Top2a promotes the development of social behavior via PRC2 and H3K27me3

Yijie Geng[1], Tejia Zhang[1], Sean C. Godar[1], Brock R. Pluimer[1], Devin L. Harrison[2], Anjali K. Nath[3], Jing-Ruey Joanna Yeh[4], Iain A. Drummond[5], Marco Bortolato[1], Randall T. Peterson[*1]

[1]Department of Pharmacology and Toxicology, College of Pharmacy, University of Utah, Salt Lake City, UT 84112, USA

[2]The Graduate Program in Biophysical Sciences, The University of Chicago, Chicago, IL 60637, USA

[3]Division of Cardiovascular Medicine, Beth Israel Deaconess Medical Center, Boston, MA 02115, USA

[4]Cardiovascular Research Center, Department of Medicine, Massachusetts General Hospital, Harvard Medical School, Charlestown, MA 02129, USA

[5]Davis Center for Aging and Regeneration, MDI Biological Laboratory, Bar Harbor, ME 04609, USA

*Corresponding author

SUMMARY

Human infants exhibit innate social behaviors at birth, yet little is understood about the embryonic development of sociality. We screened 1120 known drugs and found that embryonic inhibition of topoisomerase IIα (Top2a) resulted in lasting social deficits in zebrafish. In mice, prenatal Top2 inhibition caused behavioral defects related to core symptoms of autism, including impairments in social interaction and communication. Mutation of Top2a in zebrafish caused downregulation of a set of genes highly enriched for genes associated with autism in humans. Both the Top2a-regulated and autism-associated gene sets possess binding sites for polycomb repressive complex 2 (PRC2), a regulatory complex responsible for H3K27 trimethylation. Moreover, both gene sets are highly enriched for H3K27me3. Inhibition of PRC2 component Ezh2 rescued social deficits caused by Top2 inhibition. Therefore, Top2a is a key component of an evolutionarily conserved pathway that promotes the development of social behavior through PRC2 and H3K27me3.

INTRODUCTION

Sociality is broadly conserved across the animal kingdom, facilitating cooperation, reproduction, and protection from predation. Human infants exhibit an innate social preference behavior at birth: infants less than an hour old fix their gaze on human face-like images longer than other images[1]. Throughout life, this simple innate social drive serves as a foundation for the development of complex and versatile adult social behaviors. On the other hand, social dysfunction is a hallmark of several neurodevelopmental disorders[2] such as autism spectrum disorder. Deficits in early social behavior are linked to later diagnoses of autism[3]. The onset of autism can be traced back to the prenatal stage[4-6], which coincides with the development of innate social behavior.

Despite its importance, little is known about the embryonic development of sociality. While many genes conferring a small amount of autism risk have been identified by genome-wide association studies (GWAS)[7-10], it is much less clear how they each contribute to autism etiology. Apart from these genetic factors, environmental factors are estimated to account for ~40% of autism risk[11,12], yet few have been identified. In this study, we investigated the development of sociality in zebrafish (*Danio rerio*) using a chemical genetics approach. The zebrafish has become an important model organism for behavioral studies due to its behavioral and neurophysiological similarities to humans[13-17]. It possesses an evolutionarily conserved social brain network[15], enabling us to acquire insights into social behavior development of potential relevance for mammals[18-21]. Like in humans, a robust social preference behavior emerges early in zebrafish[22-25]. To restrict this investigation to developmental events occurring during the embryonic stage, we conducted a chemical screen that allows us to precisely apply and withdraw chemicals to and from zebrafish embryos during embryonic development. We screened an annotated library containing chemicals with known biological targets to help rapidly identify downstream biological targets.

Using an automated high-throughput behavioral analysis system which we named Fishbook, we screened 1120 known drugs and found that embryonic inhibition of topoisomerase II α (Top2a) resulted in social deficits in juvenile zebrafish. In mice, prenatal Top2 inhibition caused postnatal behavioral defects that are specifically related to the two core symptoms of autism: social communication impairments and increased perseveration. In fact, several known environmental risk factors of autism are Top2 inhibitors[26-29], indicating that Top2 inhibition may be a common mechanism through which environmental insults contribute to autism risk.

To elucidate the Top2-mediated mechanism of action, we performed RNA-seq of Top2a mutant zebrafish and observed downregulation of a set of genes highly enriched for autism risk genes. Using a custom upstream analysis pipeline, we found that both the Top2a-dependent gene set and a set of autism-associated genes maintained by the Simons Foundation for Autism Research (SFARI genes hereafter) are selectively targeted by polycomb repressive complex 2 (PRC2), a regulatory complex responsible for depositing H3K27 trimethylation (H3K27me3). In corroboration of this finding, we found both gene sets to be highly enriched for H3K27me3. Strikingly, chemical inhibition of the PRC2 component Ezh2 rescued social deficits caused by Top2 inhibition in zebrafish, indicating that Top2a likely functions by antagonizing PRC2/H3K27me3-mediated gene silencing. These findings identify Top2a as a key component of an evolutionarily conserved pathway that promotes the development of social behavior through PRC2 and H3K27me3.

RESULTS

Fishbook: a scalable social behavior assay system

Zebrafish develop a robust social preference for age-matched conspecifics at 3 weeks of age[22]. To quantitatively assess this behavior, we developed a scalable and fully automated assay system named Fishbook. Briefly, a test arena (Figure 1A) was manufactured by 3D printing (Figure 1B). Each arena (Figure 1C, red rectangle) is a long rectangular lane printed using nontransparent material, divided into 3 parts by two transparent windows (4): the longer middle part serves as the test compartment (2), one of the two smaller end-compartments contains a live fish as social stimulus (1), and the other end-compartment is left empty (3) (Supplemental Video). To maximize throughput, we grouped 44 arenas together (Figure 1D), and constructed a unique telecentric lens-based imaging system which enables us to simultaneously image all 44 arenas while avoiding optical obstruction by the non-transparent arena walls (Figure 1E). The Fishbook system is capable of screening >1000 fish per day.

A social score with values ranging between -1 and 1 was defined to quantitatively measure social behavior (Figure 1F). Wild-type (WT) fish typically spend a significant amount of time near the social stimulus fish, generating high social scores; without a social stimulus, test subjects swim randomly, giving rise to an average social score close to 0 (Figure 1G-1I). Consistent with a previous report[22], this preference is lost in the dark (Figure S1A), indicating that it is driven primarily by visual cues. Social score is also drastically reduced when social stimulus fish is replaced by a pebble of similar size (Figure S1B), suggesting that this preference behavior is highly specific to social cues and cannot be explained solely by curiosity toward a novel object. To validate its power in recognizing social deficits, we examined fish possessing loss-of-function mutations of genes linked to autism and schizophrenia, including *Nrg1*[27] (Figures 1K & S1C) and *Disc1*[28] (Figures 1L & S1D). Social scores are significantly reduced in both mutants, demonstrating that the Fishbook assay can effectively detect social defect caused by a single genetic mutation.

A systematic screen identifies fluoroquinolones as inhibitors of social development

Using Fishbook, we screened the Prestwick library which contains 1120 compounds with known biological targets. Briefly, embryos were exposed to chemicals during the first 3 days of development (0 - 3 days post fertilization, or dpf), raised to 3 weeks of age, and tested in the Fishbook assay (Figure 2A). Hits were defined as having an average social score < 0.1. Due to the relatively large variation expected for a behavior-based assay, to reduce the number of false positive hits, primary screening was performed in duplicates and hits were confirmed by repeat testing. Four confirmed hits were found, including flumequine, lomefloxacin, ofloxacin, and oxolinic acid (Figures 2B & S2A). Importantly, these four compounds were the only validated hits from the screen, and all belonged to the same class of antibiotics named fluoroquinolones. A previous publication reported modulation of zebrafish shoaling behavior by long term exposure to a mixture of 3 fluoroquinolones (ofloxacin, ciprofloxacin, and enrofloxacin) and 3 tetracyclines, suggesting a potential link between fluoroquinolones and zebrafish social behavior which is in line with our observation, but did not examine the effect of fluoroquinolones alone without tetracyclines or provide insights on their molecular mechanisms[30].

Top2 inhibition is responsible for the social deficit phenotype

Although fluoroquinolones are best known as antibiotics targeting the prokaryotic type-II topoisomerases, they also inhibit the eukaryotic topoisomerase II (Top2), albeit at higher doses. For example, the IC_{50} of ofloxacin and lomefloxacin against HeLa cell Top2 are 365 μM and 613 μM[31], and their effective doses in the Fishbook assay are 200 μM and 100 μM, respectively. Prior studies have also suggested a link

between Top2 and autism[32-36] (see Discussion for details). We thus hypothesized that fluoroquinolones inhibit the development of sociality by targeting zebrafish Top2.

To test this hypothesis, we examined several structurally diverse eukaryotic Top2 inhibitors (Figure S2B). They target different steps of the Top2 catalytic cycle[37,38] (Figure 2C) and can be grouped into two categories based on their mechanisms of action, including Top2 poisons (ellipticine, amsacrine, etoposide, and mitoxantrone), and Top2 catalytic inhibitors (aclarubicin, ICRF-193, and sobuzoxane). Through embryonic exposure (Figure 2D), all 7 structurally diverse Top2 poisons and inhibitors induced social deficits (Figures 2E & 2F). We also tested several environmental chemicals which were known to inhibit eukaryotic Top2 activity, including chlorpyrifos, chlorpyrifos oxon, and genistein (see Discussion for more information), and observed similar social deficits (Figure S2C). These results demonstrated that inhibition of zebrafish Top2 can indeed induce social deficits.

Top2 inhibition does not function by inhibiting transcription or genomic stability

Apart from its role in facilitating replication, Top2 plays important parts in promoting transcription and genomic stability[39]. We therefore investigated whether the effect of Top2 inhibition can be attributed to deficits in one of these two apparent mechanisms. We first examined two RNA polymerase II inhibitors, actinomycin D and triptolide, but did not observe social deficits (Figures S2D & 2E). We then tested two DNA double strand break (DSB) inducers, bleomycin and hydroxyurea, and again found that neither compound induced social deficits (Figure S2F). This result is also supported by the fact that while Top2 poisons induce DSBs, catalytic inhibitors do not, yet both classes of Top2 inhibitors cause social deficits (Figures 2E & 2F). Together, these data support the possibility that Top2 inhibition functions through a mechanism other than general transcription inhibition or DNA instability.

Top2a, but not Top2b, is required for the development of social behavior

We next investigated which isoform, Top2a or Top2b, was responsible for the social deficit phenotype. We first applied two splice-blocking morpholino oligonucleotides (MO) (Figure S3A). Compared to permanent genomic editing methods such as CRISPR, MO's transient-acting nature better mimics effects of the Top2 inhibitors. Note that maternal Top2a mRNAs are present during the first 24 hours of embryogenesis[40] and will escape inhibition by Top2a-MO. Following injection, Top2a-MO induced social deficits, which were rescued by co-injecting *hTOP2A* mRNA (Figures 3A & S3B). Top2b-MO, however, failed to induce social deficits (Figure 3B). A Top2a-selective inhibitor sodium salicylate also induced social deficits (Figure 3C) when applied at a dose known to confer its isoform selectivity[38].

We then examined two loss-of-function (LoF) mutants of Top2a (*can4*[41]) and Top2b (*noto*[42]). Because homozygous mutants are lethal, we established breeding colonies using heterozygous (het) mutants. Incrossing het mutants generated mixed populations composed of ~ 1/3 WT and 2/3 het offspring. They were assayed blind in Fishbook, i.e., with no prior knowledge of each fish's genetic background, and genotyped individually. Het Top2a mutant (*can4*$^{+/-}$) exhibited reduced social score compared to WT sibling (Figure 3D), while het Top2b mutant (*noto*$^{+/-}$) did not show reduction in social score (Figure 3E).

Finally, we tested whether transient overexpression of Top2a can rescue the Top2 inhibitor-induced social deficit. A potential concern is that overexpressed WT hTOP2A protein may simply bind to and "soak up" a Top2 inhibitor, leaving less free compound to interact with its real target, resulting in false positive rescue. We therefore generated a hTOP2A mutant (Y165S) which is functionally active but resistant to ICRF-193[43]. Transient overexpression of hTOP2AY165S by mRNA injection successfully rescued ICRF-193 (Figures 4L & S3C), demonstrating that Top2a alone can rescue Top2 inhibition. Together, these results suggest that Top2a is the isoform responsible for the social deficit phenotype.

Prenatal inhibition of Top2 induces social deficits and other autism-related behavioral defects in mice

To determine whether this Top2-mediated mechanism is conserved in mammals, we assessed the effect of prenatal Top2 inhibition in mice. Pregnant mice received daily intraperitoneal (i.p.) injections of ICRF-193 or vehicle control from 7 to 18 days post coitum (dpc). Offspring grew to adult stage without observable differences in gross morphology or body weight (Figures S3D) compared to vehicle control or no-treatment groups. Behavioral assays were conducted 2 – 10 months after birth (Figure 4A).

We first conducted a reciprocal social interaction assay[44], in which 2 sex and age matched mice, including a test subject and a social stimulus mouse, were placed into a novel cage and video recorded. The number and duration of interactions between a test subject and the control mouse were manually scored by a blinded reviewer. ICRF-193-treated mice showed significantly reduced total duration, total number, and mean duration (Figures 4B-4D) of social investigations compared to the vehicle control. We then performed a novel object exploration assay to determine if the ICRF-193 mice lack interest in or even avoid novel stimuli in general. ICRF-193 mice did not show significantly altered object exploration compared to controls, as quantified by the total duration, total number, and mean duration (Figures 4E-4G) of their interactions with the novel object (a wooden cube). We also observed no significant changes in motion patterns in an open field assay, as measured by the time spent in the center of the test arena and average velocity (Figures 4H-4I), indicating intact motor capabilities and normal stress responses for the ICRF-193 mice. These results indicate that social behaviors are selectively targeted by prenatal Top2 inhibition, while exploratory behaviors, motor activity, and stress response remain intact.

We then investigated other autism-related behavioral deficits including impaired communication and perseveration. We examined ultrasonic vocalization (USV) of the pups for potential social communication deficits. When briefly separated from their mothers, ICRF-193-treated pups exhibited significantly reduced numbers of USVs compared to controls (Figure 4J), indicating a significant impairment in their communication. The ICRF-193-treated pups also showed a significantly elevated mean peak frequency in their vocalizations (Figure 4K), a phenotype that was also found in established mouse models of autism[45,46]. We then examined restrictive and repetitive behaviors using T-maze[44] and hole-board[44] assays. ICRF-193 mice repeatedly explored the same arm of a T-maze, as shown by their significantly reduced number of alternative explorations and delayed onset of first alteration (Figures 4L-4M) compared to the controls. Hole-board assay also revealed significantly elevated coefficient of exploratory variation, a measure of repetitive tendencies[44], in ICRF-193 mice (Figure 4N). These results demonstrate that prenatal inhibition of Top2 induces behavioral deficits related to the core symptoms of autism in mice.

Top2a depletion selectively downregulates autism risk genes

Based on the observed link between Top2a inhibition and autism-related behavioral deficits, we investigated whether Top2a depletion can selectively modulate expression of autism risk genes using RNA-seq of 3 dpf *can4*$^{-/-}$ (Top2a) mutants (Figure 5A). Due to the presence of maternal Top2a mRNA during the first 24 hours of embryogenesis[40], *can4*$^{-/-}$ embryos are expected to be partially depleted of Top2a during the first few days of embryonic development. We found 8634 significantly (adjusted *p*-value < 0.05) upregulated genes and 8589 significantly downregulated genes in *can4*$^{-/-}$ mutants compared to WT controls. Top2a was downregulated as expected (adjusted *p*-value = 2.16E-73). We used the human orthologs of zebrafish genes for subsequent analyses, which include 5044 human orthologs for the downregulated genes (hereafter referred to as *can4*$^{-/-}$ downregulated genes, or can4Dn) and 5054 human orthologs for the upregulated genes (*can4*$^{-/-}$ upregulated genes, or can4Up) (Table S1).

To identify diseases that most closely associate with genes affected by Top2a depletion, we analyzed two disease-gene libraries DisGeNET[47] and GLAD4U[48] using over-representation analysis (ORA) to search for disease risk gene sets that are enriched (over-represented) in the can4Dn and can4Up gene lists. can4Dn is enriched for genes associated with autism and its comorbid conditions including schizophrenia, bipolar disorder, depression, and intellectual disability; these diseases consistently rank at the top of the list among the 24166 diseases in the DisGeNET library and the 3071 in GLAD4U, demonstrating a high degree of specificity for these diseases (Figures 5B & S4A). can4Up is not enriched for most of these diseases except modestly for intellectual disability in DisGeNET. We also found a similar pattern of enrichment in independently curated disease gene sets related to autism and its comorbidities (SFARI[49,50], AutismKB 2.0[51], DisGeNET[47], GLAD4U[48], BDgene[52], PsyGeNET[53], SZGene[54], SysID[55], and GEPAD[56]) (Figure S4B). Using gene set enrichment analysis (GSEA) as an orthogonal approach, we again found significant enrichment (5% FDR) of autism and its comorbidities risk genes only in the downregulated genes (Figures 5C, 5D, S4C, & S5).

Pathway (KEGG and REACTOME) and gene ontology (GO) ORA analyses also revealed enrichment of genes related to pathways that are believed to contribute to autism etiology[57-60], including axon guidance, neuronal development, glutamatergic signaling, and synaptic transmission, in can4Dn but not can4Up (Figure S6A). GSEA analysis using these libraries showed similar enrichments only in the downregulated genes (Figure S6B).

To eliminate any potential bias in our enrichment analyses caused by naturally-occurring enrichment or depletion of autism-associated genes in zebrafish, we examined the specificity of our analyses using a permutation assay. Because can4Dn and can4Up both contain close to 5000 genes, we randomly selected 1000 lists of 5000 genes out of the 14989 human genes with zebrafish orthologs and repeated the ORA analyses for these gene sets to create a null distribution. We found that can4Dn but not can4Up consistently gives significantly higher odds ratios compared to the null distribution, thus validating the specificity of the observed enrichment of autism risk genes (Figures S7-S9).

can4Dn and autism risk genes share similar upstream regulators

We analyzed a published human TOP2A ChIP-seq dataset to investigate whether TOP2A selectively binds to autism risk genes[61]. We did not find enrichment of autism risk genes in TOP2A target genes (Figure 7B, first column), consistent with the traditional view of Top2a's role as a "housekeeping" gene with no known selectivity for autism risk genes. Given that Top2a binding is not enriched in autism risk genes but Top2a-dependent regulation is, we hypothesized that some specificity factor may associate with Top2a, enabling its influence to be heightened in autism risk genes. To identify such a specificity factor, we conducted ORA analyses to identify upstream regulators that are significantly enriched in the promoter regions of can4Dn genes but not can4Up genes. We also analyzed SFARI genes in parallel as a representative set of autism risk genes. Strikingly, by analyzing a curated library based on published ChIP-seq data (ChIP-X[62,63]), we found that the SFARI genes and can4Dn are targeted by a highly similar set of upstream regulators, especially the transcription factor SUZ12 and its associating partner EZH2 (Figure 6A).

We assessed the specificity of this similarity by running a blast-style search (UES-blast) against control gene sets from a reference dataset consist of DisGeNET, GLAD4U, and an RNA-seq gene set library RNAseqGEO[47,48,63] to identify gene sets with the lowest blast scores (highest similarity rankings) to a query gene set (Figure 6B). Using SFARI genes as a query gene set for UES-blast, we found that can4Dn, together with several independently curated autism risk gene sets, rank among the highest, while can4Up ranks much lower. Removing overlapping SFARI genes from can4Dn (can4Dn-SFARI) does not significantly affect its ranking, indicating that the overlapping SFARI genes were not the sole contributors to their rankings. Autism comorbid conditions also rank high on the list, except for intellectual disability

(Figures 6C & S10A). We also acquired several neurodegenerative disorder risk gene sets including Alzheimer's disease, Parkinson's disease, and Huntington's disease from DisGeNET as negative controls due to their neurological nature, late onset, and lack of comorbidity with autism; these control gene sets rank relatively low on the list as expected. Using tSNE analysis we found that can4Dn, SFARI, and other autism risk gene sets tightly cluster together, compared to the control disease gene sets in DisGeNET (Figure 6D) or the entire reference dataset (Figure S11). These results suggest that the Top2a-dependent can4Dn and autism risk genes share a very similar and unique set of upstream regulators.

Among the strongly associated upstream regulators, we found 4 that are shared by can4Dn and SFARI genes, but not by can4Up, including SUZ12, EZH2, REST, and SMAD4 (Figure 6B). Intriguingly, SUZ12 and EZH2 have been reported to bind to TOP2A[64]. Because both SUZ12 and EZH2 are core components of the PRC2 complex, a histone methyltransferase, we explored enrichment of histone modifications in our gene sets using ORA analysis and two curated libraries containing published ChIP-seq data gathered from the ENCODE and Roadmap Epigenomics projects[63]. PRC2 methylates histone H3 on lysine 27 (i.e., H3K27me3). Remarkably, we found that can4Dn and SFARI genes, but not can4Up, are highly enriched in genes marked by H3K27me3 but not by the other 29 histone markers examined (Figure 6E). We calculated a H3K27me3 score to quantify this enrichment (Methods). can4Dn and autism risk gene sets give the highest H3K27me3 scores when compared to control gene sets from the reference dataset, while can4Up scores zero as expected (Figures 6F & S10B). Autism comorbid conditions except for intellectual disability also give high H3K27me3 scores.

PRC2 and H3K27me3 mediate Top2a-dependent gene regulation

We compared genome-wide binding patterns of TOP2A and PRC2 using published ChIP-seq data[61,65] in human K562 cells, and found that among the triple co-binding sites of SUZ12, EZH2, and H3K27me3, > 93% were also bound by TOP2A (Figure 7A). In addition, while TOP2A binding peaks alone are not enriched for autism risk genes (Figure 7B, first column), overlapping TOP2A binding peaks with EZH2, SUZ12, and H3K27me3 binding peaks, respectively, enriched their selective targeting to autism risk genes and can4Dn, but not to can4Up and the neurodegenerative disorders gene sets (Figure 7B). These results suggest that the selective regulation of autism risk genes by Top2a may be achieved through its interplay with PRC2 and H3K27me3.

