Q, Wapinski OL, et al (2019) The novel lncRNA lnc-NR2F1 is pro-neurogenic and mutated in human neurodevelopmental disorders. Elife 8:. https://doi.org/10.7554/eLife.41770

Antaki D, Guevara J, Maihofer AX, et al (2022) A phenotypic spectrum of autism is attributable to the combined effects of rare variants, polygenic risk and sex. Nat Genet 54:1284–1292. https://doi.org/10.1038/s41588-022-01064-5

Bai D, Yip BHK, Windham GC, et al (2019) Association of Genetic and Environmental Factors With Autism in a 5-Country Cohort. JAMA Psychiatry 76:1035. https://doi.org/10.1001/jamapsychiatry.2019.1411

Boukas L, Havrilla JM, Hickey PF, et al (2019) Coexpression patterns define epigenetic regulators associated with neurological dysfunction. Genome Res 29:532–542. https://doi.org/10.1101/gr.239442.118

Chen S, Francioli LC, Goodrich JK, et al (2024) A genomic mutational constraint map using variation in 76,156 human genomes. Nature 625:92–100. https://doi.org/10.1038/s41586-023-06045-0

Deneault E, White SH, Rodrigues DC, et al (2018) Complete Disruption of Autism-Susceptibility Genes by Gene Editing Predominantly Reduces Functional Connectivity of Isogenic Human Neurons. Stem Cell Reports 11:1211–1225. https://doi.org/10.1016/j.stemcr.2018.10.003

Dias CM, Walsh CA (2020) Recent Advances in Understanding the Genetic Architecture of Autism. Annu Rev Genomics Hum Genet 21:289–304. https://doi.org/10.1146/annurev-genom-121219-082309

Doan RN, Lim ET, De Rubeis S, et al (2019) Recessive gene disruptions in autism spectrum disorder. Nat Genet 51:1092–1098. https://doi.org/10.1038/s41588-019-0433-8

Gandal MJ, Haney JR, Wamsley B, et al (2022) Broad transcriptomic dysregulation occurs across the cerebral cortex in ASD. Nature 611:532–539. https://doi.org/10.1038/s41586-022-05377-7

Gao F, Cai Y, Kapranov P, Xu D (2020) Reverse-genetics studies of lncRNAs—what we have learnt and paths forward. Genome Biol 21:93. https://doi.org/10.1186/s13059-020-01994-5

Gazal S, Weissbrod O, Hormozdiari F, et al (2022) Combining SNP-to-gene linking strategies to identify disease genes and assess disease omnigenicity. Nat Genet 54:827–836. https://doi.org/10.1038/s41588-022-01087-y

Graham V, Khudyakov J, Ellis P, Pevny L (2003) SOX2 functions to maintain neural progenitor identity. Neuron 39:749–65. https://doi.org/10.1016/s0896-6273(03)00497-5

Grove J, Ripke S, Als TD, 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

Haddad Derafshi B, Danko T, Chanda S, et al (2022) The autism risk factor CHD8 is a chromatin activator in human neurons and functionally dependent on the ERK-MAPK pathway effector ELK1. Sci Rep 12:22425. https://doi.org/10.1038/s41598-022-23614-x

Iakoucheva LM, Muotri AR, Sebat J (2019) Getting to the Cores of Autism. Cell 178:1287–1298. https://doi.org/10.1016/j.cell.2019.07.037

Kalkman H (2012) A review of the evidence for the canonical Wnt pathway in autism spectrum disorders. Mol Autism 3:10. https://doi.org/10.1186/2040-2392-3-10

Karczewski KJ, Francioli LC, Tiao G, 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

Katayama Y, Nishiyama M, Shoji H, et al (2016) CHD8 haploinsufficiency results in autistic-like phenotypes in mice. Nature 537:675–679. https://doi.org/10.1038/nature19357

Kawase H, Ago Y, Naito M, et al (2019) mS-11, a mimetic of the mSin3-binding helix in NRSF, ameliorates social interaction deficits in a prenatal valproic acid-induced autism mouse model. Pharmacol Biochem Behav 176:1–5. https://doi.org/10.1016/j.pbb.2018.11.003

Kochinke K, Zweier C, Nijhof B, et al (2016) Systematic Phenomics Analysis Deconvolutes Genes Mutated in Intellectual Disability into Biologically Coherent Modules. Am J Hum Genet 98:149–64. https://doi.org/10.1016/j.ajhg.2015.11.024

