A Functional Schizophrenia-associated genetic variant near the *TSNARE1* and *ADGRB1* genes.

Marah H. Wahbeh*[1], Rachel J. Boyd*[1], Christian Yovo[1], Bailey Rike[1], Andrew S. McCallion[1,2], and Dimitrios Avramopoulos[1,•]

1. McKusick-Nathans Department of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA;

2. Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

* Equal contributors

• Corresponding Author

KEYWORDS: gene regulation, genome editing, CRISPR, genetic association, schizophrenia, psychosis, *TSNARE1*, *ADGRB1,* synapse, neurons, axons, neurodevelopment, zebrafish, induced pluripotent stem cells.

ABSTRACT

Recent collaborative genome wide association studies (GWAS) have identified >200 independent loci contributing to risk for schizophrenia (SCZ). The genes closest to these loci have diverse functions, supporting the potential involvement of multiple relevant biological processes; yet there is no direct evidence that individual variants are functional or directly linked to specific genes. Nevertheless, overlap with certain epigenetic marks suggest that most GWAS-implicated variants are regulatory. Based on the strength of association with SCZ and the presence of regulatory epigenetic marks, we chose one such variant near *TSNARE1* and *ADGRB1*, rs4129585, to test for functional potential and assay differences that may drive the pathogenicity of the risk allele. We observed that the variant-containing sequence drives reporter expression in relevant neuronal populations in zebrafish. Next, we introduced each allele into human induced pluripotent cells and differentiated 4 isogenic clones homozygous for the risk allele and 5 clones homozygous for the non-risk allele into neural precursor cells. Employing RNA-seq, we found that the two alleles yield significant transcriptional differences in the expression of 109 genes at FDR <0.05 and 259 genes at FDR <0.1. We demonstrate that these genes are highly interconnected in pathways enriched for synaptic proteins, axon guidance, and regulation of synapse assembly. Exploration of genes near rs4129585 suggests that this variant does not regulate *TSNARE1* transcripts, as previously thought, but may regulate the neighboring *ADGRB1*, a regulator of synaptogenesis. Our results suggest that rs4129585 is a functional common variant that functions in specific pathways likely involved in SCZ risk.

1. INTRODUCTION

Schizophrenia (SCZ) is a chronic, neuropsychiatric disorder generally characterized by a set of symptoms that include delusions and hallucinations; disorganized speech and behavior; and cognitive impairments, such as mental, emotional, and social deficits[1]. SCZ is estimated to impact over 24 million individuals worldwide, with particularly high prevalence, incidence, and burden in the United States[2]. In addition to the mental, social, occupational, and educational duress associated with SCZ, affected individuals have a reduced life expectancy by about 12-15 years relative to the general population[3–5].

While the genetic contribution to SCZ risk is high, with a heritability of 64%-80% [6–8], patients exhibit variable phenotypes and complex inheritance patterns. This is indicative of an environmental contribution to disease risk mediated by gene-by-environment interactions[7,9]. Current treatment strategies for SCZ include the administration of antipsychotic drugs. However, many clinical trials for antipsychotics have failed or yielded negative-to-moderate effects in SCZ patients, who exhibit variable responses to treatment depending on age of onset, sex, geographic location, and other factors[10]. This clinical heterogeneity is thought to reflect the contribution of varying underlying biological processes among patients[11]. Therefore, in order to develop more effective or targeted therapeutics, it is imperative that we understand the genetic architecture of this disorder.

In recent years, collaborative large genome wide association studies (GWAS) have robustly identified 287 independent loci contributing to the risk for SCZ[12]. While these discoveries are valuable, one of the primary limitations of a GWAS is that it cannot resolve the lead single nucleotide variant (SNV) driving the risk association, from other variants in linkage disequilibrium (LD). It is also increasingly acknowledged that most variants identified by GWAS are located in non-coding regions of the genome, and ~40% of the time, the haplotype blocks that contain these lead SNVs do not include coding exons[13–15]. In fact, a growing body of evidence, including expression Quantitative Trait Loci (eQTL) studies, suggest that most variants in LD with lead GWAS SNVs are regulatory in nature[15–19] and concentrate in regulatory

DNA marked by deoxyribonuclease I (DNase I) hypersensitive sites (DHSs)[20]. Nevertheless, even when variants are eQTLs for specific genes, it is often not just one gene that they regulate, and the LD conundrum remains. As a result, genes are often implicated because of their proximity to SCZ-associated variants and known involvement in synaptic biology but are seldom accompanied by direct functional evidence.