While H3K27me3 functions as a gene silencing mark[66], Top2a binding is strongly associated with open chromatin[61,67]. A FAIRE-Seq analysis revealed that Top2a inhibition by ICRF-193 in mouse embryonic stem cells results in reductions in chromatin accessibility[67], demonstrating a role for Top2a in maintaining open chromatin. These results suggest an antagonistic relationship between Top2a and H3K27me3. We therefore reasoned that Top2a inhibition may serve to derepress PRC2 on genes contributing to social development, whereas this effect might be mitigated by a PRC2 inhibitor (Figure 7C).

To test this hypothesis, we added an Ezh2 inhibitor UNC1999 and its inactive control compound UNC2400[68] alongside Top2 inhibitors during embryonic exposure and found that UNC1999 effectively ameliorates social deficits caused by flumequine (a fluoroquinolone antibiotic), ellipticine (a Top2 poison), and sodium salicylate (a Top2a-selective Top2 catalytic inhibitor), while UNC2400 shows no effective rescue (Figure 7D). Perhaps more significantly, we found that overnight exposure to UNC1999 but not UNC2400 the night before Fishbook test effectively rescued social deficits in juvenile fish treated with Top2 inhibitors during embryonic development (Figure 7E). These findings suggest that a significant portion of the social deficit phenotype may be attributed to elevated H3K27me3 that was introduced by Top2a inhibition at the embryonic stage and persisted to the juvenile stage.

DISCUSSION

Association studies have now identified hundreds of genetic variants that each contribute a small amount of autism risk[7-10]. Given the large number of genes involved and the inability of any one gene to explain more than a small amount of increased risk, one might ask whether the many autism risk genes are regulated independently or are coordinately regulated in some way. Evidence of coordination might point toward shared mechanisms of autism etiology, as well as give hope that autism spectrum disorders of diverse genetic backgrounds might be treatable by targeting their shared mechanisms.

Although genetic factors are clearly important contributors to ~60% of autism risk, environmental factors appear to explain an additional ~40% of risk[11,12]. Despite the importance of identifying environmental factors contributing to autism, few have been identified. Interest in identifying environmental risk factors of autism is intense; even a fraudulent claim linking vaccinations and autism prompted lasting changes in vaccine utilization[69]. The difficulty of linking environmental factors with neurobehavioral development may be largely attributed to the massive number of environmental factors and the limited throughput of current research methodologies. Only a fraction of the chemicals in our environment have been examined for potential adverse effects on neurobehavioral development[70,71]. Current experimental methods mainly rely on rodent models. Although these models can often recapitulate complex behavioral phenotypes that relate to human diseases, the time, space, and financial resources required can be prohibitive for conducting large scale experimental screens. Models enabling large scale testing of the impact of small molecules on social development are needed.

The zebrafish possesses an evolutionarily conserved social brain network[15]. Many components of this network have been studied in rodent models for social-related behaviors, such as social reward[72], aggression[73,74], mating[75,76], social learning[77], and social memory[78], which paved the way for using zebrafish to investigate social behavior and autism. The Fishbook platform enables systematic evaluation of environmental chemicals to identify those affecting social behavior development. In a systematic small molecule screen, we identified Top2a inhibitors as profound disruptors of social development in zebrafish and mice. Importantly, disruption of Top2a preferentially alters the expression of many of the genes previously associated with autism in humans. Hence, Top2a may serve as a common regulatory factor controlling a large set of genes known to confer autism risk, and as such may also serve as a link between genetic and environmental contributors to autism.

Prior studies have implicated Top2a in autism via different approaches. One study identified *TOP2A* as a member of a core autism risk gene network consisting of 234 genes by analyzing whole exome sequencing data[34]. Another revealed *TOP2A* as one of the 23 hub-genes related to brain maldevelopment in autism toddlers[35]. The Top2 inhibitor ICRF-193 has been shown to inhibit transcription of extremely long genes that are associated with autism in cultured mouse cortical neurons, embryonic stem (ES) cells and ES-cell-derived neurons[33]. In corroboration of both this finding and ours, a FAIRE-Seq analysis found that ICRF-193 caused reduction in chromatin accessibility in ES cells, especially in long silent genes[67]. Interestingly, several Top2 inhibitors were found to reactivate the dormant allele of the autism and Angelman syndrome gene *Ube3a* in primary cortical neurons from mice by reducing transcription of an imprinted antisense RNA *Ube3a-ATS*[32].

Fluoroquinolones are generally avoided during pregnancy due to their association with developmental toxicity in animals. Despite this consensus, fluoroquinolones are occasionally prescribed to pregnant woman[79,80], likely due to providers being unaware of the pregnancy or unaware of the potential risk. The results presented here may call for additional investigation into whether prenatal fluoroquinolone exposure in humans increases autism risk. It is noteworthy that a number of known or suspected environmental risk factors of autism are Top2 inhibitors, such as Bisphenol A (BPA)[26,81,82], polychlorinated biphenyls (PCB)[27,28,83-85], chlorpyrifos and chlorpyrifos oxon[29,86-88], and genistein[89-91]. Top2 inhibitors are widely present in the environment, and their possible link to autism has previously been proposed[92]. Chemicals with Top2 inhibiting activities that are present in the environment but have

not yet been clearly associated with autism include flavonoids[90,93], alternariol[94], Ginkgo biloba leaf extract[95], and lignans[96-98], to name a few. Environmental pollution by drugs with Top2 inhibitor activities has also been reported[99-103]. Nevertheless, it is not known whether or not Top2a inhibition is a meaningful contributor to autism prevalence in humans. Further epidemiological studies will be needed.

It is perhaps surprising that Top2a, an enzyme often regarded as performing 'housekeeping' functions, might have a specific effect on development of sociality. In fact, as described here, Top2a binds DNA throughout the genome, with no detected preference for genes associated with autism. The situation changes, however, when considering Top2a colocalized with the PRC2. Colocalization of Top2a, PRC2, and the PRC2-deposited epigenetic mark H3K27me3 is highly enriched among autism risk genes, including those discovered through human GWAS studies and those identified in zebrafish in this study. Hence, Top2a appears to influence the expression of diverse genes that together regulate social development by antagonizing PRC2/H3K27me3 mediated gene silencing. In alignment with this hypothesis, mutations in both H3K27 demethylases, KDM6A and KDM6B, have been associated with autism[104,105], linking defects in a H3K27me3 antagonizing mechanism with autism etiology. Why so many autism risk genes would be co-regulated by the Top2a/PRC2 mechanism is unclear. One possibility is that many genes are required for successful social development, and that being able to simultaneously regulate their expression during development is advantageous.

The effects of Top2a inhibition on sociality appear to be durable: zebrafish and mice treated transiently with Top2a inhibitors during development exhibit social deficits weeks later in the case of zebrafish and months later in the case of mice. This durability may reflect a permanent developmental defect, for example if Top2a inhibition during development alters the structure of the brain in such a way that normal sociality cannot be recovered even after Top2a activity is restored. Alternatively, the social deficits may reflect gene expression changes that are preserved by epigenetic marks but might theoretically be reversed by a second epigenetic intervention. Remarkably, we found that manipulation of the Top2a/PRC2 pathway using the Ezh2 inhibitor UNC1999 could rescue social deficits caused by Top2 inhibition, not only when co-treated during development but even in older animals. In particular, overnight exposure to UNC1999 rescued social deficits caused by embryonic Top2 inhibition, suggesting that a short-term modulation of this pathway at the juvenile stage may be sufficient to boost pro-social gene expression. This phenomenon, if conserved in mammals, may offer hope that Top2a-mediated social deficits might be reversible through epigenetic reprogramming.

In summary, this study reveals a crucial role for Top2a in promoting the development of social behavior in zebrafish and mice. Our results suggest that Top2a is indispensable for maintaining an upstream regulatory network that selectively controls the expression and epigenetic status of a large subset of autism risk genes. Top2a likely functions by antagonizing PRC2 and H3K27me3 mediated gene silencing. Further analyses are warranted to examine potential links between prenatal exposures to Top2-inhibiting chemicals and autism risk in humans through epidemiological and toxicological approaches.

METHODS

The Fishbook system and Social Score

The basic unit (test arena) of this system is a 3D printed, 10 mm deep, 8.5 mm wide, and 80 mm long rectangular chamber. Each chamber is divided into three compartments by two transparent acrylic windows (1.5 mm thick): a 60 mm long middle testing chamber to place the test subject, and two 8.5 mm long end chambers to place the social stimulus fish or remain empty, respectively. The walls are 1.5 mm thick. The 3D printing model was created using Tinkercad (Autodesk). These arenas were 3D printed using white PLA (polylactic acid) at 100 % infill. Printed arenas were glued onto 3/16" thick white translucent (43% light transmission) acrylic sheets (US Plastic) using a silicone sealer (Marineland). Transparent acrylic windows were precision cut to 8.5 mm × 10 mm pieces using a laser cutter and inserted into printed slots in the arena and fastened using silicone sealer.

The key component of our imaging system is a 322 mm diameter bi-telecentric lens (Opto Engineering) with an IR (850 nm) bandpass filter (Opto Engineering). Videos were taken by a Blackfly 5.0 MP Mono USB3 Vision camera (PointGrey) at 7.5 frames per second (fps). Ambien light was provided using white LEDs (EnvironmentalLights). The arenas were illuminated from below with infrared (850 nm) LEDs (EnvironmentalLights). Infrared LEDs were cooled by a heat sink (H S Marston). Structural supports and enclosure were custom built using parts purchased from Thorlabs, McMaster Carr, and US Plastic.

Test subjects were individually placed inside each test chamber using a plastic transfer pipette with its tip cut off to widen the opening. Their visual access to social stimulus fish were temporarily blocked by a 3D printed white comb-like structure placed in front of the social stimulus compartment (Figures S1E & S1F). Once all test subjects were placed into the test arena array, the array was placed inside the imaging station and the combs were removed to visually expose the social stimulus fish to the test subjects. After a 5 min acclimation period, a 7.5 min (for screening and validation) or 10 min test session was video recorded.

Videos were streamed through the software Bonsai[106]. Videos were analyzed in real-time during recording, and the frame-by-frame x and y coordinates of each fish relative to its own test compartment were exported as a CSV file. Data were analyzed using custom scripts (Python) to calculate social scores and generate tracking plots. Social score was defined as a fish's average Y-axis position for all frames. We designated the middle of each test chamber as the origin of Y-axis. We then assigned a value of 1 to the end of the chamber close to the social stimulus fish, and a value of -1 to the other end of the chamber which is close to the empty control compartment. Therefore, in this coordinate system, all social scores have values between -1 and 1. A higher social score demonstrates a shorter average distance between a test fish and a social stimulus fish during a test, which suggests a stronger social preference.

Zebrafish embryonic chemical treatment

For high-throughput screening, fertilized eggs were transferred into 96-well square-well plates at 12 eggs per well. Each well was filled with 500 μl volume E3 medium. E3 medium was supplemented with penicillin/streptomycin to prevent bacterial contamination. Compounds in the Prestwick chemical library (all dissolved in DMSO) were screened at 1:500 dilution for a final concentration of 20 μM. Negative controls were treated with an equal volume of DMSO. Each compound was added to 2 wells. All or 50% of medium and drugs were changed at 1- and 2-day post fertilization (dpf), respectively. Larvae typically hatch between 2-3 dpf. At 3 dpf, all hatched and live but unhatched larvae treated by the same compound were collected and transferred to a petri dish containing fresh E3 medium without penicillin/streptomycin. At 5-7 dpf, larvae from each petri dish were transferred to a separate nursery tank to be raised to 3 weeks of age for Fishbook assay.

For individual compound treatments, each chemical's general toxicity level was first estimated by

measuring overall embryonic lethality after 3 days of treatment (0-3 dpf). Most biologically active small molecules are toxic to zebrafish embryos when applied at high doses and toxicity is often a good benchmark assessment of effective compound exposure. Each compound was first applied to the embryos from 0-3 dpf by dissolving in E3 medium at a wide range of doses, often ranging from 0.1 µM to 1 mM with 10-fold serial dilutions. Two to three additional rounds of assessments were then conducted to narrow down the dose ranges and identify the LD_{50} of each compound. Once toxicity is assessed, compound treatments began at LD_{50} and simultaneously increased and decreased 2-3 times serially by two-fold (at $LD_{50}\times4$, $LD_{50}\times2$, $LD_{50}/2$, $LD_{50}/4$, etc.). We have also found that some compounds, when applied at a certain dose, showed little to no toxicity at the embryonic stage, but caused massive lethality at a later stage once these drug-treated larvae entered nursery. For these compounds, further adjustments of compound doses were required to identify their new LD_{50} values with nursery survival taken into consideration. Most inhibitors tested effectively induced social deficits when applied at LD_{50}. If a compound showed no effect in inducing social deficits at lower doses, at least one dose at > LD_{50} must be successfully tested before a conclusion could be made regarding the compound's effectiveness. All compounds were dissolved in DMSO (Sigma-Aldrich) or water. The final DMSO concentration was never higher than 1%.

The Peterson lab previously discovered that the logarithm value of a compound's octanol:water partition coefficient (logP) strongly correlates with the compound's embryonic permeability[107,108]. A logP value higher than 1 predicts good absorption, whereas compounds with a logP value lower than 1 are often poorly absorbed by the zebrafish embryo. This principle has been confirmed by the author's own observations, as in our hands, compound with logP values lower than 1 often failed to cause any detectable toxicity even when applied at extremely high doses in E3 medium. This problem can be overcome by injecting compounds directly into embryos at the 1- or 2-cell stage[107]. Therefore, compounds with logP values lower than 1, including ICRF-193, etoposide, bleomycin, and hydroxyurea, were directly injected into embryos. Their respective dosages were determined by first injecting a range of doses, often ranging from 0.1 µM to 10 mM with 10-fold dilutions in between, to assess their general toxicity levels as quantified by the percentage of embryonic lethality induced at each dosage. The doses were then narrowed down and refined to identify the LD_{50} of each compound. Finally, each drug was injected at its LD_{50} dose and at several doses higher and lower (with 2-fold differences) than LD_{50}.

High-throughput screening and validation

A total of 2 rounds of screening and 1 round of validation were performed. The first round exposed 24 embryos to each compound. Treatment groups with social scores lower than 0.1 were identified. These compounds were then hand-picked for the second round of screening, which exposed 48 embryos to each compound. Treatment groups that again inhibited social score to < 0.1 were identified. These compounds were purchased from commercial sources and used to treat 100 embryos. Compounds that significantly reduced social score compared to DMSO control were identified as hits.

Chemical library and other compounds

The Prestwick library (Prestwick Chemical) contains 1,120 approved drugs dissolved in DMSO at a stock concentration of 10 mM. Other compounds were purchased from the following sources. From Cayman Chemical: flumequine, oxolinic acid, etoposide, mitoxantrone (hydrochloride), ellipticine, bleomycin (sulfate), hydroxyurea, actinomycin D, triptolide, UNC1999. From Sigma-Aldrich: lomefloxacin hydrochloride, ofloxacin, sodium salicylate. From ApexBio: amsacrine. From Santa Cruz: ICRF-193, sobuzoxane, aclarubicin (aclacinomycin A). From Tocris: UNC2400. All individually purchased compounds were dissolved in DMSO or water. Chemical structures were generated using PubChem

Sketcher.

Zebrafish juvenile chemical treatment and Fishbook testing

For the juvenile rescue experiments, 20 dpf zebrafish were collected from nursery tanks. Fish that received the same treatment at the embryonic stage but raised in different tanks were pooled together and then sorted into deep petri dishes (such as a 25 mm deep petri dish) containing 40 ml E3 medium mixed with rotifers as fish feed and 5 µl AmQuel (Kordon) to remove harmful ammonia excreted by fish. 10-15 fish were sorted into each dish. Compounds were then added to each dish. Dishes were incubated at 28 °C in a larvae incubator overnight. On 21 dpf, each dish is fed lightly with rotifers in the morning. Fishbook assay was conducted ~1-2 hours after feeding. Immediately before plating fish in each petri dish into the Fishbook test arena, the content of the petri dish is poured through a nylon tea strainer so that all liquid passes through and fish are kept in the tea strainer. The tea strainer containing fish is then consecutively dipped into 3 petri dishes containing E3 to wash the residual chemical away from the fish. The fish is then poured into a petri dish containing clean E3 and transferred into the Fishbook test arena using a plastic transfer pipette for testing.

Site-directed mutagenesis

Construct containing full-length human *TOP2A* cDNA sequence were purchased from Genecopoeia. Site-directed mutagenesis were conducted using QuikChange Site-Directed Mutagenesis Kit (Agilent). The forward and reverse mutagenic primers for *TOP2AY165S* are 5'- aatttggctccagagccatttcgaccacctgtcact -3' and 5'- agtgacaggtggtcgaaatggctctggagccaaatt-3'.

Morpholino injection and validation

Splice-blocking morpholinos targeting zebrafish Top2a (5'- GACATTCATCATAAACTCACCCAGA -3') and Top2b (5'- ATGCAGTAATCTTACCAAGGATCTC -3') were purchased from Gene Tools. To validate their knockdown efficiencies, injected embryos were collected for mRNA extraction using RNeasy Mini Kit (Qiagen) at 1 dpf, followed by reverse transcription using QuantiTect Reverse Transcription Kit (Qiagen) and PCR using primers that flank the targeted exons. Top2a-MO knockdown efficiency was examined using the forward and reverse primers 5'- GGCTCTCAGTAAGCCCAAGAA -3' and 5'- TACTGGAGGTCAGGAGTTGGC -3', with the full-length amplicon predicted to be 401 bp and splice-blocked amplicon 310 bp long. Top2b-MO knockdown efficiency was examined using the forward and reverse primers 5'- AAATGAATGGCCGGGGAGATG -3' and 5'- TGATGGTTGTCATGTTCTTGTCTCT -3', with the full-length amplicon predicted to be 297 bp and splice-blocked amplicon 206 bp long. The toxicity of the morpholinos were examined and the doses determined by following a protocol described in the section "Zebrafish embryonic chemical treatment".

mRNA overexpression and validation

mRNA *in vitro* syntheses were conducted using the mMESSAGE mMACHINE T7 Transcription Kit (Invitrogen). Linearized constructs containing *TOP2A* and *TOP2AY165S* genes were used as templates. Following synthesis, mRNAs were A-tailed using E. coli Poly(A) Polymerase (NEB) to improve translation efficiency. *TOP2A* and *TOP2AY165S* mRNAs were injected to the yolk of 1- to 2-cell stage embryos. For validation of protein overexpression, ~200-300 embryos were dechorionated 3-6 hours after injection by 5 min incubation in 1 mg/ml pronase (Roche) in E3 medium in a 2% agarose-coated petri

dish and washed with excess amount of fresh E3. Dechorionated embryos were transferred to a 1.5 ml tube and deyolked by pipetting up and down 3 to 5 times using a P200 tip in 200 µl ice-cold PBS. 1 ml ice-cold PBS was added, and the resulting mixture was centrifuged at 300 g for 30 seconds. The supernatant was discarded, and the cell pellet was washed with an additional ml of ice-cold PBS. Cells were lysed in 20 µl RIPA buffer (Santa Cruz Biotechnology) supplemented with cOmplete Mini EDTA-free protease inhibitor tablet (Roche), 5 mM sodium orthovanadate (New England Biolabs), and 10 mM sodium fluoride (Sigma-Aldrich), for 50 min on ice with vortexing every 10 min. The resulting lysate was centrifuged at 14000 g for 15 min at 4°C. The entire volume of supernatant (typically corresponding to ~30 µg protein) was saved and subsequently combined with 5× Laemmli buffer prior to gel loading for Western blot analysis. Protein concentration was measured by Bradford (Bio-Rad) assay. Anti-TOP2A (Cell Signaling D10G9) and anti-β-actin (Cell Signaling D6A8) antibodies were used for detection. Signal was detected by ECL.

Zebrafish mutagenesis and genotyping

The complete protocol used to generate and genotype the mutant lines *nrg1*$^{-/-}$ and *disc1*$^{-/-}$ was previously described[109]. Briefly, frameshift mutations were engineered into the *nrg1* and *disc1* genes of TuAB strain zebrafish at exons 2 and 4 respectively, using CRISPR-Cas9 mediated genome editing. The CRIPSR gRNA spacer sequences were 5'- GGCCGAGGGAGTGGTGCTGG -3' (*nrg1*) and 5'- GGATACATGCGGTCTGAGCC -3' (*disc1*). The primers used for genotyping were 5'- TAGGCCGAGGGAGTGGTGCTGG -3' and 5'- AAACCCAGCACCACTCCCTCGG -3' for *nrg1*, and 5'- TAGGCTCAGACCGCATGTATCC -3' and 5'- AAACGGATACATGCGGTCTGAG -3' for *disc1*. A 20 bp deletion (GGATACATGCGGTCTGAGCC) was detected in exon 2 of *disc1* gene, resulting in a premature stop codon (Figure S1E). A 22 bp deletion (CCAGCACCACTCCCTCGGCCAA) was detected in exon 4 of *nrg1* gene, also resulting in a premature stop codon (Figure S1F).

The mutant line *can4* was genotyped using the forward and reverse primers 5'- CTGCAGAAACCCTGTTAAG -3' and 5'- AGGGGATTGACCTCTCGTTG -3'[41]. The mutant line *noto* was genotyped using the forward and reverse primers 5'- TCGGTTCCAAATGTGCTCTCT -3' and 5'- AGCCTTGCAAGCCCTAATCAT -3'[42]. PCR and subsequent Sanger sequencing were performed using the same primers. Genomic DNA extracted from whole fish following Fishbook assays were purified (Zymo Research) before PCR to remove PCR inhibitors in the fish tissues.

Mouse prenatal injection

Assessing pregnancy according to the presence of vaginal plugs can be unreliable[110]. We therefore assessed pregnancy based on weight gain following a reported protocol[110] with modifications. We considered the day after overnight pairwise breeding to be potentially 0 days post coitum (dpc). Females were first weighed on 0 dpc, then single housed until 7 dpc when they were weighed again. Females with weight gains > 1.0 gram were considered pregnant. Drug and vehicle control were injected daily through intraperitoneal injection from 7 dpc to 18 dpc. Births were given on 19 or 20 dpc. ICRF-193 was first dissolved in 100% ethanol (Fisher Scientific), then diluted in saline (Baxter) to a final concentration of 0.01 µg/µl. Diluted drug was injected at a volume of 10 µl per gram body weight to achieve a final dosage of 0.1 mg/kg.