Konermann S, Lotfy P, Brideau NJ, et al (2018) Transcriptome Engineering with RNA-Targeting Type VI-D CRISPR Effectors. Cell 173:665-676.e14. https://doi.org/10.1016/j.cell.2018.02.033

Lachmann A, Giorgi FM, Lopez G, Califano A (2016) ARACNe-AP: gene network reverse engineering through adaptive partitioning inference of mutual information. Bioinformatics 32:2233–2235. https://doi.org/10.1093/bioinformatics/btw216

Lambert SA, Jolma A, Campitelli LF, et al (2018) The Human Transcription Factors. Cell 172:650–665. https://doi.org/10.1016/J.CELL.2018.01.029

Lee KE, Seo J, Shin J, et al (2014) Positive feedback loop between Sox2 and Sox6 inhibits neuronal differentiation in the developing central nervous system. Proc Natl Acad Sci U S A 111:2794–9. https://doi.org/10.1073/pnas.1308758111

Lek M, Karczewski KJ, Minikel E V., et al (2016) Analysis of protein-coding genetic variation in 60,706 humans. Nature 536:285–291. https://doi.org/10.1038/nature19057

Li C, Fleck JS, Martins-Costa C, et al (2023) Single-cell brain organoid screening identifies developmental defects in autism. Nature 621:373–380. https://doi.org/10.1038/s41586-023-06473-y

Liao C, Moyses-Oliveira M, De Esch CEF, et al (2023) Convergent coexpression of autism-associated genes suggests some novel risk genes may not be detectable in large-scale genetic studies. Cell Genomics 3:100277. https://doi.org/10.1016/j.xgen.2023.100277

Liberzon A, Birger C, Thorvaldsdóttir H, et al (2015) The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst 1:417–425. https://doi.org/10.1016/j.cels.2015.12.004

Luo S, Lu JY, Liu L, et al (2016) Divergent lncRNAs Regulate Gene Expression and Lineage Differentiation in Pluripotent Cells. Cell Stem Cell 18:637–652. https://doi.org/10.1016/j.stem.2016.01.024

Luo T, Liu P, Wang X-Y, et al (2018) Effect of the autism-associated lncRNA Shank2-AS on architecture and growth of neurons. J Cell Biochem. https://doi.org/10.1002/jcb.27471

Margolin AA, Nemenman I, Basso K, et al (2006) ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC Bioinformatics 7 Suppl 1:S7. https://doi.org/10.1186/1471-2105-7-S1-S7

Marshall CR, Noor A, Vincent JB, et al (2008) Structural variation of chromosomes in autism spectrum disorder. Am J Hum Genet 82:477–88. https://doi.org/10.1016/j.ajhg.2007.12.009

Mattick JS, Amaral PP, Carninci P, et al (2023) Long non-coding RNAs: definitions, functions, challenges and recommendations. Nat Rev Mol Cell Biol 24:430–447. https://doi.org/10.1038/s41580-022-00566-8

Noor A, Whibley A, Marshall CR, et al (2010) Disruption at the PTCHD1 Locus on Xp22.11 in Autism spectrum disorder and intellectual disability. Sci Transl Med 2:49ra68. https://doi.org/10.1126/scitranslmed.3001267

Obradovic A, Chowdhury N, Haake SM, et al (2021) Single-cell protein activity analysis identifies recurrence-associated renal tumor macrophages. Cell 184:2988-3005.e16. https://doi.org/10.1016/J.CELL.2021.04.038

Page L, Brin S, Motwani R, Winograd T (1999) The PageRank Citation Ranking□: Bringing Order to the Web. The Web Conference

Parikshak NN, Swarup V, Belgard TG, et al (2016) Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. Nature 540:423–427. https://doi.org/10.1038/nature20612

Paulsen B, Velasco S, Kedaigle AJ, et al (2022) Autism genes converge on asynchronous development of shared neuron classes. Nature 602:268–273. https://doi.org/10.1038/s41586-021-04358-6

Polioudakis D, de la Torre-Ubieta L, Langerman J, et al (2019) A Single-Cell Transcriptomic Atlas of Human Neocortical Development during Mid-gestation. Neuron 103:785-801.e8. https://doi.org/10.1016/j.neuron.2019.06.011