While identifying reliable genetic associations is a significant first step towards understanding the etiology of SCZ, functional studies linking specific variants to the regulation of specific genes and biological processes is an important next step towards unlocking their translational potential. Understanding how SCZ associated variants are involved in regulatory activities will promote the generation of innovative disease treatments and prophylactics. Technological advancements in cell engineering and genome editing using CRISPR/Cas9[21] – and the ability to generate human induced pluripotent stem cells (hiPSCs) that can be differentiated to a variety of cell types – have created a new toolkit for modeling human diseases[22]. Neurons that carry variants implicated in disease risk can now be generated and compared to neurons that carry non-risk alleles at the same position, on an otherwise isogenic background. This strategy can be used to not only show the functionality of non-coding, potentially regulatory variants, but also to observe its consequences in a disease-relevant context. One of the most informative ways to characterize and phenotype such allelic differences is to study the transcriptome, in which allele-specific effects may be identified that impact the target gene, but also point to disrupted transcriptional programs and their pathways.

We employ this strategy in the work we present here, where we test the functionality of one such non-coding, SCZ-associated, SNV based on the strength of its association to SCZ and the presence of open chromatin marks in a disease-relevant, neuronal population. We first employ a well-established transgenic zebrafish reporter assay[23–25] to confirm the regulatory potential of the sequence within which this variant resides. We then use precise genome editing to generate isogenic hiPSC-derived neural precursor cells (NPCs) that possess the risk allele and non-risk allele and compare the transcriptional differences between

NPCs with the SCZ risk allele versus the non-risk allele. Our results support the functionality of this SCZ risk variant, identify transcriptional changes associated with possession of this variant, and unearth additional downstream effects within SCZ-related pathways in the transcriptome.

2. MATERIAL AND METHODS

2.1. Variant Prioritization

We prioritized characterization of variant rs4129585 based on the information available at initiation of the project. The second version of SCZ GWAS data was available at the time from the SCZ working group of the psychiatric genetics consortium (PGC) (https://pgc.unc.edu/for-researchers/working-groups/schizophrenia-working-group/)[26], which indicated that rs4129585 was the top-ranking SNV at that locus near the *TSNARE1* and *ADGRB1* genes (p= 2.03x10^{-13}, risk allele = C). This variant was reported by the PICS fine-mapping algorithm (pics2.ucsf.edu) to be the most likely driver of SCZ association at that locus. Using the UCSC genome browser, we found that rs4129585 overlaps with a chromatin immunoprecipitation signal from restrictive element-1 silencing transcription factor / neuron-restrictive silencing factor (REST/NRSF) and was shown to drive reporter gene expression in K562 cells (p=2.4 x10^{-4}) in a massively parallel reporter assay (MPRA) we recently reported[27]. Of note, another nearby SNV, rs13262595, is in very high LD with rs4129585 (LD, r^2=0.996 from https://ldlink.nih.gov/), which was also strongly associated with SCZ risk. At the time, this SNV showed slightly weaker association (p= 2.85x10^{-13}, risk allele = G); however, in most recent PGC data[12], rs13262595 had slightly stronger signal (p= 5.109x10^{-18}) compared to rs4129585 (p= 7.062x10^{-18}). It also overlaps with many chromatin marks, making it another good candidate to drive SCZ risk at this locus. This SNV was not found to drive reporter gene expression in our published MPRA[27]. Nevertheless, rs13262595 remains a target of our future investigations.

121 2.2. Functional Assessment using Zebrafish Transgenesis

122 To evaluate the regulatory potential of rs4129585 and its surrounding sequence, we employed
123 Tol2 transposon-mediated transgenesis according to our standard protocol[23–25], with minor modifications.
124 Briefly, primers were designed to amplify a human non-coding sequence containing rs4129585 (**Table S1**).
125 The boundaries of this sequence (hg38 chr8:142231443-142231752) were chosen to include a ChIP-seq
126 peak surrounding rs4129585, as identified by the UCSC Genome Browser hg38 track "Transcription Factor
127 ChIP-seq Clusters (340 factors, 129 cell types) from ENCODE 3," suggestive of restrictive element-1
128 silencing transcription factor (REST)/neuron-restrictive silencing factor (NRSF) transcription factor (TF)
129 occupancy.