Social interaction assay

Social interaction was examined as described previously[44] using 2 months old mice. Briefly, each adult

mouse was introduced into a neutral and unfamiliar cage. After a 5 min acclimation period, a foreign age-, sex-, and weight-matched untreated mouse (from separate litters) was introduced into the cage as a social stimulus. Both experimental and untreated females were assayed at the metestrus or diestrus stages to minimize the influence of estrous cycle on social behavior. The test mouse was marked by an odorless black marker at the tail to differentiate from the social stimulus mouse. Testing sessions lasted 5 min and were manually scored for social interactions later. Durations of each social interaction were timed using the software BORIS[111]. All mice were scored blind to their treatment conditions. Social interaction was defined as any period of time in which the test mouse was actively investigating the social stimulus mouse. Investigations included sniffing of the social stimulus mouse toward its facial, abdominal, and anogenital areas, grooming, or closely pursuing the stranger as it explored the cage; investigation of the test mouse by the social stimulus mouse was not scored.

Novel object investigation assay

The novel object test was performed in a neutral and unfamiliar cage using 8 months old mice in the exact same manner as the social interaction assay, except with a wooden block introduced to the cage after the acclimation period instead of a social stimulus mouse. Total time spent investigating the object was quantified over a 5 min period. Objects were covered with tape during the test; tape was thoroughly removed and replaced in between tests to remove odor traces.

Ultrasonic vocalization assay

Vocalizations were tested in pups on postnatal days 7 or 8. Cages were acclimatized to the room for 30 min. Pups were individually placed on the test platform in a sound-proof cabinet. Ultrasonic vocalizations were measured for 5 min using an ultrasonic vocalization detector (Avisoft Bioacoustics).

T-maze assay

T-maze assay was performed as previously described[112] using 3 months old mice. Each session consisted of eight consecutive trials. In each trial, the mouse was placed in the start compartment of a T-maze. After 1 min before the first trial and 15 sec for the subsequent trials, the door was lifted, and the mouse was left free to explore the two arms of the maze. As soon as the animal entered (with all four paws) one of the two alternative arms (left or right), the door of that compartment was closed for 30 s to confine the animal. The animal was removed between trials and the T-maze was quickly cleaned with 70% alcohol and dried to remove any odor traces that may affect the performance in the next trial.

Hole-board assay

Hole-board assay was conducted in a 16-hole square apparatus as previously described[44] using 3 months old mice. Each session lasted 10 min and the total number of head pokes was recorded. The occurrence of perseverative behaviors (manifested as the tendency to explore the same holes) was measured via the coefficient of exploratory variation, calculated as the ratio of the standard deviation of the number of head pokes for each hole over the mean for each mouse. This index[113] provides a quantitative estimation of the dispersion of the probability distribution of exploratory activity with regard to perseverative behaviors.

Open field assay

Testing was conducted as described previously[114] using 10 months old mice. The open field consisted of a grey Plexiglas square arena (40 × 40 cm) enclosed by 4 black walls (40 cm high). On the floor, two zones of equal area were defined: a central square zone of 28.28 cm per side, and a concentric peripheral zone including the area within 11.72 cm from the walls. Mice were placed in the central zone at the beginning of each assay and their behaviors were each monitored for 10 min. Animal motions were tracked using Bonsai. Percentage of time spent in the center zone and average velocity were measured using a custom written Python script.

Animal husbandry

All animal husbandry and experiment protocols were approved by and carried out in accordance with the Institutional Animal Care and Use Committee at Massachusetts General Hospital or University of Utah.

Fertilized eggs (up to 10,000 embryos per day) were collected from group mating of EkkWill strain zebrafish (*Danio rerio*) (EkkWill Waterlife Resources). Embryos were raised in HEPES (10 mM) buffered E3 medium at 28 °C, with or without compound treatment, during the first 3 days. At 3 days post fertilization (dpf), chorion debris was removed, and larvae were transferred into petri dishes containing fresh E3 medium. At 5-7 dpf, larvae were transferred into nursery tanks and raised at 28 °C on a 14/10 hr on/off light cycle.

Male C57BL/6J mice (*Mus musculus*) were obtained from Jackson Laboratories. Mice were group-housed at no more than 5 per cage and were maintained on a 12/12 hr light/dark cycle. The room temperature was maintained at 20-23 °C with relative humidity at approximately 50%. Food and water were available *ad libitum* for the duration of the study, except during testing. All tests were conducted during the light phase of the light dark cycle.

Statistical analysis

Graphs were generated using GraphPad Prism or Python. Data were analyzed using 2-tailed Student's t test. Bonferroni correction was used to adjust for multiple comparison, except for ORA conducted using the enrichR R package. P values less than 0.05 were considered significant.

RNA-seq and data analysis

We conducted RNA-seq on 3 dpf *can4*$^{-/-}$ mutants and their WT siblings. Sequencings and initial data processing were conducted at Huntsman Cancer Institute High-Throughput Genomics and Bioinformatic Analysis Shared Resource. We first incrossed heterozygous *can4* mutants to produce a mixture of *can4*$^{+/-}$ (heterozygous), *can4*$^{-/-}$ (homozygous), and *can4*$^{+/+}$ (wild type; WT) embryos. We then lysed 3 dpf larvae individually by rigorously pipetting in 55 µl DNA/RNA Shield (Zymo Research) at RT followed by overnight incubation at 4°C. We extracted genomic DNA from 5 µl lysate of each larvae using Zymo 96 genomic DNA purification kit and genotyped each sample by PCR and Sanger sequencing. The remaining lysates were stored at -80°C while awaiting genotyping results. Once genotyped, larvae with the same genetic background were pooled for mRNA extraction using Direct-zol RNA Miniprep Kit with TRI Reagent from Zymo Research. We prepared 5 biological replicates for each genetic background – *can4*$^{-/-}$ and WT – with each sample containing mRNA extracted from 9 larvae. Libraries were prepared using the Illumina TruSeq Stranded Total RNA Library Prep Ribo-Zero Gold Kit. Sequencing was conducted in NovaSeq with 2 x 150 bp runs at 50 million reads per sample. Sequencing reads were aligned using STAR[115]. Quality control was performed using MultiQC (https://multiqc.info/). Initial data processing and analyses were performed using the R packages DESeq2[116] and hciR

(https://github.com/HuntsmanCancerInstitute/hciR). Data were normalized using the regularized log (rlog) counts method in DESeq2. The *can4*$^{-/-}$ mutant data was compared to WT data to identify genes with significantly altered expressions. The raw and processed RNA-seq data have been deposited in the Gene Expression Omnibus under accession number GSE181730.

Gene sets and gene set libraries

The SFARI[49,50] gene set was acquired from the archived online SFARI gene database (https://gene-archive-dev.sfari.org/tools/). The AutismKB 2.0[51] gene set was downloaded from the AutismKB website (http://db.cbi.pku.edu.cn/autismkb_v2/download.php). The BDgene[52] gene set was downloaded from the BDgene website (http://bdgene.psych.ac.cn/directSearch.do?type=gene&keyword=). The SZGene[54] gene set was acquired from http://www.szgene.org/.The SysID[55] gene set was downloaded from SysID database website (https://sysid.cmbi.umcn.nl/table/human-gene-info). PsyGeNET[53] gene sets for depressive disorders (PsyGeNET Depressive Disorder), bipolar disorder (PsyGeNET Bipolar Disorder), and schizophrenia spectrum and other psychotic disorders (PsyGeNET Schizophrenia) were downloaded from the PsyGeNET website (http://www.psygenet.org/web/PsyGeNET/menu/downloads). The GEPAD (Genomics England's PanelApp database)[56] gene set was acquired from the cited publication.

The gene set libraries DisGeNET, RNAseqGEO (RNA-Seq_Disease_Gene_and_Drug_Signatures_from_GEO), ChIP-X (ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X), ENCODE_HM (ENCODE_Histone_Modifications_2015), and Epigenomics_Roadmap_HM (Epigenomics_Roadmap_HM_ChIP-seq) were downloaded from the Enrichr[63] website (https://amp.pharm.mssm.edu/Enrichr/#stats). The GLAD4U disease gene set library was downloaded from the WebGestalt[117] website (http://www.webgestalt.org/).

Over-representation analysis (ORA)

ORA using DisGeNET, KEGG (KEGG_2019_Human), REACTOME (Reactome_2016), and GO term (biological processes, molecular function, and cellular component) libraries (GO_Biological_Process_2018, GO_Molecular_Function_2018, GO_Cellular_Component_2018) were performed in R using the enrichR package (https://cran.r-project.org/web/packages/enrichR/index.html). Results were ranked based on adjusted *p*-value.

Other ORA assays were performed in R using a one-sided Fisher's exact test with 95% confidence calculated according to the R function fisher.test, and Bonferroni correction was used to adjust for multiple comparison in these tests. Total number of human coding genes was set at 20438 at the time of analysis (GRCh38.p13; https://uswest.ensembl.org/Homo_sapiens/Info/Annotation).

Permutation analysis was conducted by randomly selecting 5000 genes from all 14989 human orthologs of zebrafish genes for each round of permutation. The null distribution consisted of odds ratios calculated for 1000 permutations. The 14989 human orthologs of zebrafish genes were identified using the R package biomaRt[118,119], based on zebrafish genome assembly GRCz11 and human genome assembly GRCh38.

Gene set enrichment analysis (GSEA)

GSEA analyses for DisGeNET, GLAD4U , GO, KEGG, and REACTOME libraries were performed using the web-based gene set analysis toolkit[117] (WebGestalt; http://www.webgestalt.org/). GSEA

analysis for the independent disease gene sets were performed using the GSEA software[120,121].

Upstream Enrichment Signature (UES) and UES-blast

UES and UES-blast analyses were conducted in R using custom scripts. For each test gene set, we first run an ORA analysis to find significantly enriched upstream regulators using the ChIP-X library (ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X). As suggested by its full name, this library contains gene sets acquired by analyzing ChIP-X data in the ENCODE and ChEA databases. We then calculate a significance score for each significantly enriched (adjusted p-value < 0.05) upstream regulator as follows: significance score = - log(adjusted p-value). Significance scores for all significantly enriched upstream regulators of a given gene set are normalized to between 0 and 1. All nonsignificant upstream regulators are assigned a significance score of 0. With a score assigned to each upstream regulator, we created a signature for each gene set which we named Upstream Enrichment Signature (UES).

We constructed a reference dataset containing >1700 UESs from analyzing published disease risk genes including DisGeNET and GLAD4U, and an RNA-seq gene set library RNAseqGEO. To avoid variations caused by gene sets containing too few genes, only gene sets containing ≥500 genes from the DisGeNET library and ≥250 genes from the GLAD4U library were analyzed. We then conduct an unbiased search using a custom-written blast-style querying algorithm (UES-blast) to find gene sets with the lowest blast scores (highest similarity rankings) to a query gene set, e.g., SFARI genes. UES-blast was conducted by first calculating the Euclidean distances between test gene sets and the target gene set (e.g., SFARI genes), and then ranking test gene sets based on their distances to the target gene set.

Histone modifications enrichment analysis and H3K27me3 score

We assessed enrichment of histone marks in the promoter regions of genes in each test gene set using ORA analyses. For each given histone mark, the curated ENCODE_HM and Epigenomics_Roadmap_HM libraries often contain a number of query gene sets each listing genes that are targeted by this histone mark in their promoter regions in a specific cell or tissue type. For example, a query gene set named “H3K27me3 cerebellum mm9” contains all genes that are marked by H3K27me3 in the mouse cerebellum, based on analyzing published ChIP-X data. In our analysis, we run ORA analyses for each test gene set against all query gene sets in the two libraries to identify those that are enriched in our test gene set. We then count the number of significantly enriched query gene sets for each histone mark, such as the data shown in Figure 7E.

To specifically quantify enrichment of H3K27me3 in each test gene set, we calculated a H3K27me3 score as follows: H3K27me3 score = the ratio of significant H3K27me3 gene sets over all H3K27me3 gene sets in the library × the ratio of significant H3K27me3 gene sets over all significant histone mark gene sets = (number of significant H3K27me3 gene sets / number of all H3K27me3 gene sets) × (number of significant H3K27me3 gene sets / number of all significant histone marks gene sets). As an example, if a test gene set is significantly enriched for 60 H3K27me3 gene sets out of a total number of 100 H3K27me3 gene sets in the library, and at the same time also significantly enriched for another 10 assorted histone marks gene sets, then its H3K27me3 score is calculated as (60 / 100) × (60 / (60 + 10)) = 0.6 × 6/7 = 0.51. For each test gene set, its final H3K27me3 score was acquired by first calculating two separate H3K27me3 scores using the ENCODE_HM and Epigenomics_Roadmap_HM libraries, respectively, and then combining these two scores by summation. All analyses were conducted in R using custom scripts.

ChIP-seq data analyses

Genome-wide TOP2A binding data was acquired from GEO accession number GSE79593[61]; we used the DMSO-treated sample in this dataset for analyses. EZH2, SUZ12, and H3K27me3 ChIP-seq data were acquired from GEO accession numbers GSM1003576, GSM1003545, and GSM733658, respectively. To enable meaningful comparison with TOP2A binding peaks, all ChIP-seq data were acquired from the same cell line K562 cells. Peaks from the EZH2, SUZ12, and H3K27me3 that overlap with TOP2A peaks were identified using the R package GenomicRanges[122]. Genes that are targeted by the original peaks or the overlapping peaks at their promoter or gene body regions were mapped using the R package bumphunter[123,124] and exported for subsequent analyses. Venn diagram for peak overlaps was drawn using the R package ChIPseeker[125].

Code availability

Code is available on the GitHub repository at https://github.com/yijie-geng/Fishbook and is archived on Zenodo under DOI 10.5281/zenodo.5171582.

AUTHOR CONTRIBUTIONS

Y.G. conceived the study, built the equipment, designed and conducted the experiments, wrote the codes, analyzed the data, and wrote the manuscript; S.C.G. designed the rodent experiments; A.K.N. and J.J.Y. designed the CRISPR knockout experiments; T.Z., B.R.P., D.L.H., and I.A.D. conducted the experiments; M.B. designed the rodent experiments, interpreted the data, and edited the manuscript; R.T.P conceived the study, designed the experiments, interpreted the data, and wrote the manuscript. All authors contributed meaningful insights during discussions and reviewed and approved the final version of the manuscript.

ACKNOWLEDGEMENTS

The *can4* and *noto* zebrafish mutants were generously provided by Dr. Leonard Zon's lab and Dr. Herwig Baier's lab, respectively. We thank members of our research groups, in particular A. Rennekamp, Y. Jin, L. Mosher, G. Floris, and R. Cadeddu, for helpful advice. We would like to thank the University of Utah Comparative Medicine Center, the Centralized Zebrafish Animal Resource, and the Massachusetts General Hospital Cardiovascular Research Center zebrafish facility for animal husbandry. We also thank the University of Utah Huntsman Cancer Institute High-Throughput Genomics and Bioinformatic Analysis Shared Resource for genomics experiments and initial analyses. This work was supported by the L. S. Skaggs Presidential Endowed Chair and by the National Institute of Environmental Health Sciences of the National Institutes of Health under Award Number K99ES031050. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

FIGURE LEGENDS

Figure 1. The Fishbook assay.

(A, B) The design and a 3D printed array of Fishbook arenas.

(C) Screen shot of a Fishbook assay. Red rectangle: a Fishbook arena; 1: social stimulus compartment; 2: social test compartment; 3: empty control compartment; 4: transparent windows.

(D) An array of 44 arenas.

(E) The Fishbook imaging station. 1: telecentric lens; 2: white light LED panels providing ambient light; 3: sample deck; 4: 850 nm infrared LED backlight panel; 5: CCD camera.

(F) Social score is calculated as the average Y-axis position for the duration of a test.

(G-I) Fishbook with or without social stimulus fish (G), showing boxplots (H) and representative tracking plots (I) of WT fish tested with social stimulus (Social Stim; n = 44) or without social stimulus (No Stim; n = 44). In each boxplot, box encloses data points from the 25th percentile to the 75th percentile, horizontal line and cross mark the median and the mean, the lines above and below the box reach datapoints with the maximum and minimum values. ****: $p<0.0001$.

(J-L) Fishbook detects social deficits in mutants (J), comparing WT (n=24) and homozygous *nrg1* knockout (*nrg1*$^{-/-}$; n=23) fish, both from a heterozygous (*nrg1*$^{+/-}$) incross and genotyped individually after Fishbook assay (K), and WT sibling (n=209) and homozygous *disc1* knockout (*disc1*$^{-/-}$; n=122) fish (L). *: $p<0.05$, **: $p<0.01$.

Figure 2. Screening for inhibitors of social development identifies fluoroquinolones and other Top2 inhibitors.

(A) The screening workflow.

(B) Fluoroquinolones induce social deficits. Comparing DMSO (n=68, 22, 82, from left to right) with flumequine (Flum, 25 µM, n=56), lomefloxacin (Lome, 100 µM, n=60), ofloxacin (Ofl, 200 µM, n=22), and oxolinic acid (OA, 200 µM, n=62).

(C) The catalytic cycle of Top2 and inhibitors targeting different steps of the cycle.

(D-F) Eukaryotic Top2 inhibitors induce social deficits (D). Comparing fish treated with vehicle control (0 µM in each panel), Top2 poisons ellipticine (0 µM: n=48; 2 µM: n=31), amsacrine (0 µM: n=21; 200 µM: n=20), etoposide (0 µM: n=24; 10 µM: n=11), and mitoxantrone (0 µM: n=38; 5 µM: n=29) (E), and Top2 catalytic inhibitors aclarubicin (0 µM: n=22; 8 µM: n=6), ICRF-193 (0 µM: n=65; 100 µM: n=29), and sobuzoxane (0 µM: n=27; 100 µM: n=8) (F).

*: $p<0.05$, **: $p<0.01$, ***: $p<0.001$, ****: $p<0.0001$.

Figure 3. Top2a rather than Top2b is required for the development of social behavior.

(A) Top2a-MO (0.05 mM, n=44) induces social deficits compared to vehicle (H_2O, n=44), which is rescued by *hTOP2A* mRNA (250 ng/µl, Top2a-MO+A, n=44).

(B) Top2b-MO (0.05 mM, n=50) does not induce social deficits compared to vehicle (H_2O, n=71).

(C) Top2a selective inhibitor sodium salicylate (100 µM) induces social deficits.

(D) Heterozygous *can4* mutant (*can4*$^{+/-}$, n=20) exhibits social deficits compared to WT (n=20). Both mutant and WT were acquired from a heterozygous (*can4*$^{+/-}$) incross and genotyped individually after Fishbook assay.

(E) Heterozygous *noto* mutant (*noto*$^{+/-}$, n=32) shows no social deficits compared to WT (n=21). Both mutant and WT were acquired from a heterozygous (*noto*$^{+/-}$) incross and genotyped individually after Fishbook assay.

(F) ICRF-193 (I, 0.1 mM, n=23) induced social deficits compared to the vehicle control ethanol (E, 1%, n=50), which is rescued by *hTOP2AY165S* mRNA (250 ng/µl, I+A^{Y165S}, n=31).

ns: not significant, *: p<0.05, **: p<0.01, ****: p<0.0001.

Figure 4. Prenatal inhibition of Top2 induces social communication impairments and increased perseveration in mouse.

(A) Experimental design for prenatal ICRF-193 injection and subsequent experiments for the offspring.

(B-D) ICRF-193 (I, n=16) reduces total duration (B), total number (C), and mean duration (D) of social investigation compared to vehicle control ethanol (E, n=20).

(E-G) ICRF-193 (n=16) does not reduce total duration (E), total number (F), and mean duration (G) of novel object investigation compared to ethanol (n=19).

(H, I) ICRF-193 (n=16) does not change percentage of time spent in the center zone (H) and velocity (I) of mice in open field assay compared to ethanol (n=18).

(J, K) ICRF-193 (n=16) reduces total number (J) but increases mean peak frequency (K) of ultrasonic vocalization compared to ethanol (n=20).

(L, M) ICRF-193 (n=16) reduces total number of alterations (number of times a mouse enters an alternative arm compared to the previous trial) (L) while increases number of trials prior to the first alteration (M) compared to ethanol (n=20) during a T-maze assay consisted of 8 trials.

(N) ICRF-193 (n=16) increases coefficient of variation in hole-board assay compared to ethanol (n=20).

ns: not significant, *: p<0.05, **: p<0.01, ***: p<0.001, ****: p<0.0001.

Figure 5. RNA-seq reveals downregulation of autism risk genes caused by Top2a depletion.

(A) RNA-seq experimental design.

(B) can4Dn but not can4Up is selectively enriched for autism and its comorbidities risk genes. For each cell, color represents odds ratio, and numbers represent ranking out of 24166 diseases in DisGeNET library (before slash) and adjusted p-value (after slash).

(C, D) GSEA analysis using the DisGeNET library shows enrichment for autism and its comorbidities risk genes (C) and neurological conditions risk genes (D) in downregulated. Value inside each bar represents FDR (C). Significance: FDR<5%.

Figure 6. Upstream analyses find PRC2 and H3K27me3 as common upstream regulators of can4Dn and autism risk genes.

(A) SFARI genes and can4Dn, but not can4Up, share very similar enrichment profiles for upstream regulators. Value in each cell represents the significance score (Methods). In row labels, c: ChEA, e: ENCODE, indicating source of ChIP-seq data.

(B) UES analysis and UES-blast workflow. ORA analysis finds significantly enriched upstream regulators for a test gene set using reference gene sets from the ChIP-X library. A significance score is calculated for each upstream regulator (Methods), creating a signature for each test gene set which we named Upstream Enrichment Signature (UES). A blast-style querying algorithm (UES-blast) then identifies test gene sets with the most similar UES compared to a query gene set, e.g., SFARI genes. UR: upstream regulator.

(C) UES-blast (Methods) shows that can4Dn, can4Dn-SFARI, and several independently curated autism risk gene sets rank high among control gene sets against the query gene set SFARI genes, indicating that they share a very similar UES with the SFARI genes.

(D) tSNE clustering of can4Dn, can4Dn-SFARI, autism risk gene sets, and DisGeNET gene sets (only those containing > 500 genes). Letters label autism comorbidities risk gene sets, I: intellectual disability; D: depression; S: schizophrenia; B: bipolar disorder.

(E) Genes in the SFARI and can4Dn gene sets, but not in can4Up, are specifically enriched for the H3K27me3 mark.

(F) can4Dn, can4Dn-SFARI, and autism risk gene sets possess the highest H3K27me3 scores when compared to the control gene sets.

Figure 7. PRC2 and H3K27me3 mediate the Top2a-dependent gene regulation.

(A) TOP2A binding peaks significantly overlap with EZH2, SUZ12, and H3K27me3.

(B) Among all EZH2, SUZ12, and H3K27me3 binding peaks, those overlap with TOP2A are more enriched for autism and its comorbidities risk genes and can4Dn, but not for can4Up and neurodegenerative disorder risk gene sets. Significance was calculated by paired t test. NDD: neurodegenerative disorders.