Price E, Fedida LM, Pugacheva EM, et al (2023) An updated catalog of CTCF variants associated with neurodevelopmental disorder phenotypes. Front Mol Neurosci 16:1185796. https://doi.org/10.3389/fnmol.2023.1185796

Ramaswami G, Won H, Gandal MJ, et al (2020) Integrative genomics identifies a convergent molecular subtype that links epigenomic with transcriptomic differences in autism. Nat Commun 11:4873. https://doi.org/10.1038/s41467-020-18526-1

Rockowitz S, Zheng D (2015) Significant expansion of the REST/NRSF cistrome in human versus mouse embryonic stem cells: potential implications for neural development. Nucleic Acids Res 43:5730–43. https://doi.org/10.1093/nar/gkv514

Rodin RE, Dou Y, Kwon M, et al (2021) The landscape of somatic mutation in cerebral cortex of autistic and neurotypical individuals revealed by ultra-deep whole-genome sequencing. Nat Neurosci 24:176–185. https://doi.org/10.1038/s41593-020-00765-6

Roopra A (2020) MAGIC: A tool for predicting transcription factors and cofactors driving gene sets using ENCODE data. PLoS Comput Biol 16:e1007800. https://doi.org/10.1371/journal.pcbi.1007800

San Roman AK, Skaletsky H, Godfrey AK, et al (2024) The human Y and inactive X chromosomes similarly modulate autosomal gene expression. Cell genomics 4:100462. https://doi.org/10.1016/j.xgen.2023.100462

Satterstrom FK, Kosmicki JA, Wang J, et al (2020) Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism Article Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. Cell 1–17. https://doi.org/10.1016/j.cell.2019.12.036

Shepherdson JL, Hutchison K, Don DW, et al (2024) Variants in ZFX are associated with an X-linked neurodevelopmental disorder with recurrent facial gestalt. Am J Hum Genet 111:487–508. https://doi.org/10.1016/j.ajhg.2024.01.007

Subramanian A, Tamayo P, Mootha VK, et al (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences 102:15545–15550. https://doi.org/10.1073/pnas.0506580102

Sugathan A, Biagioli M, Golzio C, et al (2014) CHD8 regulates neurodevelopmental pathways associated with autism spectrum disorder in neural progenitors. Proc Natl Acad Sci U S A 111:E4468-77. https://doi.org/10.1073/pnas.1405266111

Szklarczyk D, Kirsch R, Koutrouli M, et al (2023) The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res 51:D638–D646. https://doi.org/10.1093/nar/gkac1000

Trost B, Thiruvahindrapuram B, Chan AJS, et al (2022) Genomic architecture of autism from comprehensive whole-genome sequence annotation. Cell 185:4409-4427.e18. https://doi.org/10.1016/j.cell.2022.10.009

Valencia AM, Sankar A, van der Sluijs PJ, et al (2023) Landscape of mSWI/SNF chromatin remodeling complex perturbations in neurodevelopmental disorders. Nat Genet 55:1400–1412. https://doi.org/10.1038/s41588-023-01451-6

Villegas VE, Zaphiropoulos PG (2015) Neighboring gene regulation by antisense long non-coding RNAs. Int J Mol Sci 16:3251–66. https://doi.org/10.3390/ijms16023251

Vlahos L, Obradovic A, Worley J, et al (2021) Systematic, Protein Activity-based Characterization of Single Cell State. bioRxiv 2021.05.20.445002

Voineagu I, Wang X, Johnston P, et al (2011) Transcriptomic analysis of autistic brain reveals convergent molecular pathology. Nature 474:380–4. https://doi.org/10.1038/nature10110

Wang Y, Chen S, Li W, et al (2020) Associating divergent lncRNAs with target genes by integrating genome sequence, gene expression and chromatin accessibility data. NAR Genom Bioinform 2:lqaa019. https://doi.org/10.1093/nargab/lqaa019

Wei J, Lotfy P, Faizi K, et al (2023) Deep learning and CRISPR-Cas13d ortholog discovery for optimized RNA targeting. Cell Syst 14:1087-1102.e13. https://doi.org/10.1016/j.cels.2023.11.006

Weinschutz Mendes H, Neelakantan U, Liu Y, et al (2023) High-throughput functional analysis of autism genes in zebrafish identifies convergence in dopaminergic and neuroimmune pathways. Cell Rep 42:112243. https://doi.org/10.1016/j.celrep.2023.112243