130 Next, we ligated 5' *att*B sequences to each primer set and amplified our chosen sequence by PCR.
131 Instead of gel extraction, leftover PCR product was purified using the DNA Clean & Concentrator™-5 kit
132 (Zymo Research #D4067). The *attB* PCR product was cloned into an empty pDONR™221 vector, and the
133 resulting construct underwent recombination with a pGW_cfosEGFP destination vector, such that the
134 conserved non-coding sequence was placed upstream of a *cfos* minimal promoter that drives EGFP
135 expression and flanked by Tol2 transposons that promote efficient genomic integration of the reporter
136 construct. Purified recombinant DNA was injected into 1–2-cell zebrafish embryos, as previously
137 described[23]. Injections were performed in zebrafish embryos of the strain AB, which were obtained from
138 the Zebrafish International Resource Center (http://zfin.org; ZFIN ID: ZDB-GENO-960809-7) and raised in
139 our facility. Embryos were screened for EGFP expression at 4- and 6-days post fertilization (d.p.f.) using a
140 Nikon AZ100 Multizoom fluorescence microscope, and the embryos with the strongest EGFP signal were
141 selected to be raised to sexual maturity.

142 After 3 months, ≥5 sexually mature adults were crossed with wild-type AB fish and F_1 embryos
143 were screened for EFGP to confirm germline transmission of the transposon-mediated insertion, again
144 using a Nikon AZ100 Multizoom fluorescence microscope at 6 d.p.f. To ensure that position effects of

individual transgene insertions did not confound our results[23], F_1 embryos were screened from each of 5 independent parental clutches, and ≥3 representative F_1 founder fish with consistent EGFP expression were selected from each clutch and raised to sexual maturity. At 3 months of age, F_1 founders were again crossed with wild-type AB fish. The resulting F_2 embryos were collected and raised in embryo media (E3) that contained 200 µM 1-phenyl 2-thiourea (PTU; Sigma Aldrich #P7629) to inhibit pigmentation[28]. This E3/PTU media was replaced daily.

At 6.d.p.f., ethyl-3-aminobenzoate methanesulfonate salt (Tricaine/MS-222; Sigma Aldrich #E10521) was prepared at a concentration of 4 mg/mL, buffered to pH7 using 1M Tris (pH 9; 4% v/v), and added to E3 medium to anaesthetize zebrafish larvae during the imaging process. Imaging was performed using the GFP filter and auto exposure settings on a Nikon AZ100 Multizoom fluorescence microscope, and GFP expression patterns among founder zebrafish were analyzed (**Table S2**), after which zebrafish larvae were transferred to a new dish containing fresh E3 media.

2.3. Cell line used for CRISPR editing and cell culture.

The iPSC line MH0180966 used for CRISPR/Cas9 editing and further experiments was derived from a female individual of European descent without a psychiatric diagnosis and provided to us by the stem cell repository at Rutgers University (NIMH/RUCDR). We received the line at passage number 12. The iPSC line was treated as we have outlined previously[21,29]. Briefly, the cells were grown on culture plates coated with 5µg/mL laminin (ThermoFisher BioLamina #LN521) and incubated in a humidified 5% CO_2 incubator at 37°C for ≥2 hours. Once seeded, the cells were maintained in StemFlex media with supplement (ThermoFisher Gibco #A3349401) and 1% penicillin–streptomycin (Penn/Strep; Gibco # 15140122) in a humidified 5% CO_2 incubator at 37°C. When thawing, passaging, or transfecting, 10µM Y-27632 Rho-kinase inhibitor (ROCKi; Tocris #1254) was added to culture media. Cells at 70-80% confluence were passaged using 1X Phosphate-buffered saline (PBS) and accutase (MilliporeSigma #A6964) every ~4-5 days. Cell

stocks were first frozen at -80°C in StemFlex media with 10% Dimethylsulfoxide (DMSO) for 24h before storage in liquid nitrogen. iPSC genotyping was done by PCR and Sanger sequencing using primers listed in the Supplementary Information file of Feuer & Wabeh, 2022 (see Online Methods; *TSNARE1* experiment)[21].