(C) A working model for how PRC2 and H3K27me3 mediate the Top2a-dependent gene regulation and social development.

(D) Embryonic (0-3 dpf) co-treatment of 10 µM UNC1999 (U99) with 100 µM sodium salicylate (SS, n=35), 15 µM flumequine (Flum, n=37), and 0.5 µM ellipticine (Ellip, n=44), respectively, rescues social deficits caused by SS (n=23), Flum (n=37), and Ellip (n=29) alone compared to DMSO control (D; left to right: n=21, 16, 15). UNC1999's inactive analog UNC2400 (U00) does not rescue social deficits (left to right: n=38, 19, 42).

(E) Social deficits caused by embryonic treatment of SS (n=38), Flum (n=45), and Ellip (n=61) compared to DMSO (left to right: n=23, 34, 47) are rescued by overnight exposure to U99 (left to right: n=44, 39, 56) but not U00 (left to right: n=42, 41, 67) at the juvenile stage (20-21 dpf).

ns: not significant, *: $p<0.05$, ***: $p<0.001$, ****: $p<0.0001$.

SUPPLEMENTAL FIGURE LEGENDS

Figure S1.

(A) Boxplot comparing social scores of WT fish in light (n = 87) or dark (n = 44). ****: $p < 0.0001$.

(B) Boxplot comparing social scores of WT fish in response to a social stimulus (fish; n = 22) or pebble (n = 21). ***: $p < 0.001$.

(C) DNA and protein sequences of WT and *nrg1* knockout fish.

(D) DNA and protein sequences of WT and *disc1* knockout fish.

(E) 3D printing design for the view-blocking comb.

(F) Image of a comb placed on a Fishbook test arena. The comb is not inserted in this image.

Figure S2.

(A) Chemical structures of fluoroquinolones.

(B) Chemical structures of eukaryotic Top2 inhibitors.

(C) Boxplot showing social scores of fish treated with DMSO (n=44), chlorpyrifos (C; 5 µM; n=57), chlorpyrifos oxon (CO; 1 µM; n=30), and genistein (G; 8 µM; n=22).

(D) Boxplot comparing social scores of fish treated with DMSO (n = 51) or actinomycin D (20 µM; n = 40).

(E) Boxplot comparing social scores of fish treated with DMSO (n = 34) or triptolide (0.2 µM; n = 44).

(F) Boxplot comparing social scores of fish treated with DMSO (n = 97), bleomycin (1 µM; n = 12), or hydroxyurea (HU; 10 mM, n = 21; 20 mM, n = 7).

ns: not significant, ***: $p<0.001$, ****: $p<0.0001$.

Figure S3.

(A) DNA gel image showing RT-PCR result of 1 dpf embryos injected with Top2a-MO (0.05 mM), Top2b-MO (0.05 mM), or un-injected (Controls 1 & 2). n = ~ 80 – 100 embryos for each condition. *: splice-blocked amplicon of Top2a. **: splice-blocked amplicon of Top2b.

(B, C) Western blot images showing protein overexpression following injection of 250 ng/µl *hTOP2A* (B) and *hTOP2AY165S* (C) mRNAs in zebrafish embryos. HEK293 cell lysate was used as positive controls. Uninjected zebrafish embryos were used as negative controls. β-actin was used as a loading control.

(D) Boxplot comparing the body weights of mice treated with ethanol (male: n = 13, female: n = 7), ICRF193 (male: n = 11, female: n = 5), and no-treatment control (male: n = 13, female: n = 13) at 2 months of age. ns: not significant.

Figure S4.

(A) can4Dn but not can4Up is selectively enriched for autism and its comorbidities risk gene sets from the GLAD4U library. For each cell, color represents odds ratio, and numbers represent ranking out of

3071 diseases in GLAD4U library (before slash) and adjusted *p*-value (after slash).

(B) Heatmap showing ORA analysis comparing can4Dn and can4Up using several independent disease risk gene sets related to autism and its comorbid disorders. Color in each cell represents odds ratio values. Number in each cell represents adjusted *p*-value.

(C) Bar chart showing normalized enrichment scores from GSEA analysis for RNA-seq data using the DisGeNET library.

Figure S5.

(A, B) GSEA analysis of RNA-seq data using the GLAD4U library shows enrichment for autism and its comorbidities risk genes (A & B) and neurological conditions risk genes (B) in downregulated but not upregulated genes. Value inside each bar represents FDR (A). Significance: FDR<5%.

(C) Bar charts showing normalized enrichment scores (NESs) from GSEA analysis for RNA-seq data using the GLAD4U library. Left: hits with positive NESs; right: hits with negative NESs.

(D) Bar chart showing normalized enrichment scores from GSEA analysis for RNA-seq data using several independent disease gene sets related to autism and its comorbid disorders. Significance: FDR < 5%.

Figure S6.

(A) Heatmap showing ORA analysis comparing can4Dn and can4Up using KEGG, REACTOME, and GO libraries. Color in each cell represents odds ratio.

(B) Bar chart showing GSEA analysis of KEGG, Reactome, and GO libraries using the RNA-seq data.

Figure S7.

(A) Histogram showing the null distribution of odds ratios generated by 1000 permutations by randomly selecting 5000 genes out of all human orthologs of zebrafish genes and conduct ORA analysis using selected gene sets from the DisGeNET library. The odds ratios for can4Dn (red) and can4Up (blue) are marked by colored vertical lines, and their respective *p*-values are shown as numbers in legends. A *p*-value <0.008 was considered significant to correct for multiple comparisons.

(B) Histogram showing the null distribution of odds ratios generated by 1000 permutations by randomly selecting 5000 genes out of all human orthologs of zebrafish genes and conduct ORA analysis using selected gene sets from the GLAD4U library. The odds ratios for can4Dn (red) and can4Up (blue) are marked by colored vertical lines, and their respective *p*-values are shown as numbers in legends. A *p*-value <0.008 was considered significant to correct for multiple comparisons.

Figure S8.

Histogram showing the null distribution of odds ratios generated by 1000 permutations by randomly selecting 5000 genes out of all human orthologs of zebrafish genes and conduct ORA analysis using selected gene sets from several independent disease gene sets related to autism and its comorbid conditions. The odds ratios for can4Dn (red) and can4Up (blue) are marked by colored vertical lines, and

their respective *p*-values are shown as numbers in legends. A *p*-value <0.005 was considered significant to correct for multiple comparisons.

Figure S9.

Histogram showing the null distribution of odds ratios generated by 1000 permutations by randomly selecting 5000 genes out of all human orthologs of zebrafish genes and conduct ORA analysis using selected gene sets from KEGG, Reactome, and GO libraries. The odds ratios for can4Dn (red) and can4Up (blue) are marked by colored vertical lines, and their respective *p*-values are shown as numbers in legends. A *p*-value <0.003 was considered significant to correct for multiple comparisons.

Figure S10.

(A) A detailed look at UES-blast rankings of the key gene sets in Figure 6C. Bar chart shows the UES-blast scores of autism risk gene sets (green), can4Dn (red), can4Dn-SFARI (orange), autism comorbid conditions risk gene sets (yellow; D: depression, B: dipolar disorder, S: schizophrenia, I: intellectual disability), can4Up (dark blue), and neurodegenerative disorders risk gene sets (blue). Numbers in the x-axis labels before gene set names represent the UES-blast rankings of each labeled gene set.

(B) A detailed look at H3K27me3 score rankings of the key gene sets in Figure 6F. Bar chart shows the H3K27me3 scores of autism risk gene sets (green), can4Dn (red), can4Dn -SFARI (orange), autism comorbid conditions risk gene sets (yellow; D: depression, B: dipolar disorder, S: schizophrenia, I: intellectual disability), can4Up (dark blue), and neurodegenerative disorders risk gene sets (blue). Numbers in the x-axis labels before gene set names represent the H3K27me3 score rankings of each labeled gene set.

Figure S11.

tSNE clustering of can4Dn, can4Dn-SFARI, autism risk gene sets, neurodegenerative disorders risk gene sets, and all control gene sets from the reference dataset. Letters label autism comorbidities risk gene sets, I: intellectual disability; D: depression; S: schizophrenia; B: bipolar disorder.

Supplemental Video.

Video recording of a Fishbook test. Played at 4× the actual speed.

Table S1.

The can4Dn and can4Up gene lists.

REFERENCES

1 Fantz, R. L. Pattern Vision in Newborn Infants. *Science* **140**, 296-297, doi:10.1126/science.140.3564.296 (1963).

2 Kennedy, D. P. & Adolphs, R. The social brain in psychiatric and neurological disorders. *Trends Cogn Sci* **16**, 559-572, doi:10.1016/j.tics.2012.09.006 (2012).

3 Elder, J. H., Kreider, C. M., Brasher, S. N. & Ansell, M. Clinical impact of early diagnosis of autism on the prognosis and parent-child relationships. *Psychol Res Behav Manag* **10**, 283-292, doi:10.2147/PRBM.S117499 (2017).

4 Willsey, A. J. *et al.* Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* **155**, 997-1007, doi:10.1016/j.cell.2013.10.020 (2013).

5 Li, M. *et al.* Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science* **362**, doi:10.1126/science.aat7615 (2018).

6 Schork, A. J. *et al.* A genome-wide association study of shared risk across psychiatric disorders implicates gene regulation during fetal neurodevelopment. *Nat Neurosci* **22**, 353-361, doi:10.1038/s41593-018-0320-0 (2019).

7 Ruzzo, E. K. *et al.* Inherited and De Novo Genetic Risk for Autism Impacts Shared Networks. *Cell* **178**, 850-866 e826, doi:10.1016/j.cell.2019.07.015 (2019).

8 Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet* **51**, 431-444, doi:10.1038/s41588-019-0344-8 (2019).

9 Satterstrom, F. K. *et al.* Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell* **180**, 568-584 e523, doi:10.1016/j.cell.2019.12.036 (2020).

10 Coe, B. P. *et al.* Neurodevelopmental disease genes implicated by de novo mutation and copy number variation morbidity. *Nat Genet* **51**, 106-116, doi:10.1038/s41588-018-0288-4 (2019).

11 Gaugler, T. *et al.* Most genetic risk for autism resides with common variation. *Nat Genet* **46**, 881-885, doi:10.1038/ng.3039 (2014).

12 Sandin, S. *et al.* The familial risk of autism. *JAMA* **311**, 1770-1777, doi:10.1001/jama.2014.4144 (2014).

13 Panula, P. *et al.* The comparative neuroanatomy and neurochemistry of zebrafish CNS systems of relevance to human neuropsychiatric diseases. *Neurobiol Dis* **40**, 46-57, doi:10.1016/j.nbd.2010.05.010 (2010).

14 Norton, W. H. Toward developmental models of psychiatric disorders in zebrafish. *Front Neural Circuits* **7**, 79, doi:10.3389/fncir.2013.00079 (2013).

15 Geng, Y. & Peterson, R. T. The zebrafish subcortical social brain as a model for studying social behavior disorders. *Dis Model Mech* **12**, doi:10.1242/dmm.039446 (2019).

16 Campbell, P. D. & Granato, M. Zebrafish as a tool to study schizophrenia-associated copy number variants. *Dis Model Mech* **13**, doi:10.1242/dmm.043877 (2020).

17 Ijaz, S. & Hoffman, E. J. Zebrafish: A Translational Model System for Studying Neuropsychiatric Disorders. *J Am Acad Child Adolesc Psychiatry* **55**, 746-748, doi:10.1016/j.jaac.2016.06.008 (2016).

18 Tang, W. *et al.* Genetic Control of Collective Behavior in Zebrafish. *iScience* **23**, 100942, doi:10.1016/j.isci.2020.100942 (2020).

19 Stednitz, S. J. *et al.* Forebrain Control of Behaviorally Driven Social Orienting in Zebrafish. *Curr Biol* **28**, 2445-2451 e2443, doi:10.1016/j.cub.2018.06.016 (2018).

20 Chou, M. Y. *et al.* Social conflict resolution regulated by two dorsal habenular subregions in zebrafish. *Science* **352**, 87-90, doi:10.1126/science.aac9508 (2016).

21 Hoffman, E. J. *et al.* Estrogens Suppress a Behavioral Phenotype in Zebrafish Mutants of the Autism Risk Gene, CNTNAP2. *Neuron* **89**, 725-733, doi:10.1016/j.neuron.2015.12.039 (2016).
22 Dreosti, E., Lopes, G., Kampff, A. R. & Wilson, S. W. Development of social behavior in young zebrafish. *Front Neural Circuits* **9**, 39, doi:10.3389/fncir.2015.00039 (2015).
23 Hinz, F. I., Aizenberg, M., Tushev, G. & Schuman, E. M. Protein synthesis-dependent associative long-term memory in larval zebrafish. *J Neurosci* **33**, 15382-15387, doi:10.1523/JNEUROSCI.0560-13.2013 (2013).
24 Engeszer, R. E., Barbiano, L. A., Ryan, M. J. & Parichy, D. M. Timing and plasticity of shoaling behaviour in the zebrafish, Danio rerio. *Anim Behav* **74**, 1269-1275, doi:10.1016/j.anbehav.2007.01.032 (2007).
25 Hinz, R. C. & de Polavieja, G. G. Ontogeny of collective behavior reveals a simple attraction rule. *Proc Natl Acad Sci U S A* **114**, 2295-2300, doi:10.1073/pnas.1616926114 (2017).
26 Liang, H., Wu, X., Yalowich, J. C. & Hasinoff, B. B. A three-dimensional quantitative structure-activity analysis of a new class of bisphenol topoisomerase IIalpha inhibitors. *Mol Pharmacol* **73**, 686-696, doi:10.1124/mol.107.041624 (2008).
27 Bender, R. P., Lehmler, H. J., Robertson, L. W., Ludewig, G. & Osheroff, N. Polychlorinated biphenyl quinone metabolites poison human topoisomerase IIalpha: altering enzyme function by blocking the N-terminal protein gate. *Biochemistry* **45**, 10140-10152, doi:10.1021/bi0524666 (2006).
28 Srinivasan, A., Robertson, L. W. & Ludewig, G. Sulfhydryl binding and topoisomerase inhibition by PCB metabolites. *Chem Res Toxicol* **15**, 497-505, doi:10.1021/tx010128+ (2002).
29 Lu, C. *et al.* Chlorpyrifos Induces MLL Translocations Through Caspase 3-Dependent Genomic Instability and Topoisomerase II Inhibition in Human Fetal Liver Hematopoietic Stem Cells. *Toxicol Sci* **147**, 588-606, doi:10.1093/toxsci/kfv153 (2015).
30 Wang, X. *et al.* Effects of beta-diketone antibiotic mixtures on behavior of zebrafish (Danio rerio). *Chemosphere* **144**, 2195-2205, doi:10.1016/j.chemosphere.2015.10.120 (2016).
31 Takei, M., Fukuda, H., Yasue, T., Hosaka, M. & Oomori, Y. Inhibitory activities of gatifloxacin (AM-1155), a newly developed fluoroquinolone, against bacterial and mammalian type II topoisomerases. *Antimicrob Agents Chemother* **42**, 2678-2681 (1998).
32 Huang, H. S. *et al.* Topoisomerase inhibitors unsilence the dormant allele of Ube3a in neurons. *Nature* **481**, 185-189, doi:10.1038/nature10726 (2011).
33 King, I. F. *et al.* Topoisomerases facilitate transcription of long genes linked to autism. *Nature* **501**, 58-62, doi:10.1038/nature12504 (2013).
34 Norman, U. & Cicek, A. E. ST-Steiner: A Spatio-Temporal Gene Discovery Algorithm. *Bioinformatics*, doi:10.1093/bioinformatics/btz110 (2019).
35 Pramparo, T. *et al.* Cell cycle networks link gene expression dysregulation, mutation, and brain maldevelopment in autistic toddlers. *Mol Syst Biol* **11**, 841, doi:10.15252/msb.20156108 (2015).
36 Lam, C. W., Yeung, W. L. & Law, C. Y. Global developmental delay and intellectual disability associated with a de novo TOP2B mutation. *Clin Chim Acta* **469**, 63-68, doi:10.1016/j.cca.2017.03.022 (2017).
37 Pommier, Y., Leo, E., Zhang, H. & Marchand, C. DNA topoisomerases and their poisoning by anticancer and antibacterial drugs. *Chem Biol* **17**, 421-433, doi:10.1016/j.chembiol.2010.04.012 (2010).
38 Bau, J. T., Kang, Z., Austin, C. A. & Kurz, E. U. Salicylate, a catalytic inhibitor of topoisomerase II, inhibits DNA cleavage and is selective for the alpha isoform. *Mol Pharmacol* **85**, 198-207, doi:10.1124/mol.113.088963 (2014).

39 Pommier, Y., Sun, Y., Huang, S. N. & Nitiss, J. L. Roles of eukaryotic topoisomerases in transcription, replication and genomic stability. *Nat Rev Mol Cell Biol* **17**, 703-721, doi:10.1038/nrm.2016.111 (2016).
40 Sapetto-Rebow, B. *et al.* Maternal topoisomerase II alpha, not topoisomerase II beta, enables embryonic development of zebrafish top2a-/- mutants. *BMC Dev Biol* **11**, 71, doi:10.1186/1471-213X-11-71 (2011).
41 Dovey, M. *et al.* Topoisomerase II alpha is required for embryonic development and liver regeneration in zebrafish. *Mol Cell Biol* **29**, 3746-3753, doi:10.1128/MCB.01684-08 (2009).
42 Nevin, L. M., Xiao, T., Staub, W. & Baier, H. Topoisomerase IIbeta is required for lamina-specific targeting of retinal ganglion cell axons and dendrites. *Development* **138**, 2457-2465, doi:10.1242/dev.060335 (2011).
43 Deng, S. *et al.* Dexrazoxane may prevent doxorubicin-induced DNA damage via depleting both topoisomerase II isoforms. *BMC Cancer* **14**, 842, doi:10.1186/1471-2407-14-842 (2014).
44 Bortolato, M. *et al.* Monoamine oxidase A and A/B knockout mice display autistic-like features. *Int J Neuropsychopharmacol* **16**, 869-888, doi:10.1017/S1461145712000715 (2013).
45 Roy, S., Watkins, N. & Heck, D. Comprehensive analysis of ultrasonic vocalizations in a mouse model of fragile X syndrome reveals limited, call type specific deficits. *PLoS One* **7**, e44816, doi:10.1371/journal.pone.0044816 (2012).
46 Wohr, M. *et al.* Lack of parvalbumin in mice leads to behavioral deficits relevant to all human autism core symptoms and related neural morphofunctional abnormalities. *Transl Psychiatry* **5**, e525, doi:10.1038/tp.2015.19 (2015).
47 Pinero, J. *et al.* DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res* **45**, D833-D839, doi:10.1093/nar/gkw943 (2017).
48 Jourquin, J., Duncan, D., Shi, Z. & Zhang, B. GLAD4U: deriving and prioritizing gene lists from PubMed literature. *BMC Genomics* **13 Suppl 8**, S20, doi:10.1186/1471-2164-13-S8-S20 (2012).
49 Banerjee-Basu, S. & Packer, A. SFARI Gene: an evolving database for the autism research community. *Dis Model Mech* **3**, 133-135, doi:10.1242/dmm.005439 (2010).
50 Abrahams, B. S. *et al.* SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol Autism* **4**, 36, doi:10.1186/2040-2392-4-36 (2013).
51 Yang, C. *et al.* AutismKB 2.0: a knowledgebase for the genetic evidence of autism spectrum disorder. *Database (Oxford)* **2018**, doi:10.1093/database/bay106 (2018).
52 Chang, S. H. *et al.* BDgene: a genetic database for bipolar disorder and its overlap with schizophrenia and major depressive disorder. *Biol Psychiatry* **74**, 727-733, doi:10.1016/j.biopsych.2013.04.016 (2013).
53 Gutierrez-Sacristan, A. *et al.* PsyGeNET: a knowledge platform on psychiatric disorders and their genes. *Bioinformatics* **31**, 3075-3077, doi:10.1093/bioinformatics/btv301 (2015).
54 Allen, N. C. *et al.* Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database. *Nat Genet* **40**, 827-834, doi:10.1038/ng.171 (2008).
55 Kochinke, K. *et al.* Systematic Phenomics Analysis Deconvolutes Genes Mutated in Intellectual Disability into Biologically Coherent Modules. *Am J Hum Genet* **98**, 149-164, doi:10.1016/j.ajhg.2015.11.024 (2016).
56 Ilyas, M., Mir, A., Efthymiou, S. & Houlden, H. The genetics of intellectual disability: advancing technology and gene editing. *F1000Res* **9**, doi:10.12688/f1000research.16315.1 (2020).
57 Parikshak, N. N. *et al.* Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* **155**, 1008-1021, doi:10.1016/j.cell.2013.10.031 (2013).
58 McFadden, K. & Minshew, N. J. Evidence for dysregulation of axonal growth and guidance in the etiology of ASD. *Front Hum Neurosci* **7**, 671, doi:10.3389/fnhum.2013.00671 (2013).