Wenderski W, Wang L, Krokhotin A, et al (2020) Loss of the neural-specific BAF subunit ACTL6B relieves repression of early response genes and causes recessive autism. Proc Natl Acad Sci U S A 117:10055–10066. https://doi.org/10.1073/pnas.1908238117

Werling DM, Parikshak NN, Geschwind DH (2016) Gene expression in human brain implicates sexually dimorphic pathways in autism spectrum disorders. Nat Commun 7:10717. https://doi.org/10.1038/ncomms10717

Willsey AJ, Sanders SJ, Li M, 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

Willsey HR, Willsey AJ, Wang B, State MW (2022) Genomics, convergent neuroscience and progress in understanding autism spectrum disorder. Nat Rev Neurosci 23:323–341. https://doi.org/10.1038/s41583-022-00576-7

Zhang P, Omanska A, Ander BP, et al (2023) Neuron-specific transcriptomic signatures indicate neuroinflammation and altered neuronal activity in ASD temporal cortex. Proceedings of the National Academy of Sciences 120:e2206758120. https://doi.org/10.1073/pnas.2206758120

(2013) Diagnostic and statistical manual of mental disorders: DSM-5^{TM}, 5th ed. American Psychiatric Publishing, Inc., Arlington, VA, US

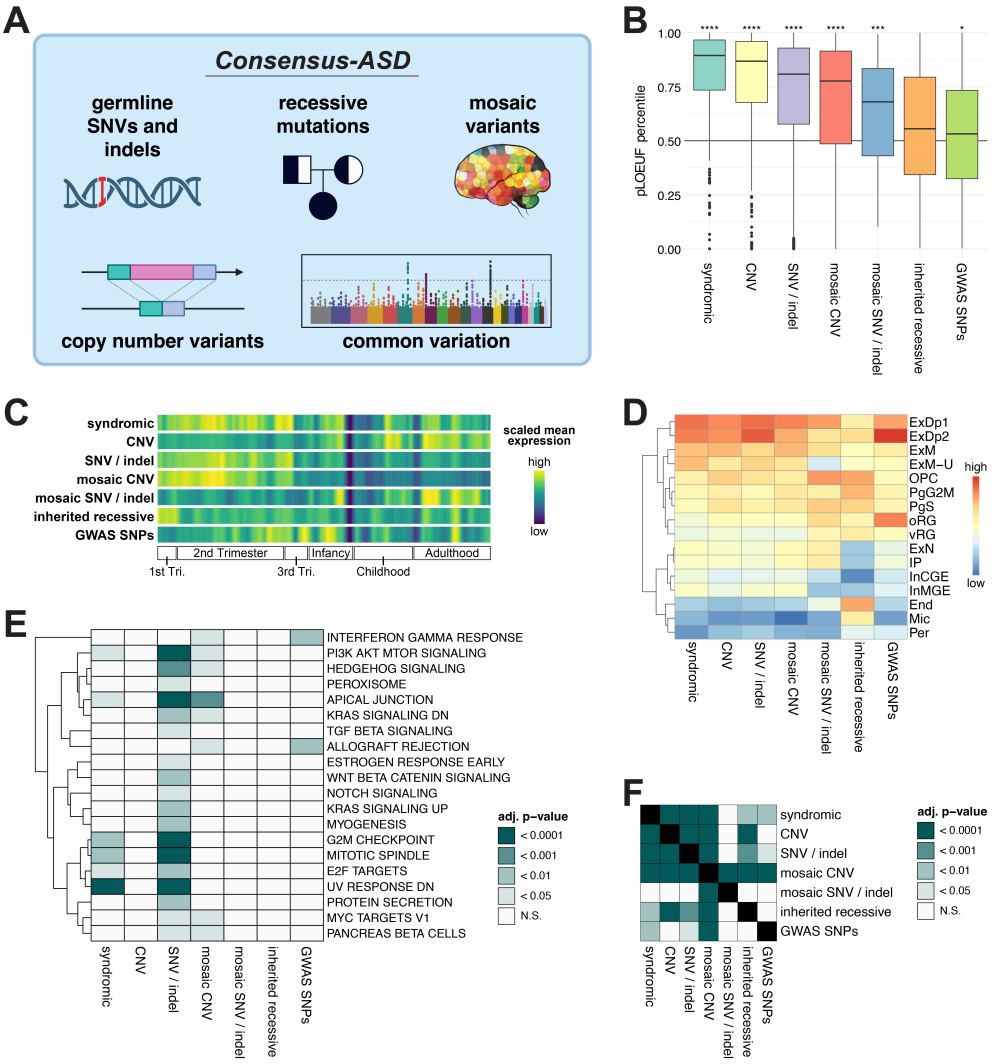

A
Consensus-ASD
germline SNVs and indels
recessive mutations
mosaic variants
copy number variants
common variation
B
pLOEUF percentile
1.00
0.75
0.50
0.25
0.00
syndromic
CNV
SNV / indel
mosaic CNV
mosaic SNV / indel
inherited recessive
GWAS SNPs
C
syndromic
CNV
SNV / indel
mosaic CNV
mosaic SNV / indel
inherited recessive
GWAS SNPs
scaled mean expression
high
low
2nd Trimester
Infancy
Adulthood
1st Tri.
3rd Tri.
Childhood
D
ExDp1
ExDp2
ExM
ExM−U
OPC
PgG2M
PgS
oRG
vRG
ExN
IP
InCGE
InMGE
End
Mic
Per
high
low
E
INTERFERON GAMMA RESPONSE
PI3K AKT MTOR SIGNALING
HEDGEHOG SIGNALING
PEROXISOME
APICAL JUNCTION
KRAS SIGNALING DN
TGF BETA SIGNALING
ALLOGRAFT REJECTION
ESTROGEN RESPONSE EARLY
WNT BETA CATENIN SIGNALING
NOTCH SIGNALING
KRAS SIGNALING UP
MYOGENESIS
G2M CHECKPOINT
MITOTIC SPINDLE
E2F TARGETS
UV RESPONSE DN
PROTEIN SECRETION
MYC TARGETS V1
PANCREAS BETA CELLS
adj. p−value
< 0.0001
< 0.001
< 0.01
< 0.05
N.S.
F
syndromic
CNV
SNV / indel
mosaic CNV
mosaic SNV / indel
inherited recessive
GWAS SNPs
adj. p−value
< 0.0001
< 0.001
< 0.01
< 0.05
N.S.