2.4. Generating Allele-Specific iPSC Lines

To induce target sequence deletion and subsequent re-insertion of the rs4129585 non-risk (A) and risk (C) alleles, we employed our pipeline for scarless genome editing, CRISPR Del/Rei[21]. Detailed methods, transfection and selection details, and summary of off-target analyses; as well as sequences for primers, single guide RNAs (sgRNAs), and single stranded oligodeoxynucleotide (ssODN), can be found in the Supplementary Information file of Feuer & Wahbeh *et al*., 2022 (see Online Methods; *TSNARE1* experiment)[21]. In brief, sgRNA oligos were designed using CRISPOR (http://crispor.tefor.net/crispor.py) to flank rs4129585 and induce a deletion between GRCh37/hg19 chr8:142231553-142231597. Restriction enzyme BbsI sticky ends were added to single-stranded DNA oligos before the oligos were annealed and cloned in pairs into pDG459 plasmids containing a puromycin resistance gene. pDG459 plasmids were transfected into iPSCs using lipofectamine-stem and successful transfection was selected for using puromycin[21]. Cells were screened for desired edits and candidate off-target effects[21] at 4 genomic locations that showed 3 or fewer mismatches with the gRNAs and were in coding sequences or overlapped with DNAse hypersensitivity sites from the encode project (www.encodeproject.org; file hg38_wgEncodeRegDnaseClustered.txt). Single cell clones with the appropriate deletion and no off-target events were chosen for expansion. Next, a synthetic sgRNA (syn-sgRNA) and an ssODN repair template were designed to reinsert either the non-risk allele or risk allele at the appropriate genomic position. As with the deletion step, syn-sgRNAs were cloned into a pX459 plasmid with a puromycin resistance gene, transfected into iPSCs, and edited cells were selected for using puromycin, before iPSCs were again

193 screened for appropriate reinsertion and lack of off target effects at 8 candidate genomic locations, as
194 above [21].
195
196 2.5 Differentiation into Neural Precursor Cells
197 Each iPSC clone was differentiated into Neural Progenitor Cells (NPCs) through directed
198 differentiation of embryoid bodies[30] (EB) with SMAD, as previously described[31,32] with minor modifications.
199 In brief, iPSCs were first plated at a density of ~300 k/well in 3 wells of 6-well laminin-coated plate.
200 Differentiation was initiated 2 days later by passaging cells at a density of 5K cells/well in a round-bottom,
201 untreated 96-well plate (Corning #3788) in 50ul/well of MTeSR media (Thermo #85850) and Rocki (Y-27632
202 dihydrochloride; BD Biosciences # 562822). Directly after plating, the plate was spun at 1,000 RPM for 3-4
203 minutes to condense the cells and facilitate EB formation. The following day, (DIV0) the presence of EBs
204 were confirmed and 50uL of Neural Induction Media (NIM) was added to each well, composed of
205 DMEM/F12 Media (Thermo #11320033) with 2% B27 supplement without vitamin A (Thermo #12587010),
206 1% GlutaMAX (Thermo # 35050061), 1% N2 (Thermo #17502048), 1% MEM NEAA (Non-Essential Amino
207 Acids; Thermo #11140050), and 1:100 BME (Beta-mercaptoethanol; Thermo #21985023) with freshly
208 added SMAD Inhibitors SB431542 (Reprocell #NC1388573) at 20uM and LDN193189 (Peprotech #1062443
209 -1mg) at 200nM. From DIV1-DIV3, 50ul of NIM was added to each well with half the concentration SMAD
210 Inhibitors as day 1 (SB431542 at 10uM, LDN193189 at 100nM) until reaching 200ul/well.
211 From DIV4 onwards, 100ul of NIM with SMAD Inhibitors was exchanged daily until EB maturity
212 (DIV6 - DIV9). EBs were passaged into 3 wells of a 6-well plate that was coated in 1:100 matrigel (Corning
213 # 47743-720) for 1 hour at 37°C and contained 2ml of NIM/well with SMAD Inhibitors. EBs were then
214 carefully removed from the 96- well plate into 1-2 wells of an uncoated 6-well plate with warm DMEM/F12,
215 before being transferred to the coated plate containing NIM. The EBs adhered to the Matrigel and were
216 fed daily from DIV7-DIV13 with NIM, SMAD inhibitors, and 10ng/mL FGF2 (Miltenyi #130-093-838) until