59 Pinto, D. *et al.* Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *Am J Hum Genet* **94**, 677-694, doi:10.1016/j.ajhg.2014.03.018 (2014).
60 Rojas, D. C. The role of glutamate and its receptors in autism and the use of glutamate receptor antagonists in treatment. *J Neural Transm (Vienna)* **121**, 891-905, doi:10.1007/s00702-014-1216-0 (2014).
61 Yu, X. *et al.* Genome-wide TOP2A DNA cleavage is biased toward translocated and highly transcribed loci. *Genome Res* **27**, 1238-1249, doi:10.1101/gr.211615.116 (2017).
62 Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128, doi:10.1186/1471-2105-14-128 (2013).
63 Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44**, W90-97, doi:10.1093/nar/gkw377 (2016).
64 Streubel, G. *et al.* The H3K36me2 Methyltransferase Nsd1 Demarcates PRC2-Mediated H3K27me2 and H3K27me3 Domains in Embryonic Stem Cells. *Mol Cell* **70**, 371-379 e375, doi:10.1016/j.molcel.2018.02.027 (2018).
65 Consortium, E. P. *et al.* Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699-710, doi:10.1038/s41586-020-2493-4 (2020).
66 Cai, Y. *et al.* H3K27me3-rich genomic regions can function as silencers to repress gene expression via chromatin interactions. *Nat Commun* **12**, 719, doi:10.1038/s41467-021-20940-y (2021).
67 Thakurela, S. *et al.* Gene regulation and priming by topoisomerase IIalpha in embryonic stem cells. *Nat Commun* **4**, 2478, doi:10.1038/ncomms3478 (2013).
68 Konze, K. D. *et al.* An orally bioavailable chemical probe of the Lysine Methyltransferases EZH2 and EZH1. *ACS Chem Biol* **8**, 1324-1334, doi:10.1021/cb400133j (2013).
69 Chatterjee, A. & O'Keefe, C. Current controversies in the USA regarding vaccine safety. *Expert Rev Vaccines* **9**, 497-502, doi:10.1586/erv.10.36 (2010).
70 Grandjean, P. & Landrigan, P. J. Developmental neurotoxicity of industrial chemicals. *Lancet* **368**, 2167-2178, doi:10.1016/S0140-6736(06)69665-7 (2006).
71 Grandjean, P. & Landrigan, P. J. Neurobehavioural effects of developmental toxicity. *Lancet Neurol* **13**, 330-338, doi:10.1016/S1474-4422(13)70278-3 (2014).
72 Hung, L. W. *et al.* Gating of social reward by oxytocin in the ventral tegmental area. *Science* **357**, 1406-1411, doi:10.1126/science.aan4994 (2017).
73 Hong, W., Kim, D. W. & Anderson, D. J. Antagonistic control of social versus repetitive self-grooming behaviors by separable amygdala neuronal subsets. *Cell* **158**, 1348-1361, doi:10.1016/j.cell.2014.07.049 (2014).
74 Leroy, F. *et al.* A circuit from hippocampal CA2 to lateral septum disinhibits social aggression. *Nature* **564**, 213-218, doi:10.1038/s41586-018-0772-0 (2018).
75 Remedios, R. *et al.* Social behaviour shapes hypothalamic neural ensemble representations of conspecific sex. *Nature* **550**, 388-392, doi:10.1038/nature23885 (2017).
76 Li, Y. *et al.* Neuronal Representation of Social Information in the Medial Amygdala of Awake Behaving Mice. *Cell* **171**, 1176-1190 e1117, doi:10.1016/j.cell.2017.10.015 (2017).
77 Allsop, S. A. *et al.* Corticoamygdala Transfer of Socially Derived Information Gates Observational Learning. *Cell* **173**, 1329-1342 e1318, doi:10.1016/j.cell.2018.04.004 (2018).
78 Hitti, F. L. & Siegelbaum, S. A. The hippocampal CA2 region is essential for social memory. *Nature* **508**, 88-92, doi:10.1038/nature13028 (2014).
79 Padberg, S. *et al.* Observational cohort study of pregnancy outcome after first-trimester exposure to fluoroquinolones. *Antimicrob Agents Chemother* **58**, 4392-4398, doi:10.1128/AAC.02413-14 (2014).

80 Yefet, E. *et al.* The safety of quinolones in pregnancy. *Obstet Gynecol Surv* **69**, 681-694, doi:10.1097/OGX.0000000000000122 (2014).
81 Stein, T. P., Schluter, M. D., Steer, R. A., Guo, L. & Ming, X. Bisphenol A Exposure in Children With Autism Spectrum Disorders. *Autism Res* **8**, 272-283, doi:10.1002/aur.1444 (2015).
82 Yu, C. *et al.* Pubertal exposure to bisphenol A disrupts behavior in adult C57BL/6J mice. *Environ Toxicol Pharmacol* **31**, 88-99, doi:10.1016/j.etap.2010.09.009 (2011).
83 Lyall, K. *et al.* Polychlorinated Biphenyl and Organochlorine Pesticide Concentrations in Maternal Mid-Pregnancy Serum Samples: Association with Autism Spectrum Disorder and Intellectual Disability. *Environ Health Perspect* **125**, 474-480, doi:10.1289/EHP277 (2017).
84 Jolous-Jamshidi, B., Cromwell, H. C., McFarland, A. M. & Meserve, L. A. Perinatal exposure to polychlorinated biphenyls alters social behaviors in rats. *Toxicol Lett* **199**, 136-143, doi:10.1016/j.toxlet.2010.08.015 (2010).
85 Panesar, H. K., Kennedy, C. L., Keil Stietz, K. P. & Lein, P. J. Polychlorinated Biphenyls (PCBs): Risk Factors for Autism Spectrum Disorder? *Toxics* **8**, doi:10.3390/toxics8030070 (2020).
86 von Ehrenstein, O. S. *et al.* Prenatal and infant exposure to ambient pesticides and autism spectrum disorder in children: population based case-control study. *BMJ* **364**, l962, doi:10.1136/bmj.l962 (2019).
87 Lan, A., Kalimian, M., Amram, B. & Kofman, O. Prenatal chlorpyrifos leads to autism-like deficits in C57Bl6/J mice. *Environ Health* **16**, 43, doi:10.1186/s12940-017-0251-3 (2017).
88 Berg, E. L. *et al.* Translational outcomes relevant to neurodevelopmental disorders following early life exposure of rats to chlorpyrifos. *J Neurodev Disord* **12**, 40, doi:10.1186/s11689-020-09342-1 (2020).
89 Mizushina, Y., Shiomi, K., Kuriyama, I., Takahashi, Y. & Yoshida, H. Inhibitory effects of a major soy isoflavone, genistein, on human DNA topoisomerase II activity and cancer cell proliferation. *Int J Oncol* **43**, 1117-1124, doi:10.3892/ijo.2013.2032 (2013).
90 Bandele, O. J. & Osheroff, N. Bioflavonoids as poisons of human topoisomerase II alpha and II beta. *Biochemistry* **46**, 6097-6108, doi:10.1021/bi7000664 (2007).
91 Marshall, B. L. *et al.* Early genistein exposure of California mice and effects on the gut microbiota-brain axis. *J Endocrinol* **242**, 139-157, doi:10.1530/JOE-19-0214 (2019).
92 Marko, D. & Boege, F. A possible link between nutritional uptake of ubiquitous topoisomerase inhibitors and autism? *Int J Dev Neurosci* **53**, 8-9, doi:10.1016/j.ijdevneu.2016.06.003 (2016).
93 Habermeyer, M. *et al.* Anthocyanidins modulate the activity of human DNA topoisomerases I and II and affect cellular DNA integrity. *Chem Res Toxicol* **18**, 1395-1404, doi:10.1021/tx050039n (2005).
94 Fehr, M. *et al.* Alternariol acts as a topoisomerase poison, preferentially affecting the IIalpha isoform. *Mol Nutr Food Res* **53**, 441-451, doi:10.1002/mnfr.200700379 (2009).
95 Zhang, Z. *et al.* Ginkgo biloba leaf extract induces DNA damage by inhibiting topoisomerase II activity in human hepatic cells. *Sci Rep* **5**, 14633, doi:10.1038/srep14633 (2015).
96 Choi, J. Y., Seo, C. S., Zheng, M. S., Lee, C. S. & Son, J. K. Topoisomerase I and II inhibitory constituents from the bark of Tilia amurensis. *Arch Pharm Res* **31**, 1413-1418, doi:10.1007/s12272-001-2125-y (2008).
97 Li, G. *et al.* Lignans from the bark of Machilus thunbergii and their DNA topoisomerases I and II inhibition and cytotoxicity. *Biol Pharm Bull* **27**, 1147-1150, doi:10.1248/bpb.27.1147 (2004).
98 Lee, Y. K. *et al.* Inhibition of DNA topoisomerases I and II and cytotoxicity by lignans from Saururus chinensis. *Arch Pharm Res* **32**, 1409-1415, doi:10.1007/s12272-009-2010-7 (2009).
99 Riaz, L. *et al.* Fluoroquinolones (FQs) in the environment: A review on their abundance, sorption and toxicity in soil. *Chemosphere* **191**, 704-720, doi:10.1016/j.chemosphere.2017.10.092 (2018).

100 Janecko, N., Pokludova, L., Blahova, J., Svobodova, Z. & Literak, I. Implications of fluoroquinolone contamination for the aquatic environment-A review. *Environ Toxicol Chem* **35**, 2647-2656, doi:10.1002/etc.3552 (2016).
101 Misik, M. *et al.* Environmental risk assessment of widely used anticancer drugs (5-fluorouracil, cisplatin, etoposide, imatinib mesylate). *Water Res* **164**, 114953, doi:10.1016/j.watres.2019.114953 (2019).
102 Heath, E., Filipic, M., Kosjek, T. & Isidori, M. Fate and effects of the residues of anticancer drugs in the environment. *Environ Sci Pollut Res Int* **23**, 14687-14691, doi:10.1007/s11356-016-7069-3 (2016).
103 Chander, V. *et al.* Pharmaceutical Compounds in Drinking Water. *J Xenobiot* **6**, 5774, doi:10.4081/xeno.2016.5774 (2016).
104 RK, C. Y. *et al.* Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nat Neurosci* **20**, 602-611, doi:10.1038/nn.4524 (2017).
105 Iossifov, I. *et al.* De novo gene disruptions in children on the autistic spectrum. *Neuron* **74**, 285-299, doi:10.1016/j.neuron.2012.04.009 (2012).
106 Lopes, G. *et al.* Bonsai: an event-based framework for processing and controlling data streams. *Front Neuroinform* **9**, 7, doi:10.3389/fninf.2015.00007 (2015).
107 Milan, D. J., Peterson, T. A., Ruskin, J. N., Peterson, R. T. & MacRae, C. A. Drugs that induce repolarization abnormalities cause bradycardia in zebrafish. *Circulation* **107**, 1355-1358 (2003).
108 Peterson, R. T. & Fishman, M. C. Discovery and use of small molecules for probing biological processes in zebrafish. *Methods Cell Biol* **76**, 569-591 (2004).
109 Hwang, W. Y. *et al.* Targeted Mutagenesis in Zebrafish Using CRISPR RNA-Guided Nucleases. *Methods Mol Biol* **1311**, 317-334, doi:10.1007/978-1-4939-2687-9_21 (2015).
110 Heyne, G. W. *et al.* A Simple and Reliable Method for Early Pregnancy Detection in Inbred Mice. *J Am Assoc Lab Anim Sci* **54**, 368-371 (2015).
111 Friard O., G. M., Fitzjohn R. BORIS: a free, versatile open‐source event‐logging software for video/audio coding and live observations. *Methods Ecol. Evol.* **7**, 1325–1330, doi:10.1111/2041-210X.12584 (2016).
112 Deacon, R. M. & Rawlins, J. N. T-maze alternation in the rodent. *Nat Protoc* **1**, 7-12, doi:10.1038/nprot.2006.2 (2006).
113 Geyer, M. A. & Paulus, M. P. Multivariate and nonlinear approaches to characterizing drug effects on the locomotor and investigatory behavior of rats. *NIDA Res Monogr* **124**, 203-235 (1992).
114 Tambaro, S., Tomasi, M. L. & Bortolato, M. Long-term CB(1) receptor blockade enhances vulnerability to anxiogenic-like effects of cannabinoids. *Neuropharmacology* **70**, 268-277, doi:10.1016/j.neuropharm.2013.02.009 (2013).
115 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).
116 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).
117 Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z. & Zhang, B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res* **47**, W199-W205, doi:10.1093/nar/gkz401 (2019).
118 Durinck, S. *et al.* BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* **21**, 3439-3440, doi:10.1093/bioinformatics/bti525 (2005).
119 Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc* **4**, 1184-1191, doi:10.1038/nprot.2009.97 (2009).

120 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).
121 Mootha, V. K. *et al.* PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* **34**, 267-273, doi:10.1038/ng1180 (2003).
122 Lawrence, M. *et al.* Software for computing and annotating genomic ranges. *PLoS Comput Biol* **9**, e1003118, doi:10.1371/journal.pcbi.1003118 (2013).
123 Jaffe, A. E. *et al.* Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. *Int J Epidemiol* **41**, 200-209, doi:10.1093/ije/dyr238 (2012).
124 Aryee, M. J. *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363-1369, doi:10.1093/bioinformatics/btu049 (2014).
125 Yu, G., Wang, L. G. & He, Q. Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382-2383, doi:10.1093/bioinformatics/btv145 (2015).

Figure 1

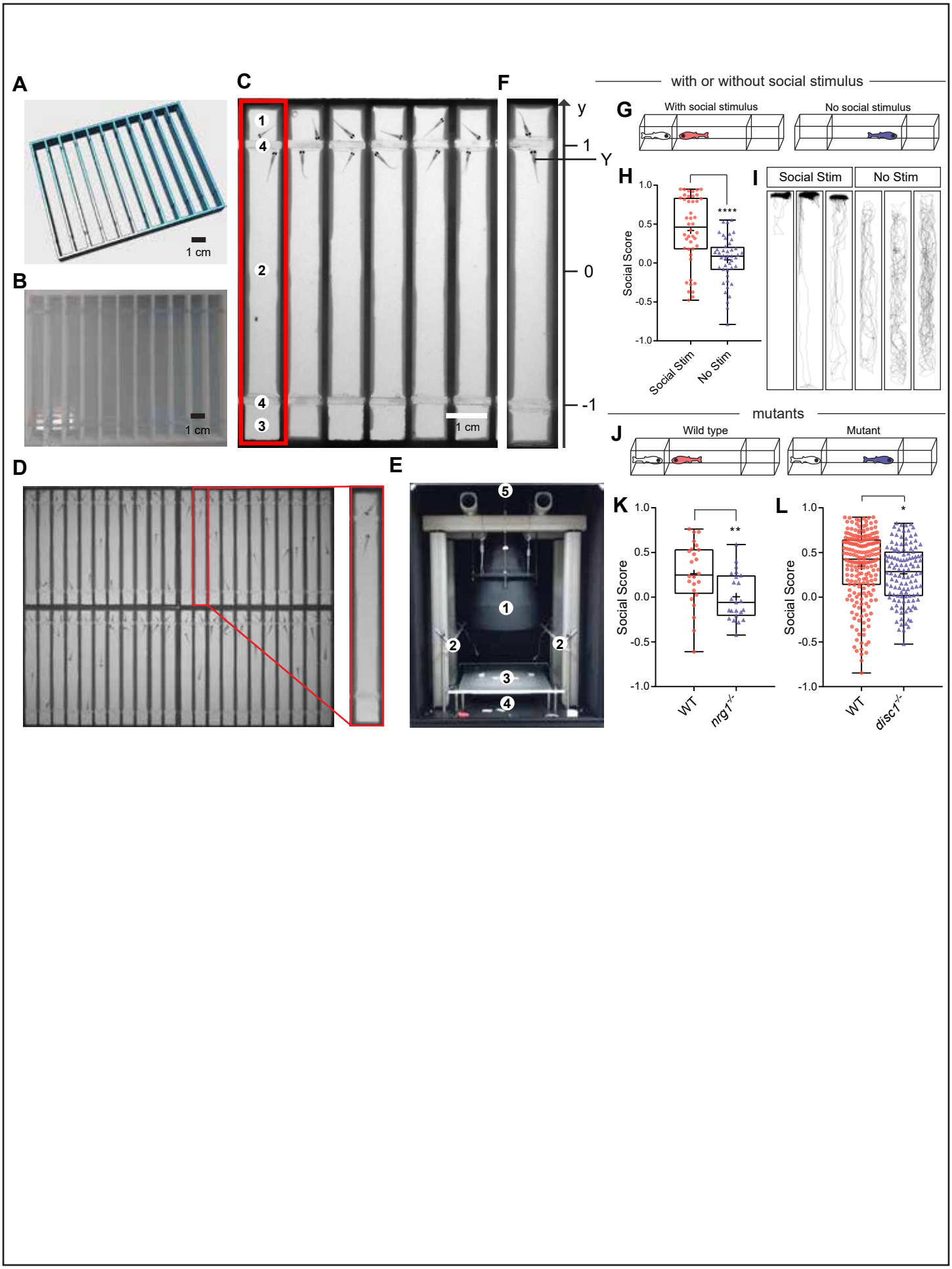

A
1 cm
B
1 cm
C
1
4
2
4
3
1 cm
F
y
1
Y
0
-1
with or without social stimulus
G
With social stimulus
No social stimulus
H
Social Score
1.0
0.5
0.0
-0.5
-1.0
****
Social Stim
No Stim
I
Social Stim
No Stim
mutants
J
Wild type
Mutant
D
E
5
1
2
2
3
4
K
Social Score
**
WT
nrg1-/-
L
Social Score
*
WT
disc1-/-

Figure 2

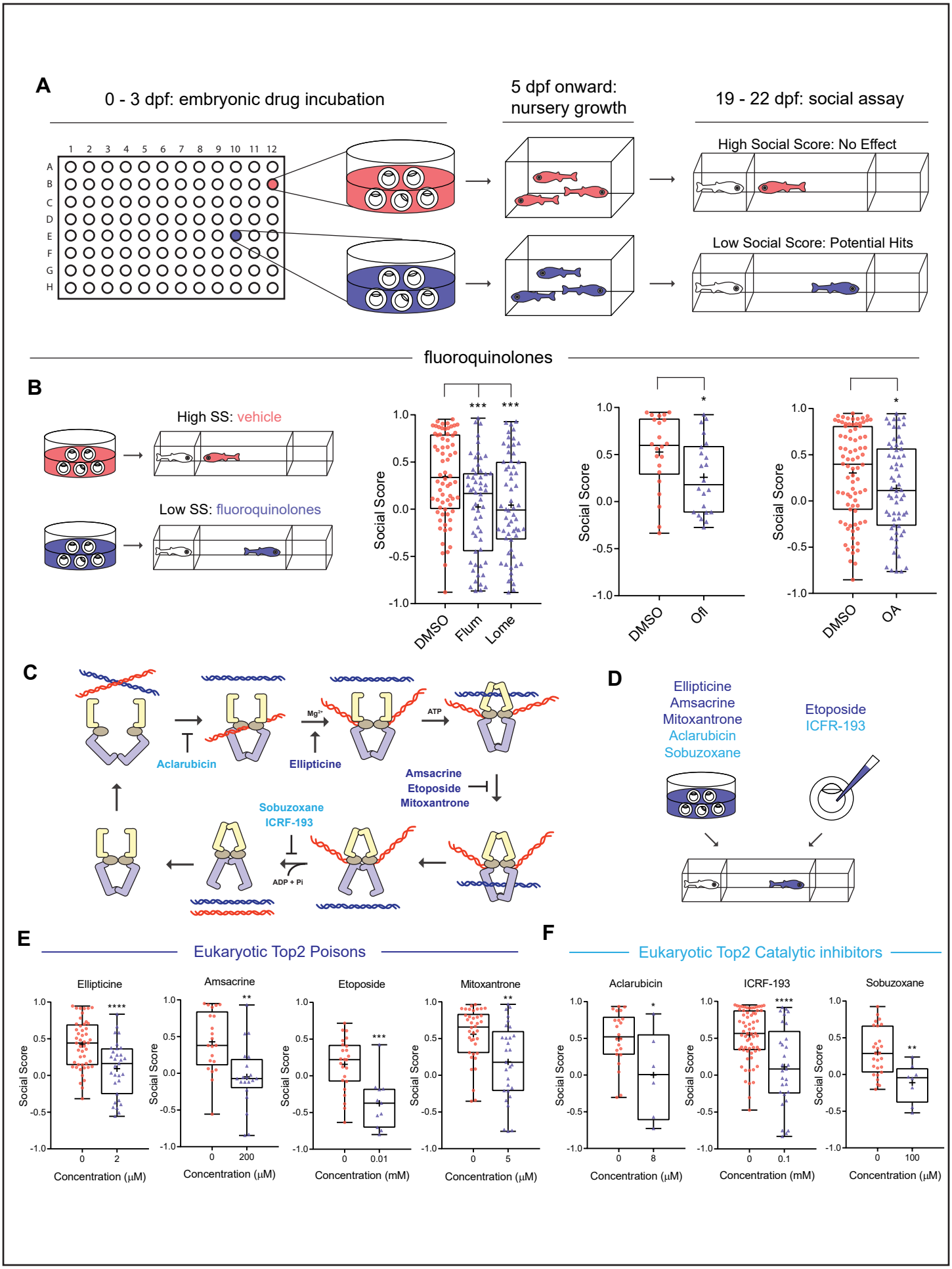

A
0 - 3 dpf: embryonic drug incubation
5 dpf onward: nursery growth
19 - 22 dpf: social assay
High Social Score: No Effect
Low Social Score: Potential Hits
fluoroquinolones
B
High SS: vehicle
Low SS: fluoroquinolones
Social Score
DMSO
Flum
Lome
Ofl
OA
C
Aclarubicin
Ellipticine
Amsacrine
Etoposide
Mitoxantrone
Sobuzoxane
ICRF-193
ADP + Pi
ATP
D
Ellipticine
Amsacrine
Mitoxantrone
Aclarubicin
Sobuzoxane
Etoposide
ICRF-193
E
Eukaryotic Top2 Poisons
Ellipticine
Amsacrine
Etoposide
Mitoxantrone
Concentration (μM)
Concentration (mM)
F
Eukaryotic Top2 Catalytic inhibitors
Aclarubicin
ICRF-193
Sobuzoxane