Fig. 1

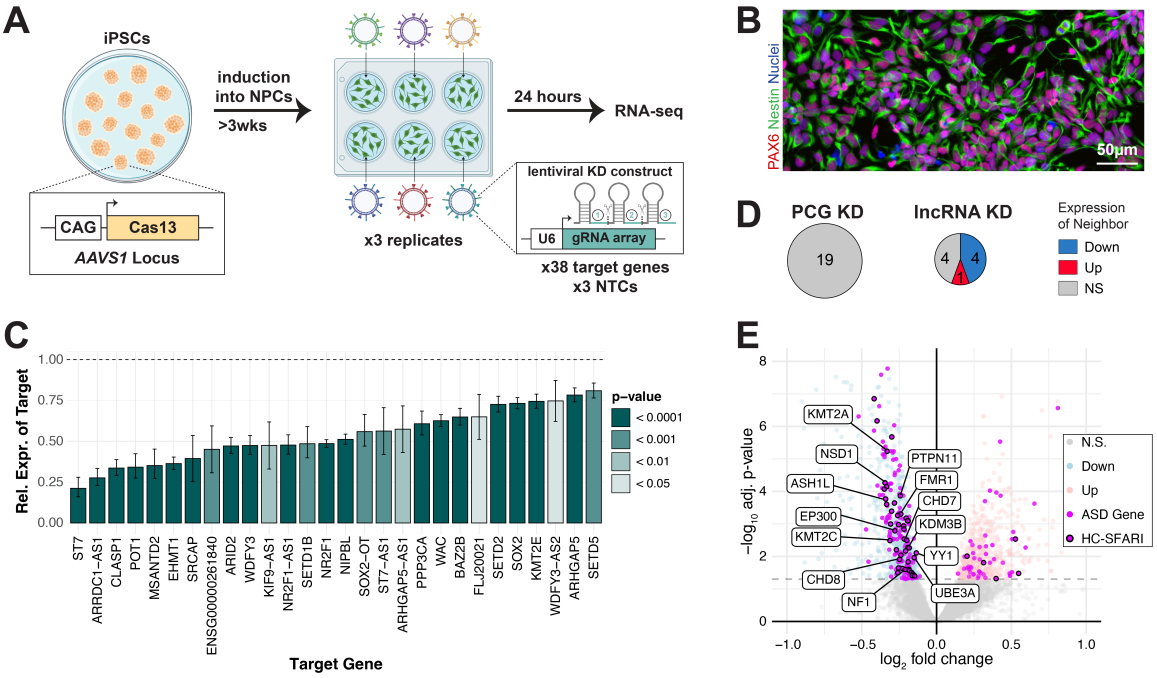