Figure 3

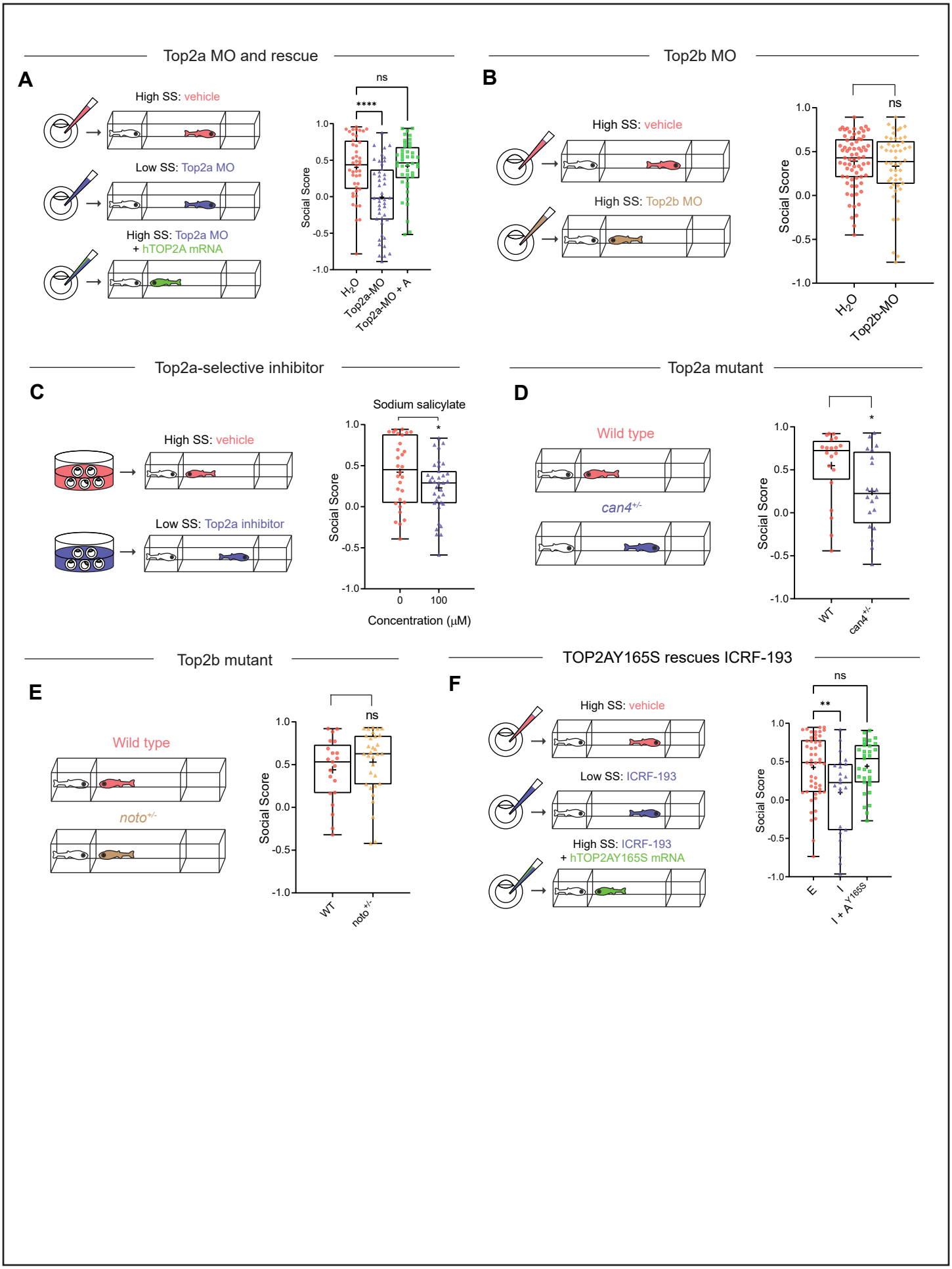

Top2a MO and rescue
A
High SS: vehicle
Low SS: Top2a MO
High SS: Top2a MO
+ hTOP2A mRNA
ns
****
Social Score
1.0
0.5
0.0
-0.5
-1.0
H2O
Top2a-MO
Top2a-MO + A
Top2b MO
B
High SS: vehicle
High SS: Top2b MO
ns
Social Score
H2O
Top2b-MO
Top2a-selective inhibitor
C
High SS: vehicle
Low SS: Top2a inhibitor
Sodium salicylate
*
Social Score
0
100
Concentration (μM)
Top2a mutant
D
Wild type
can4+/-
*
Social Score
WT
can4+/-
Top2b mutant
E
Wild type
noto+/-
ns
Social Score
WT
noto+/-
TOP2AY165S rescues ICRF-193
F
High SS: vehicle
Low SS: ICRF-193
High SS: ICRF-193
+ hTOP2AY165S mRNA
ns
**
Social Score
E
I
I + A Y165S

Figure 4

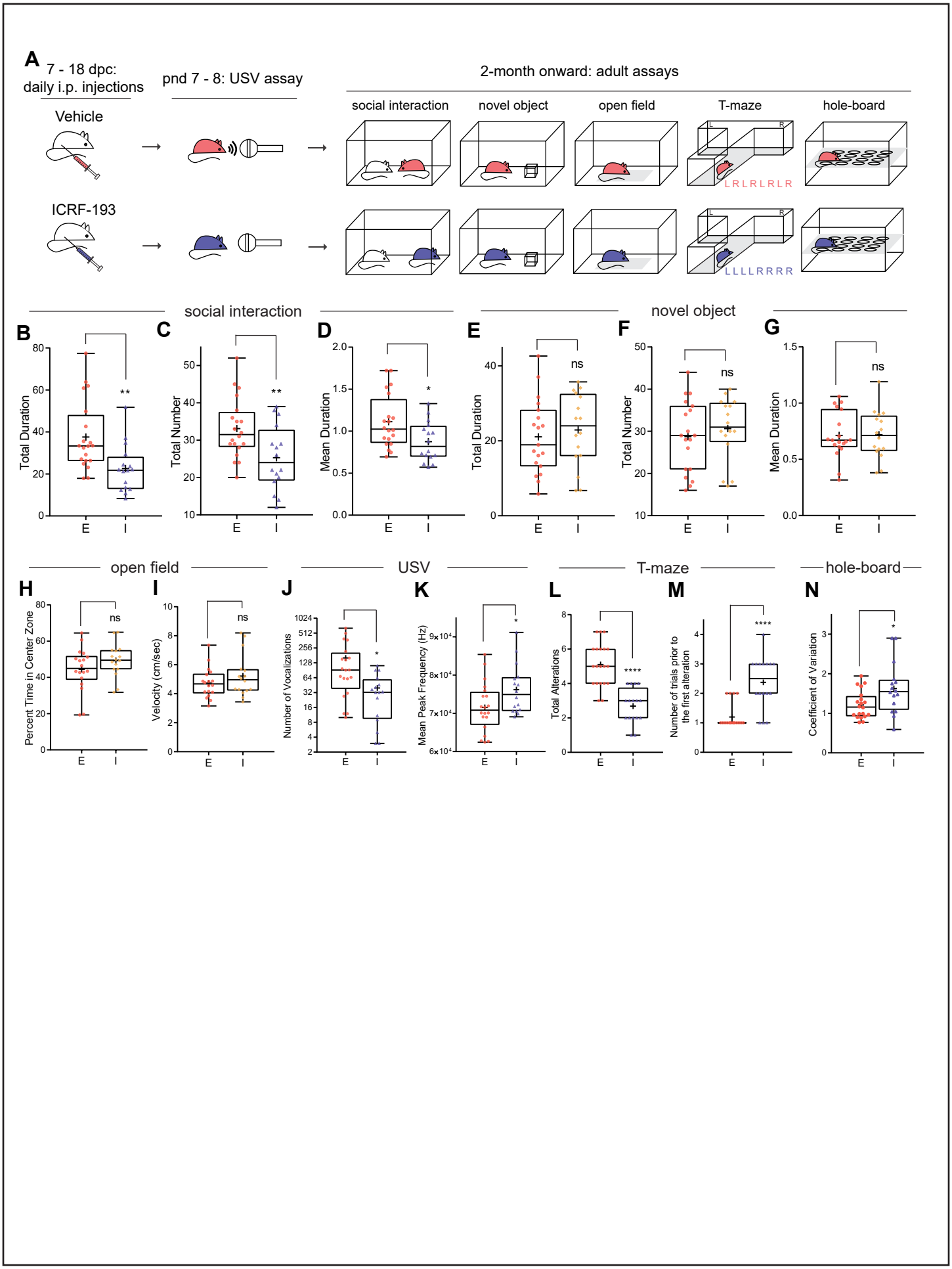

A
7 - 18 dpc:
daily i.p. injections
pnd 7 - 8: USV assay
2-month onward: adult assays
social interaction
novel object
open field
T-maze
hole-board
Vehicle
ICRF-193
L R L R L R L R
L L L L R R R R
social interaction
novel object
B
Total Duration
C
Total Number
D
Mean Duration
E
Total Duration
F
Total Number
G
Mean Duration
E
I
**
*
ns
open field
USV
T-maze
hole-board
H
Percent Time in Center Zone
I
Velocity (cm/sec)
J
Number of Vocalizations
K
Mean Peak Frequency (Hz)
L
Total Alterations
M
Number of trials prior to the first alteration
N
Coefficient of Variation
****

Figure 5

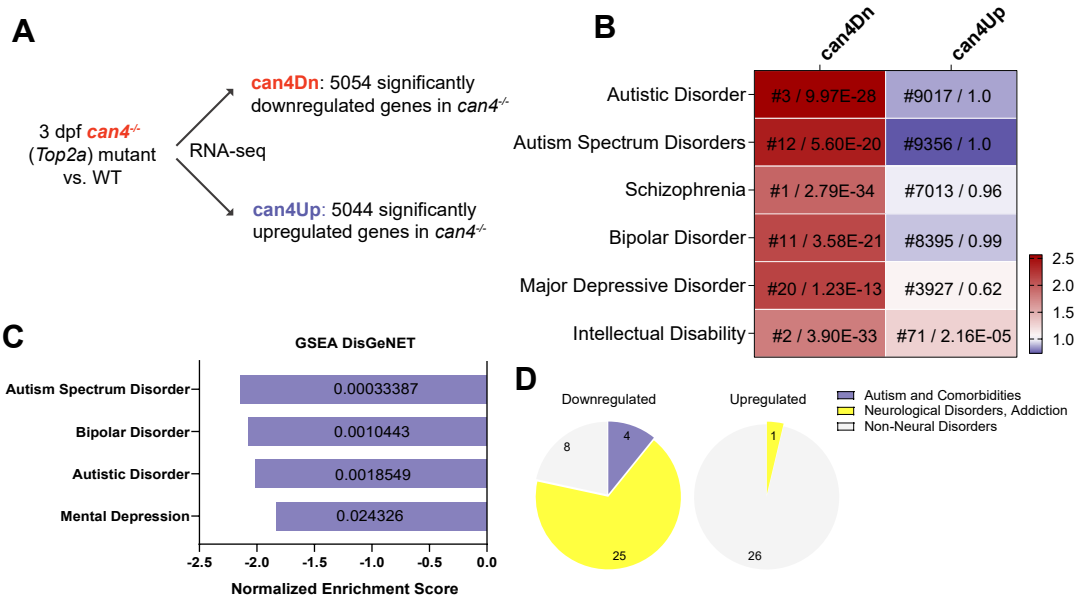

A
3 dpf can4-/- (Top2a) mutant vs. WT
RNA-seq
can4Dn: 5054 significantly downregulated genes in can4-/-
can4Up: 5044 significantly upregulated genes in can4-/-
B
can4Dn
can4Up
Autistic Disorder
Autism Spectrum Disorders
Schizophrenia
Bipolar Disorder
Major Depressive Disorder
Intellectual Disability
#3 / 9.97E-28
#9017 / 1.0
#12 / 5.60E-20
#9356 / 1.0
#1 / 2.79E-34
#7013 / 0.96
#11 / 3.58E-21
#8395 / 0.99
#20 / 1.23E-13
#3927 / 0.62
#2 / 3.90E-33
#71 / 2.16E-05
2.5
2.0
1.5
1.0
C
GSEA DisGeNET
Autism Spectrum Disorder
Bipolar Disorder
Autistic Disorder
Mental Depression
0.00033387
0.0010443
0.0018549
0.024326
-2.5
-2.0
-1.5
-1.0
-0.5
0.0
Normalized Enrichment Score
D
Downregulated
Upregulated
Autism and Comorbidities
Neurological Disorders, Addiction
Non-Neural Disorders
4
8
25
1
26

Figure 6

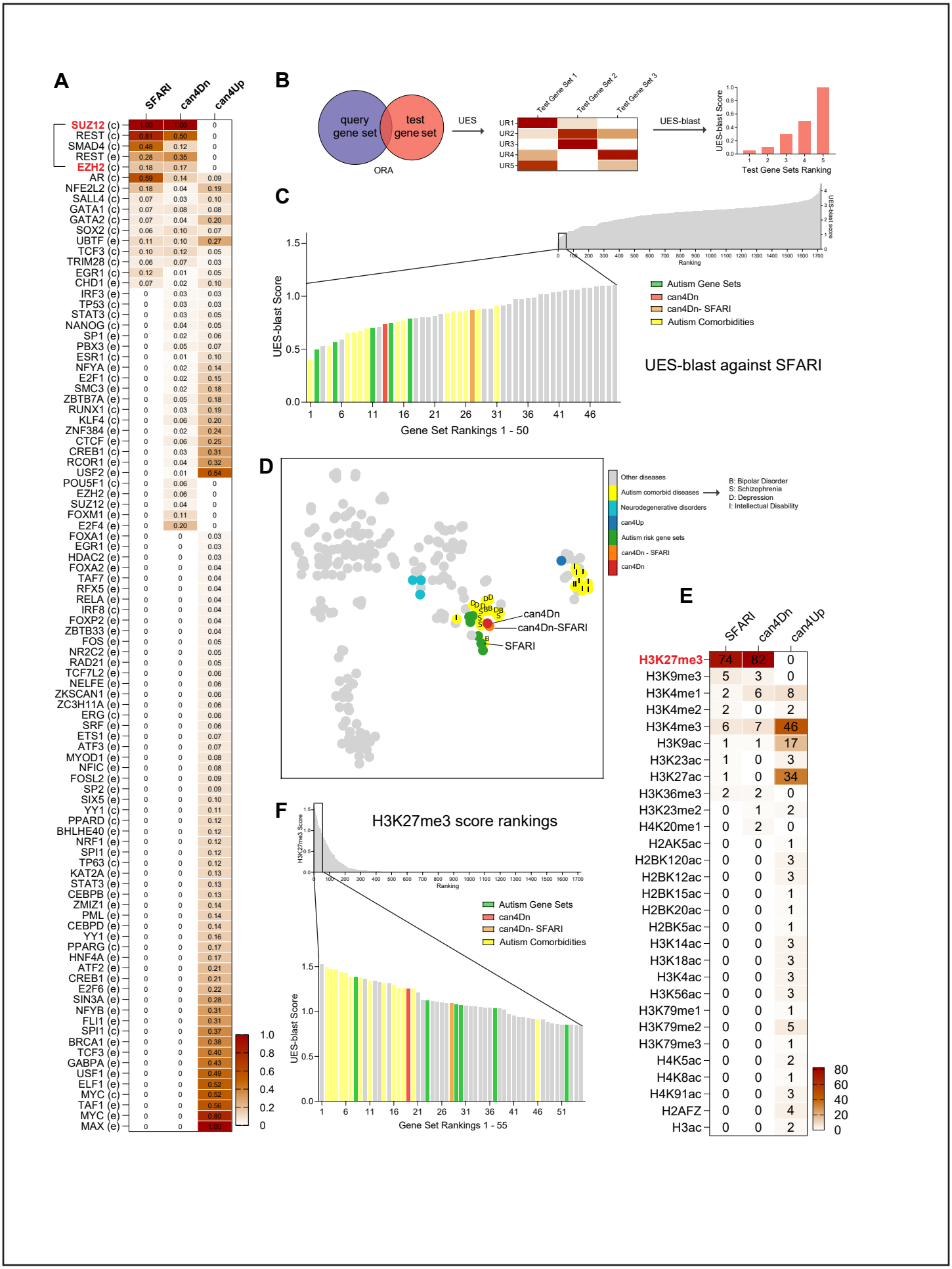
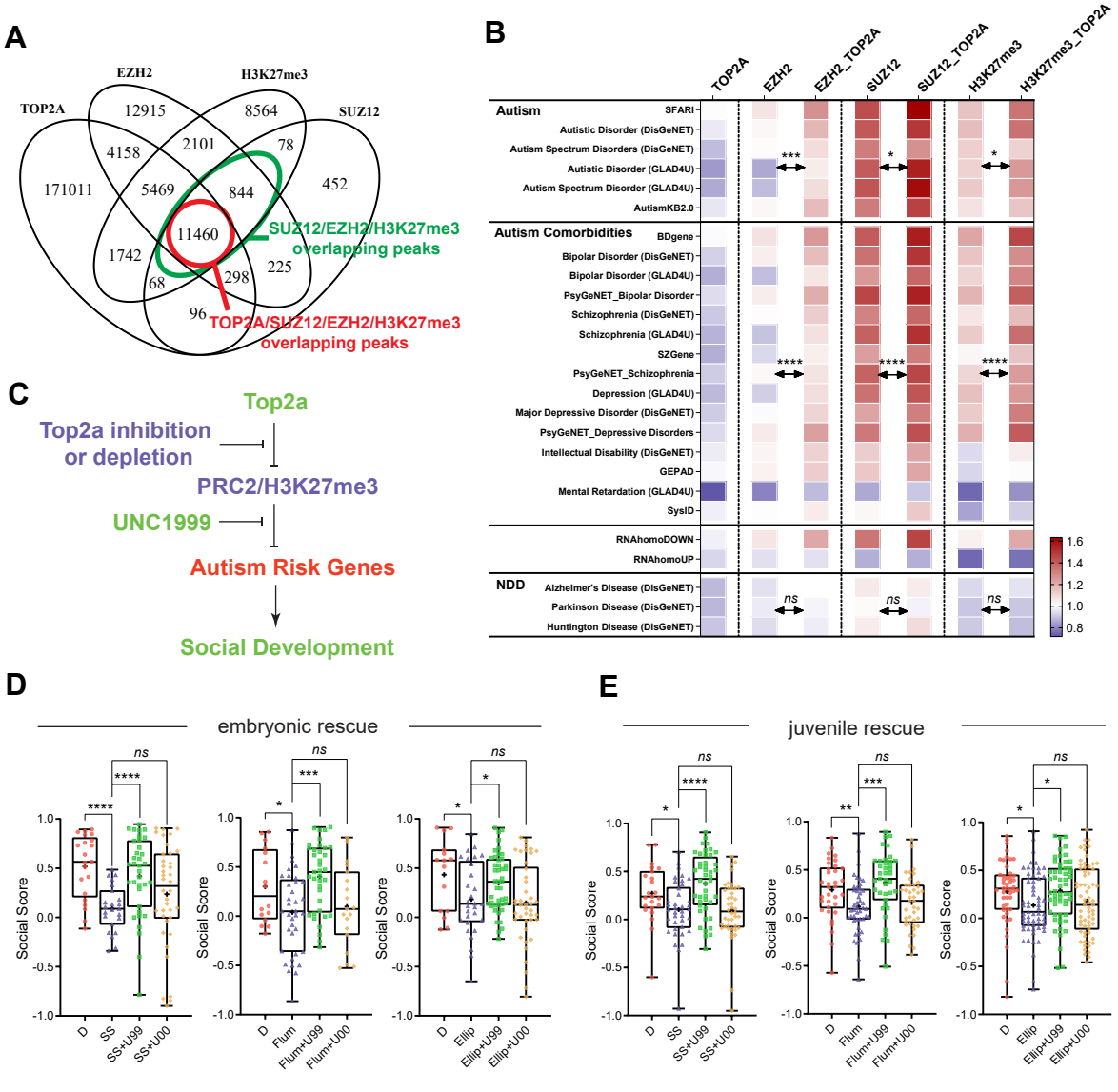
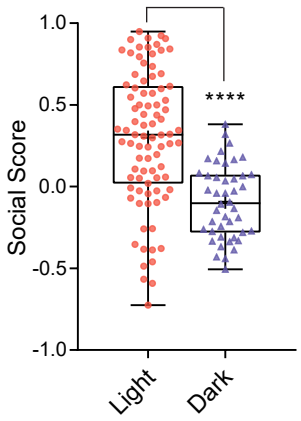

A
SFARI
can4Dn
can4Up
SUZ12 (c)
REST (c)
SMAD4 (c)
REST (e)
EZH2 (c)
AR (c)
NFE2L2 (c)
SALL4 (c)
GATA1 (c)
GATA2 (c)
SOX2 (c)
UBTF (e)
TCF3 (c)
TRIM28 (c)
EGR1 (c)
CHD1 (e)
IRF3 (e)
TP53 (c)
STAT3 (c)
NANOG (c)
SP1 (e)
PBX3 (e)
ESR1 (c)
NFYA (e)
E2F1 (c)
SMC3 (e)
ZBTB7A (e)
RUNX1 (c)
KLF4 (c)
ZNF384 (e)
CTCF (e)
CREB1 (c)
RCOR1 (e)
USF2 (e)
POU5F1 (c)
EZH2 (e)
SUZ12 (e)
FOXM1 (e)
E2F4 (e)
FOXA1 (e)
EGR1 (e)
HDAC2 (e)
FOXA2 (e)
TAF7 (e)
RFX5 (e)
RELA (e)
IRF8 (c)
FOXP2 (e)
ZBTB33 (e)
FOS (e)
NR2C2 (e)
RAD21 (e)
TCF7L2 (e)
NELFE (e)
ZKSCAN1 (e)
ZC3H11A (e)
ERG (c)
SRF (e)
ETS1 (e)
ATF3 (e)
MYOD1 (e)
NFIC (e)
FOSL2 (e)
SP2 (e)
SIX5 (e)
YY1 (c)
PPARD (c)
BHLHE40 (e)
NRF1 (e)
SPI1 (e)
TP63 (c)
KAT2A (e)
STAT3 (e)
CEBPB (e)
ZMIZ1 (e)
PML (e)
CEBPD (e)
YY1 (e)
PPARG (c)
HNF4A (e)
ATF2 (e)
CREB1 (e)
E2F6 (e)
SIN3A (e)
NFYB (e)
FLI1 (e)
SPI1 (c)
BRCA1 (e)
TCF3 (e)
GABPA (e)
USF1 (e)
ELF1 (e)
MYC (c)
TAF1 (e)
MYC (e)
MAX (e)
B
query gene set
test gene set
ORA
UES
Test Gene Set 1
Test Gene Set 2
Test Gene Set 3
UR1
UR2
UR3
UR4
UR5
UES-blast
UES-blast Score
Test Gene Sets Ranking
C
UES-blast Score
Gene Set Rankings 1 - 50
Ranking
Autism Gene Sets
can4Dn
can4Dn- SFARI
Autism Comorbidities
UES-blast against SFARI
D
Other diseases
Autism comorbid diseases
Neurodegenerative disorders
can4Up
Autism risk gene sets
can4Dn - SFARI
can4Dn
B: Bipolar Disorder
S: Schizophrenia
D: Depression
I: Intellectual Disability
can4Dn
can4Dn-SFARI
SFARI
E
SFARI
can4Dn
can4Up
H3K27me3
H3K9me3
H3K4me1
H3K4me2
H3K4me3
H3K9ac
H3K23ac
H3K27ac
H3K36me3
H3K23me2
H4K20me1
H2AK5ac
H2BK120ac
H2BK12ac
H2BK15ac
H2BK20ac
H2BK5ac
H3K14ac
H3K18ac
H3K4ac
H3K56ac
H3K79me1
H3K79me2
H3K79me3
H4K5ac
H4K8ac
H4K91ac
H2AFZ
H3ac
F
H3K27me3 score rankings
H3K27me3 Score
Ranking
Autism Gene Sets
can4Dn
can4Dn- SFARI
Autism Comorbidities
UES-blast Score
Gene Set Rankings 1 - 55