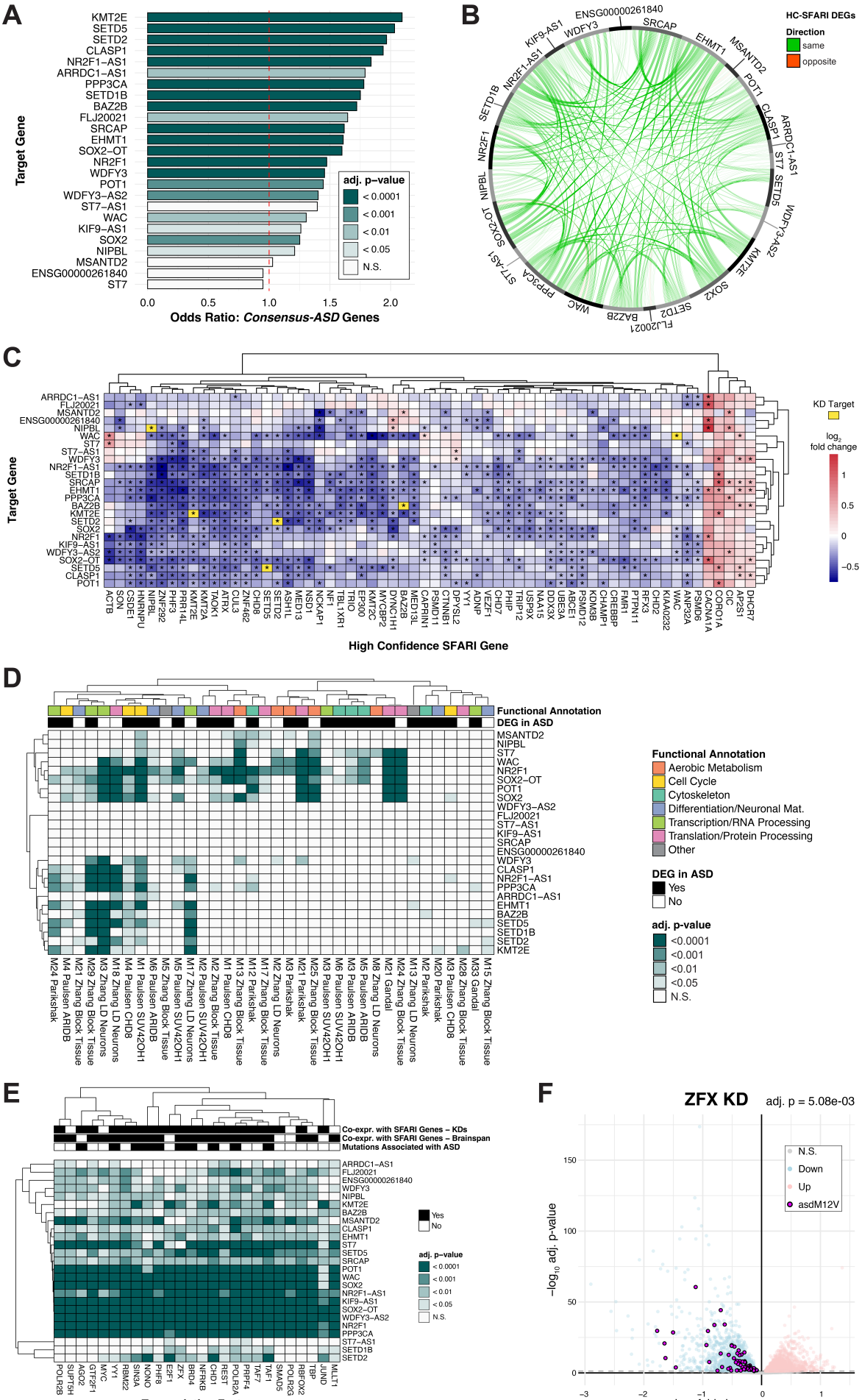