Figure 7

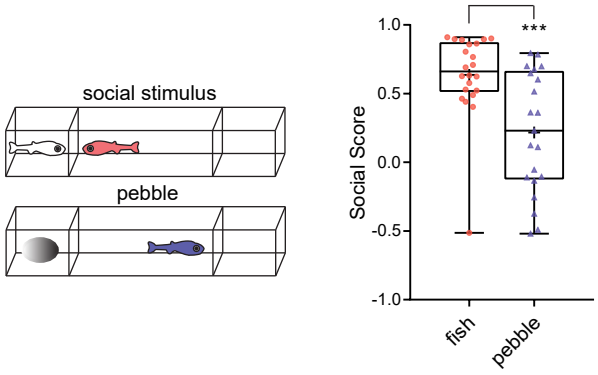

A
EZH2
H3K27me3
TOP2A
SUZ12
12915
8564
4158
2101
78
171011
5469
844
452
11460
1742
225
68
298
96
SUZ12/EZH2/H3K27me3 overlapping peaks
TOP2A/SUZ12/EZH2/H3K27me3 overlapping peaks
B
TOP2A
EZH2
EZH2_TOP2A
SUZ12
SUZ12_TOP2A
H3K27me3
H3K27me3_TOP2A
Autism
SFARI
Autistic Disorder (DisGeNET)
Autism Spectrum Disorders (DisGeNET)
Autistic Disorder (GLAD4U)
Autism Spectrum Disorder (GLAD4U)
AutismKB2.0
Autism Comorbidities
BDgene
Bipolar Disorder (DisGeNET)
Bipolar Disorder (GLAD4U)
PsyGeNET_Bipolar Disorder
Schizophrenia (DisGeNET)
Schizophrenia (GLAD4U)
SZGene
PsyGeNET_Schizophrenia
Depression (GLAD4U)
Major Depressive Disorder (DisGeNET)
PsyGeNET_Depressive Disorders
Intellectual Disability (DisGeNET)
GEPAD
Mental Retardation (GLAD4U)
SysID
RNAhomoDOWN
RNAhomoUP
NDD
Alzheimer's Disease (DisGeNET)
Parkinson Disease (DisGeNET)
Huntington Disease (DisGeNET)
ns
1.6
1.4
1.2
1.0
0.8
C
Top2a
Top2a inhibition or depletion
PRC2/H3K27me3
UNC1999
Autism Risk Genes
Social Development
D
embryonic rescue
Social Score
D
SS
SS+U99
SS+U00
Flum
Flum+U99
Flum+U00
Ellip
Ellip+U99
Ellip+U00
E
juvenile rescue

Supplemental Figure 1

A



B



C

WT DNA sequence:
CAGCCACCAGCACCACTCCCTCGGCCAAGACATCCA
nrg1 KO DNA sequence:
CAGCCA----------GACATCCA

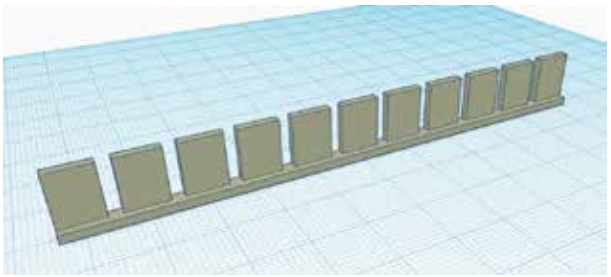
WT protein sequence (residues 259-300):
ATSTTPSAKTSSHVTPCNESEKEYCVNHGKCFTLEVTPGNIR
nrg1 KO protein sequence (predicted):
ARHPVTStop

D

WT DNA sequence:
CAGACGGCCCGGATACATGCGGTCTGAGCCGATTAACCAGCT
disc1 KO DNA sequence:
CAGACGGCCC----------GATTAACCAGCT

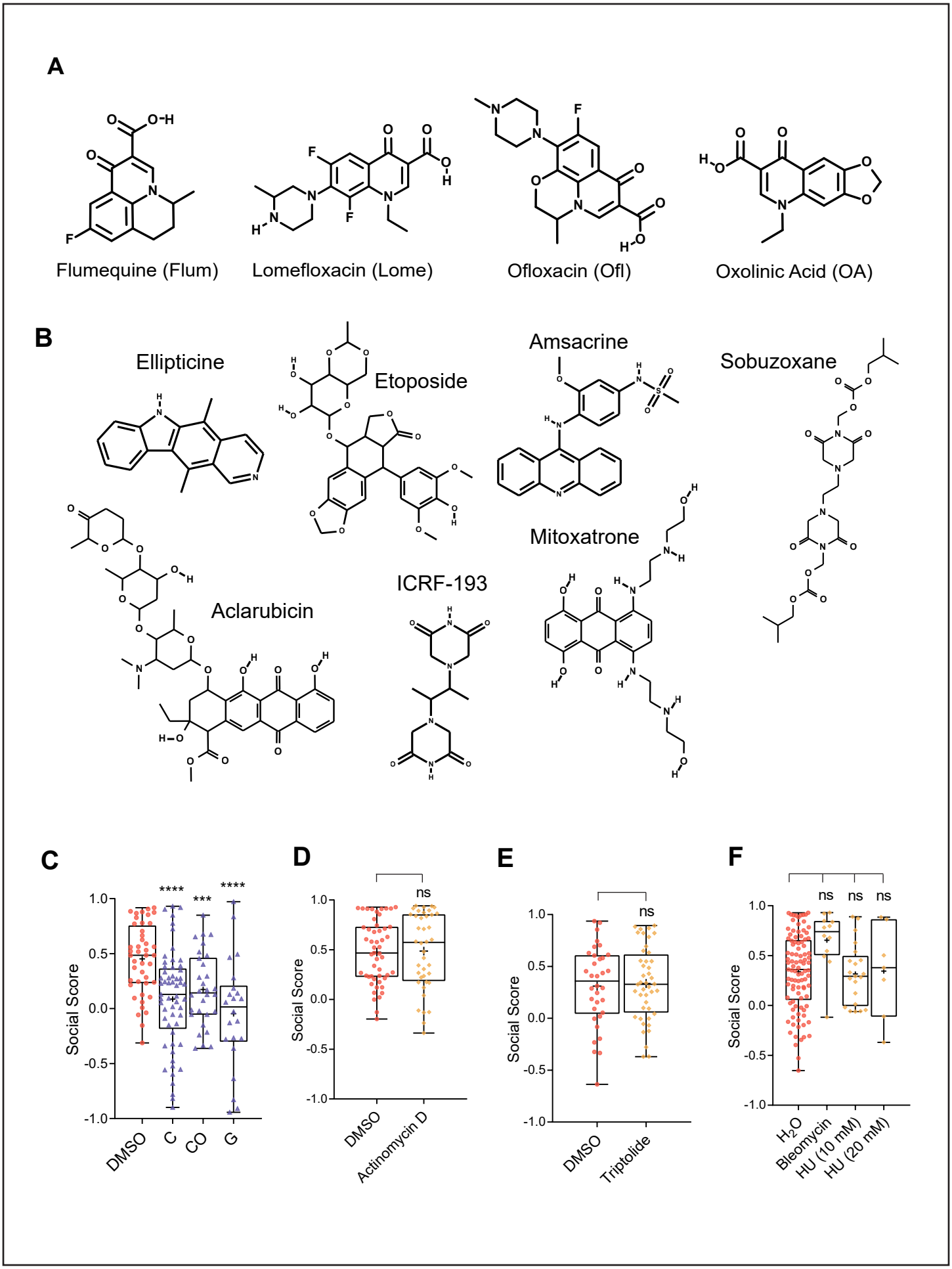
WT protein sequence (residues 32-92):
GVNPSGHRRRSFRRPGYMRSEPINQLDVAETSCDSEHHRSPISKSPAVEN
disc1 KO protein sequence (predicted):
GVNPSGHRRRSFRRPDStop