A
iPSCs
induction into NPCs
>3wks
24 hours
RNA-seq
CAG
Cas13
AAVS1 Locus
x3 replicates
lentiviral KD construct
U6
gRNA array
x38 target genes
x3 NTCs
B
PAX6 Nestin Nuclei
50µm
D
PCG KD
19
lncRNA KD
4
4
1
Expression of Neighbor
Down
Up
NS
C
Rel. Expr. of Target
1.00
0.75
0.50
0.25
0.00
p–value
< 0.0001
< 0.001
< 0.01
< 0.05
ST7
ARRDC1–AS1
CLASP1
POT1
MSANTD2
EHMT1
SRCAP
ENSG00000261840
ARID2
WDFY3
KIF9–AS1
NR2F1–AS1
SETD1B
NR2F1
NIPBL
SOX2–OT
ST7–AS1
ARHGAP5–AS1
PPP3CA
WAC
BAZ2B
FLJ20021
SETD2
SOX2
KMT2E
WDFY3–AS2
ARHGAP5
SETD5
Target Gene
E
–log10 adj. p-value
log2 fold change
KMT2A
NSD1
ASH1L
EP300
KMT2C
CHD8
NF1
PTPN11
FMR1
CHD7
KDM3B
YY1
UBE3A
N.S.
Down
Up
ASD Gene
HC-SFARI

Fig. 2

Fig. 3


A
Target Gene
KMT2E
SETD5
SETD2
CLASP1
NR2F1−AS1
ARRDC1−AS1
PPP3CA
SETD1B
BAZ2B
FLJ20021
SRCAP
EHMT1
SOX2−OT
NR2F1
WDFY3
POT1
WDFY3−AS2
ST7−AS1
WAC
KIF9−AS1
SOX2
NIPBL
MSANTD2
ENSG00000261840
ST7
adj. p−value
< 0.0001
< 0.001
< 0.01
< 0.05
N.S.
0.0
0.5
1.0
1.5
2.0
Odds Ratio: Consensus-ASD Genes
B
HC-SFARI DEGs
Direction
same
opposite
C
Target Gene
High Confidence SFARI Gene
KD Target
log2 fold change
D
Functional Annotation
DEG in ASD
Functional Annotation
Aerobic Metabolism
Cell Cycle
Cytoskeleton
Differentiation/Neuronal Mat.
Transcription/RNA Processing
Translation/Protein Processing
Other
DEG in ASD
Yes
No
adj. p-value
<0.0001
<0.001
<0.01
<0.05
N.S.
E
Co-expr. with SFARI Genes – KDs
Co-expr. with SFARI Genes – Brainspan
Mutations Associated with ASD
Yes
No
adj. p−value
< 0.0001
< 0.001
< 0.01
< 0.05
N.S.
Transcription Factor
F
ZFX KD
adj. p = 5.08e-03
N.S.
Down
Up
asdM12V
−log10 adj. p-value
log2 fold change

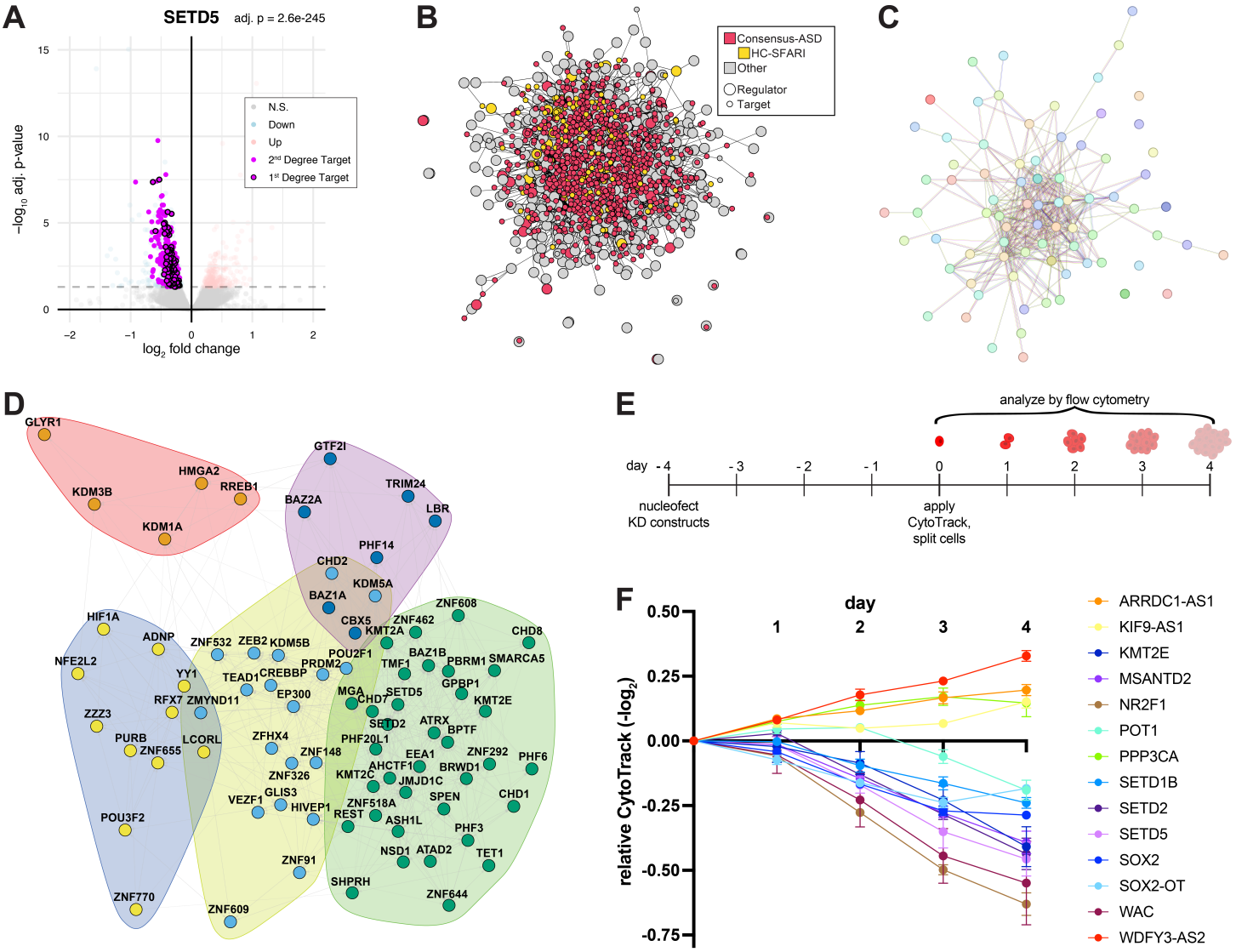

A
SETD5
adj. p = 2.6e-245
N.S.
Down
Up
2nd Degree Target
1st Degree Target
−log10 adj. p-value
log2 fold change
B
Consensus-ASD
HC-SFARI
Other
Regulator
Target
C
D
GLYR1
HMGA2
KDM3B
RREB1
KDM1A
GTF2I
TRIM24
BAZ2A
LBR
PHF14
CHD2
KDM5A
BAZ1A
ZNF608
HIF1A
ADNP
ZNF532
ZEB2
KDM5B
CBX5
ZNF462
KMT2A
CHD8
NFE2L2
YY1
TEAD1
CREBBP
PRDM2
POU2F1
BAZ1B
TMF1
PBRM1
SMARCA5
GPBP1
RFX7
ZMYND11
EP300
MGA
CHD7
SETD5
KMT2E
ZZZ3
SETD2
ATRX
BPTF
PURB
LCORL
ZFHX4
PHF20L1
ZNF655
ZNF148
EEA1
ZNF292
PHF6
AHCTF1
BRWD1
ZNF326
KMT2C
JMJD1C
GLIS3
VEZF1
HIVEP1
ZNF518A
SPEN
CHD1
REST
ASH1L
POU3F2
PHF3
NSD1
ATAD2
TET1
ZNF91
SHPRH
ZNF644
ZNF770
ZNF609
E
analyze by flow cytometry
day
-4
-3
-2
-1
0
1
2
3
4
nucleofect
KD constructs
apply
CytoTrack,
split cells
F
day
1
2
3
4
relative CytoTrack (-log2)
0.50
0.25
0.00
-0.25
-0.50
-0.75
ARRDC1-AS1
KIF9-AS1
KMT2E
MSANTD2
NR2F1
POT1
PPP3CA
SETD1B
SETD2
SETD5
SOX2
SOX2-OT
WAC
WDFY3-AS2

Fig. 4