E

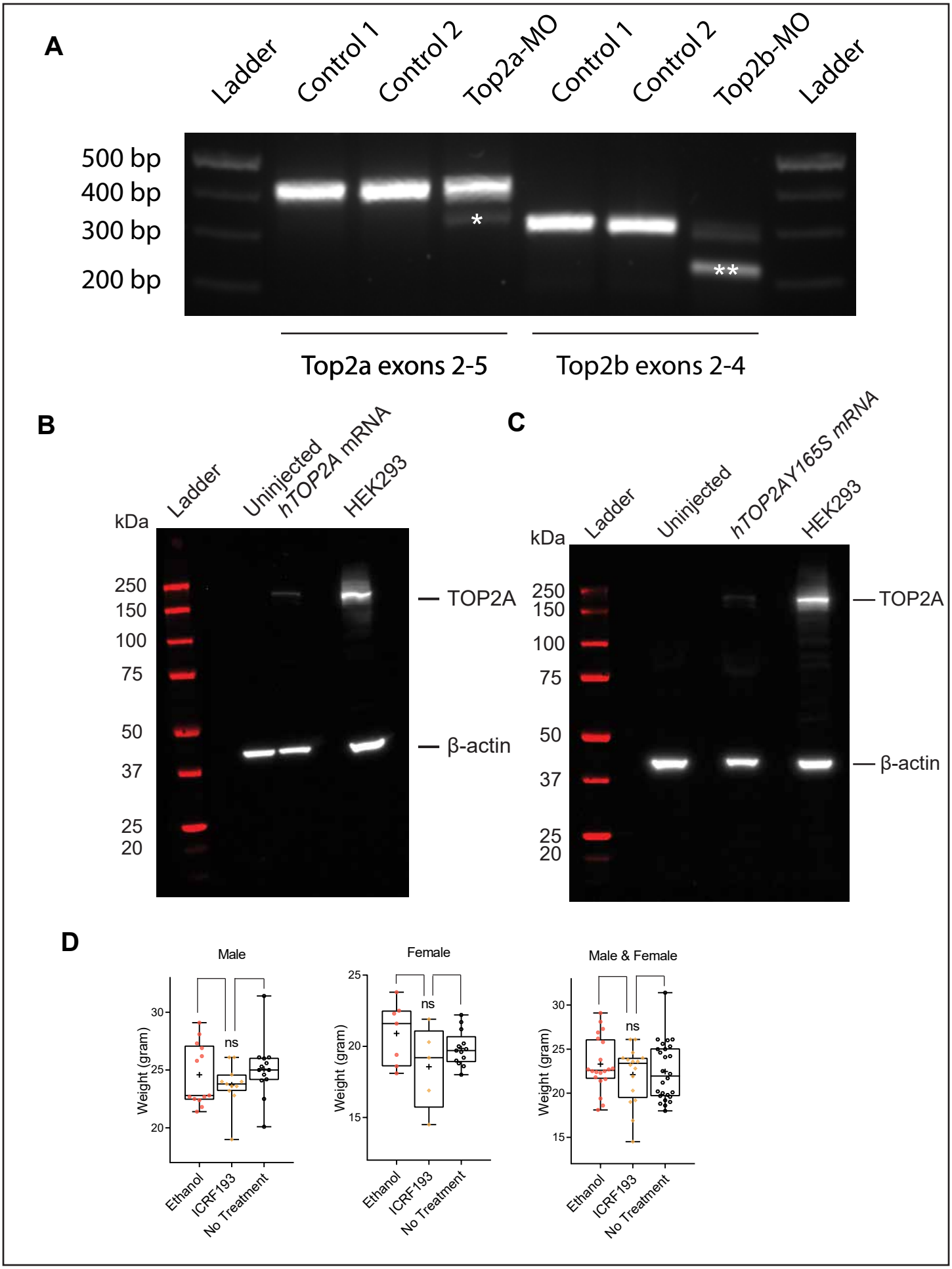


F

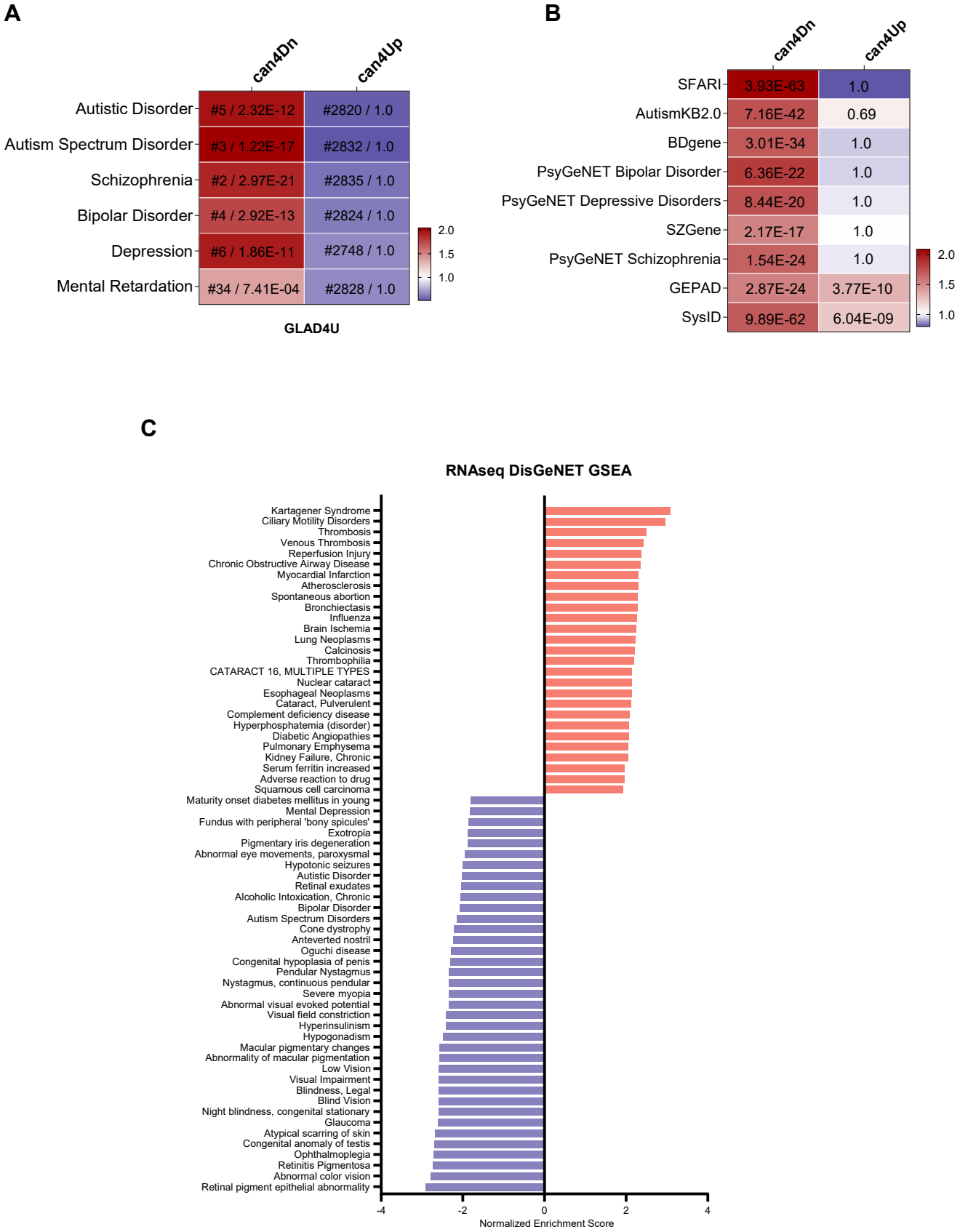
Supplemental Figure 2



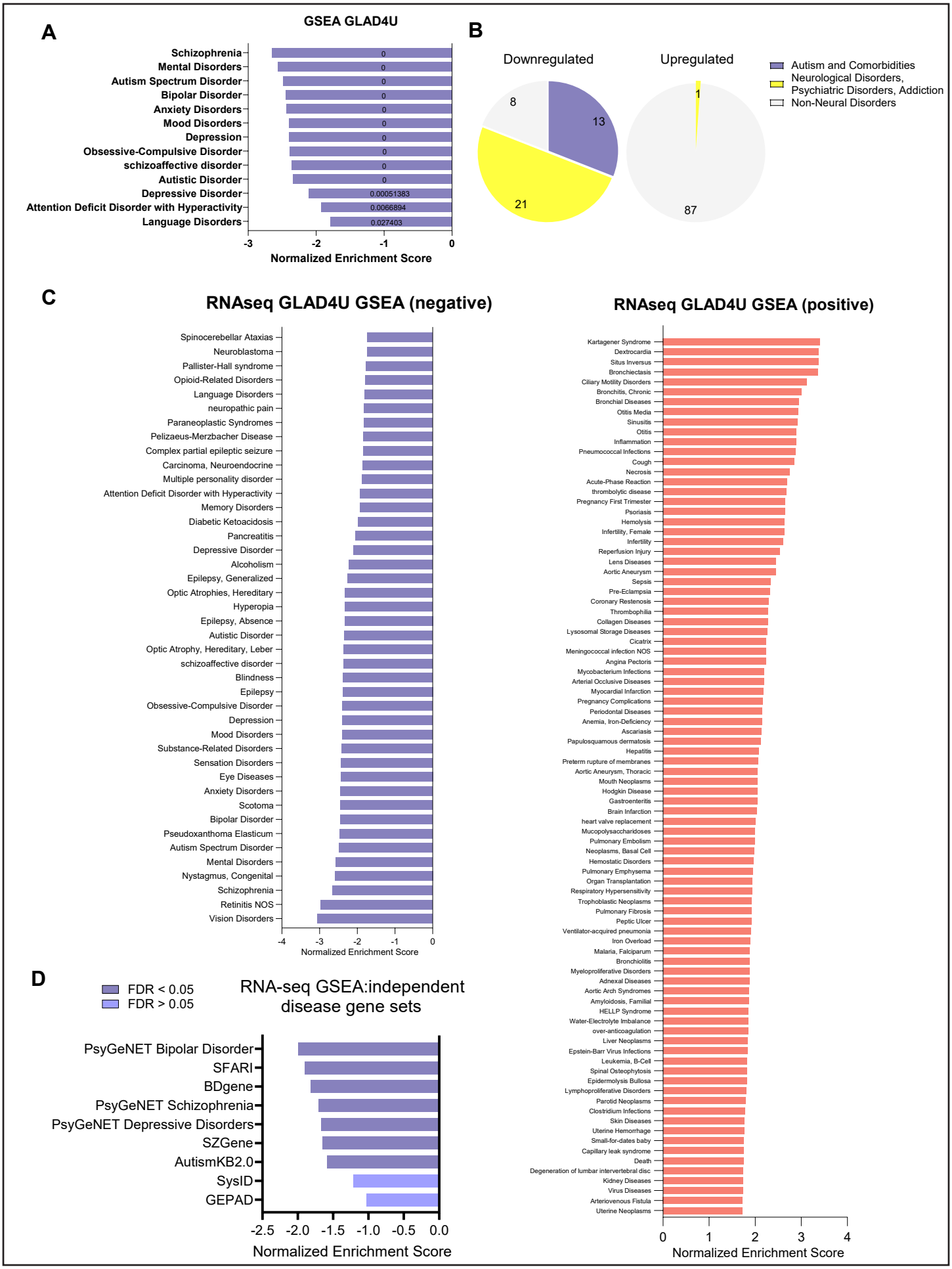
Supplemental Figure 3



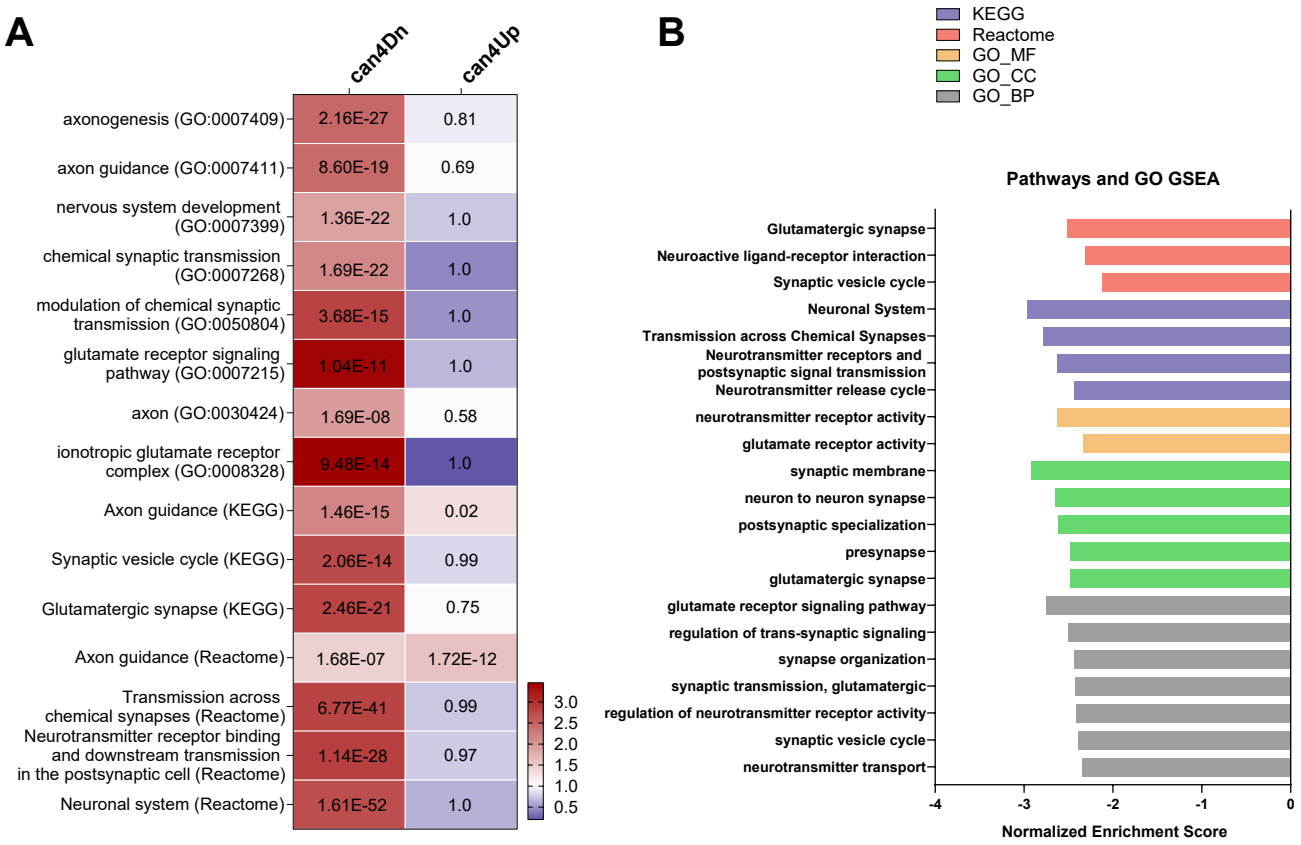
Supplemental Figure 4



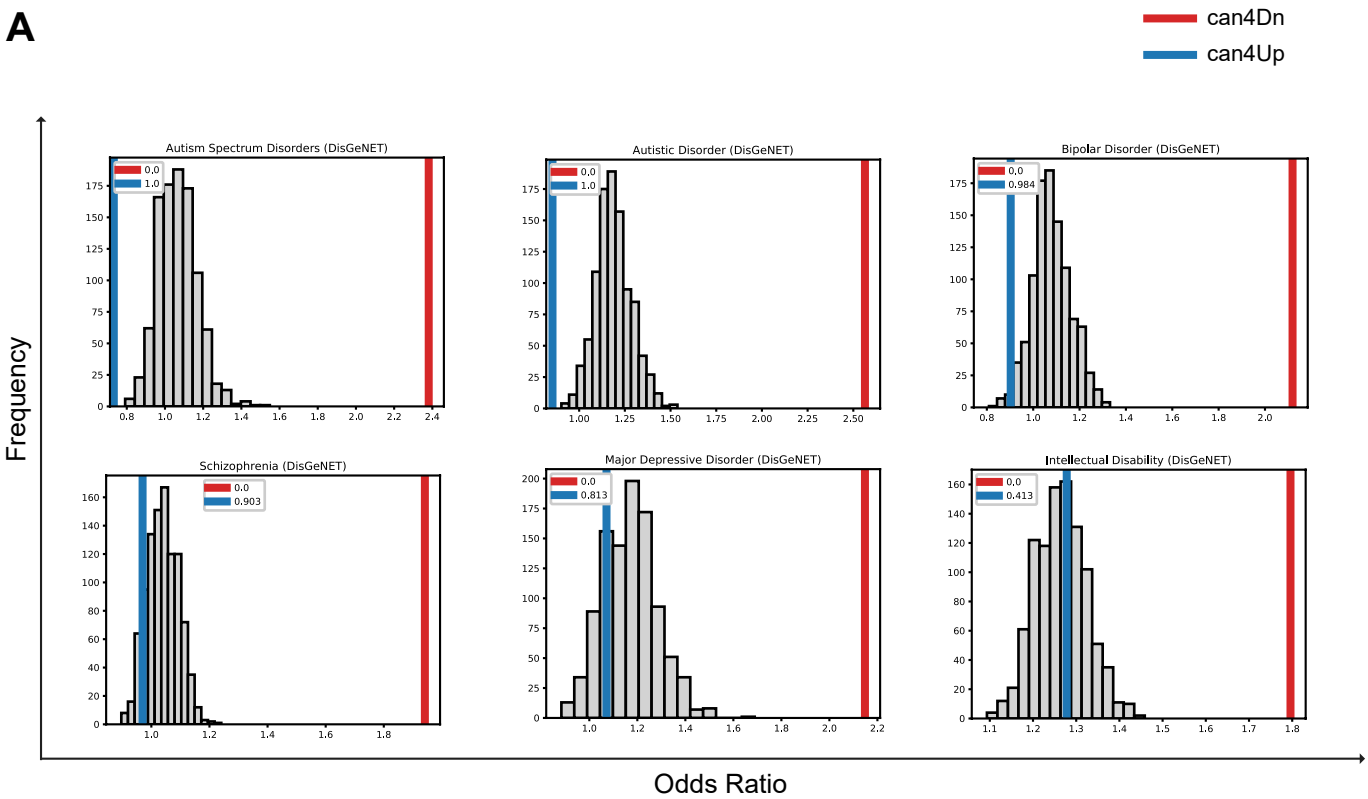
Supplemental Figure 5

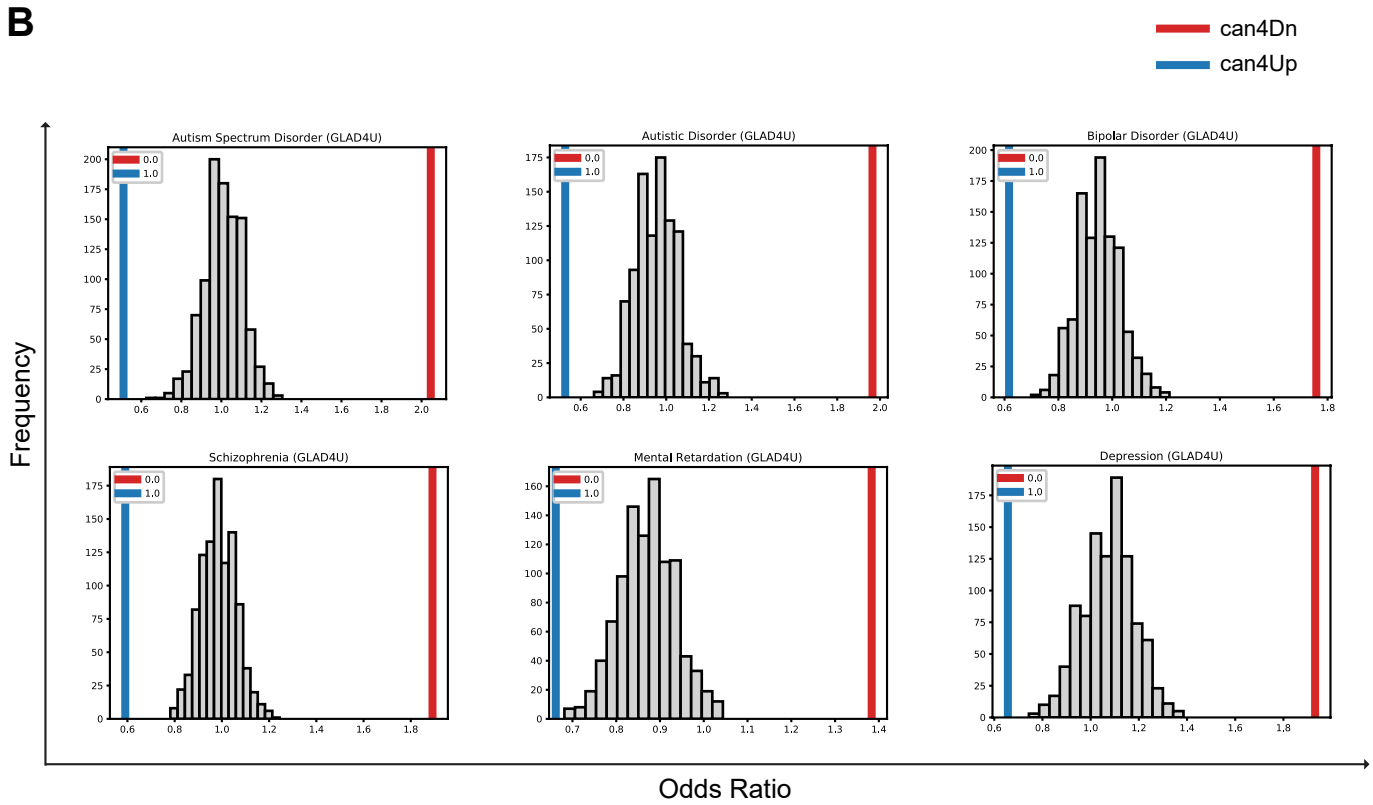


Supplemental Figure 6

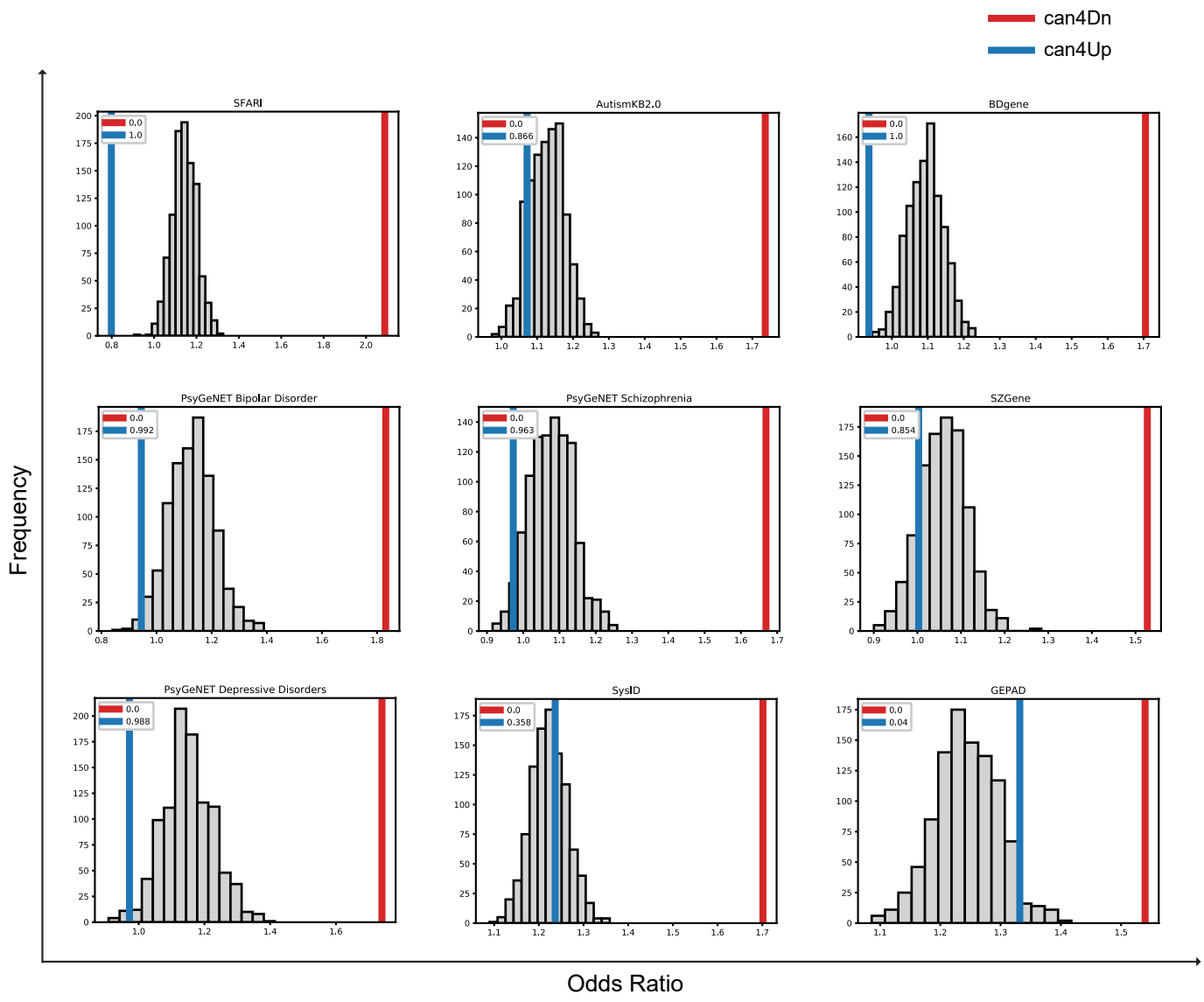


Supplemental Figure 7

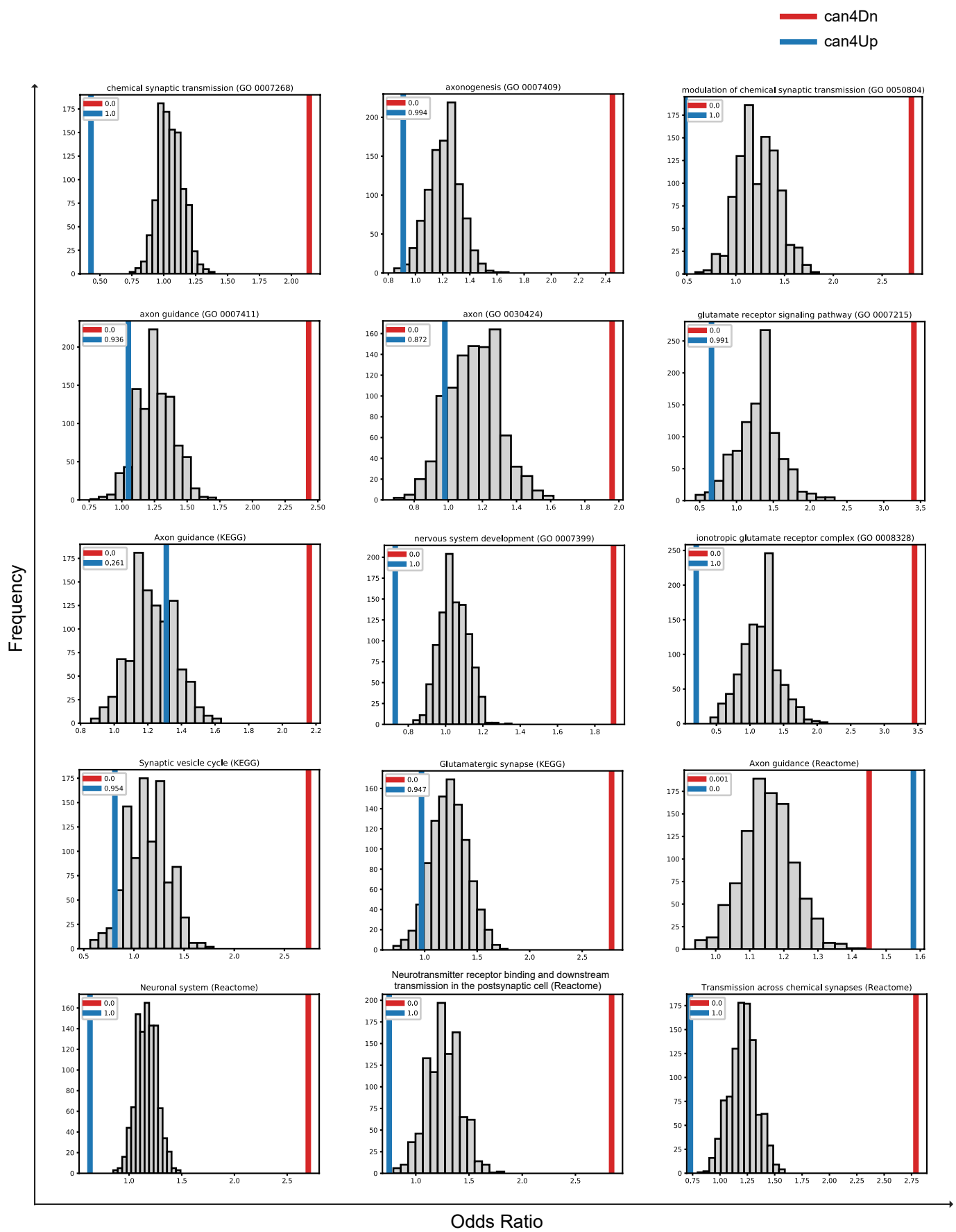




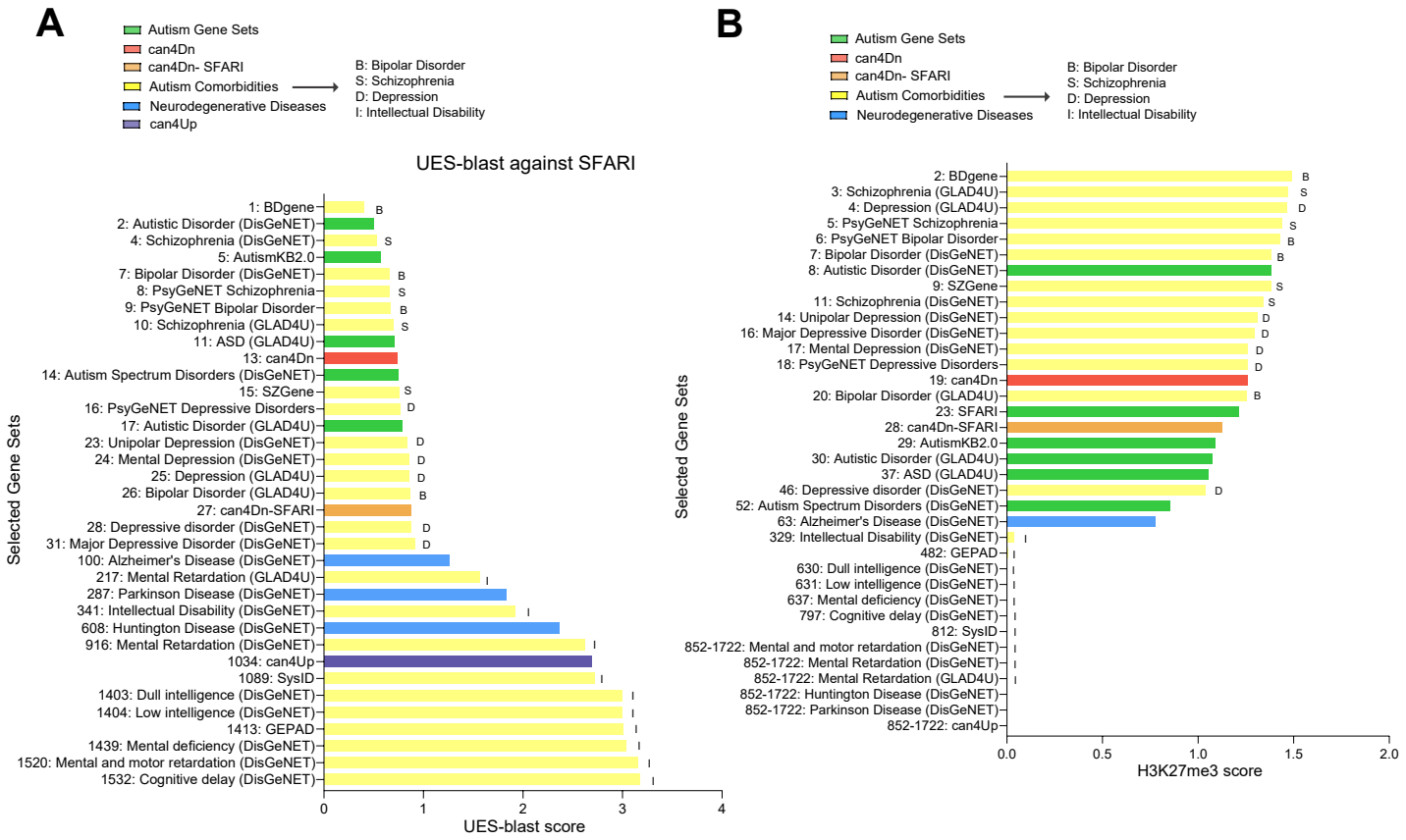
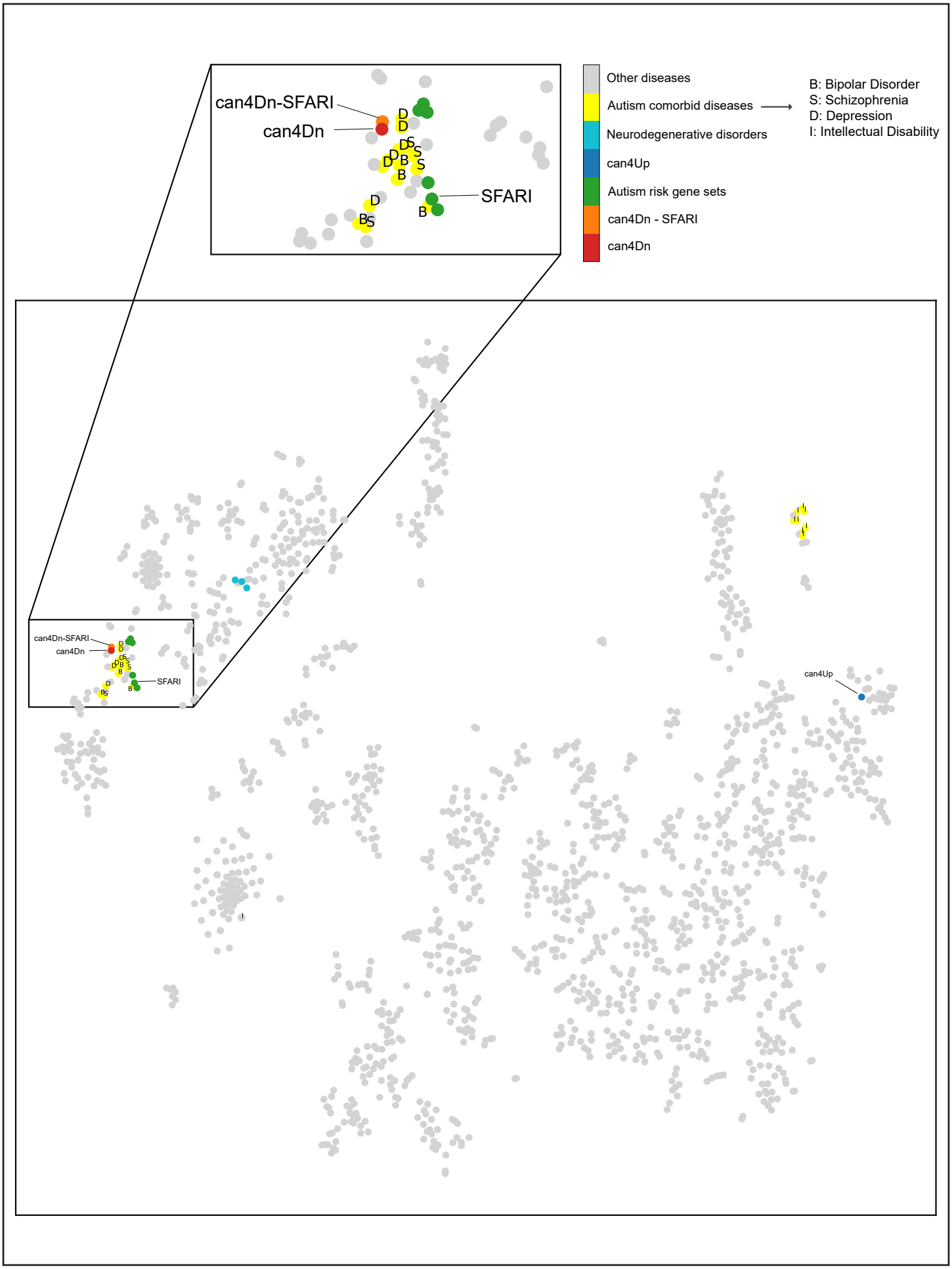
Supplemental Figure 8



Supplemental Figure 9



Supplemental Figure 10



Supplemental Figure 11