robust rosettes formed. Once formed, rosettes were manually picked and transferred into 3 wells of a 6-well plate coated in 1:100 matrigel for 1 hour at 37°C and containing 2ml of NIM/well with SMAD inhibitors. The following day onwards, cells were fed with NIM, SMAD inhibitors, and FGF2 until robust rosettes formed again. The rosettes were again lifted manually, and Accutase was added to separate the cells before they were plated into NPC media containing half DMEM/F12 and half Neurobasal Media (Thermo #21103049) with 2% B27 supplement without vitamin A, 1% GlutaMAX, 1% N2, 1% MEM NEAA and 1:100 BME. FGF2, 10ng/ml EGF (Peprotech #100-47-10ug) and Rocki were also added. Cells at this stage were fed with this NPC media without Rocki every 2 days until ~70% confluent. After 2 passages, the cells were characterized by the presence of NPC markers.

2.6 Characterization of NPCs

To confirm the identity and quality of NPCs, we visualized the expression of NPC markers and lack of expression of iPSC markers by immunocytochemistry staining. 3-4 days before staining, a 4-well chamber slide (Thermo #62407-294) was coated with 1:100 matrigel in DMEM/F12 for 1 hour (300uL per well) and ~150k NPCs were plated into each well. Slides were fed once every 2 days with NPC media until the cells were ~60% confluent. NPC media was then removed, and the cells were washed with cold 1XPBS and fixed using 4% PFA (Alfa Aesar #J61899). Cells were blocked for 1 hour using a blocking buffer composed of 1XPBS, 5% Goat Serum, and 0.3% Triton X-100 (Sigma #93443-100mL), after being washed three times with 1X PBS for 5 minutes each time. Following blocking, cells were incubated in antibody solution made up of 1XPBS, 1% BSA, and 0.3% Triton X-100, as well as different combinations of 1:200 Nestin antibody, 1:200 PAX6 antibody, 1:300 TRA-160 antibody, 1:300 SOX2 antibody, and 1:200 Nanog antibody at 4°C overnight. Slides were washed three times with cold 1X PBS for 5 minutes each time before adding antibody solution with 1:400 488 Goat α-Mouse secondary antibody and 1:400 565 Goat α -Rabbit secondary antibody. Incubation lasted 1h at 37°C before washing three times with cold 1X PBS again before

mounting with 25uL of the prolong mounting media with DAPI (Invitrogen #P36931) to each well. Slides were imaged with an Axio Observer fluorescent microscope. Next, we performed RT-qPCR in some of the samples by generating cDNA from 1ug of RNA from each clone before and after differentiation (iPSCs and NPCs) using the Superscript III kit, following the manufacturer's protocol (ThermoFisher # 18080400). cDNA was diluted 1:10 in water for RT-qPCR. We used iTaq and custom primers (**Table S3)** following the manufacturer's protocol (BioRad # 1725120). We used the CFX Connect System with denaturation at 95°C for 20 sec, amplification for 40 cycles, Denaturation at 95°C for 1 sec, annealing at 60°C for 30, and Melt Curve Analysis at 65-95°C. Normalized relative quantification and error propagation was calculated and analyzed as proposed by Taylor *et al.* (2019)[33], with results normalized to *ACTB* and *GAPHD*. A two-sample t-test was performed using R function "t.test", (alternative = "two.sided") to test for significant differences in gene expression.

2.6. RNA-sequencing

Total RNA was isolated from 1 confluent well of NPCs/line from a 6 well plate using the Zymo Quick-RNA™ Miniprep Kit (#R1054) following the manufacturer's protocol. RNA was submitted to Novogene (novogene.com) for 150-bp paired-end RNA-sequencing. All libraries passed Novogene's post-sequencing quality control, except REF_5, which was excluded from analysis. We received an average of 51.1 million reads per sample with a maximum of 74.9 and a minimum of 42.5 million reads. Reads were aligned to the human reference genome GRCh38 using Hisat2 version 2.1.0[34] and corresponding BAM files were generated using Samtools 1.9[35]. Stringtie 2.0.3[36] was used to assemble and estimate the abundance of transcripts based on GRCh38 human gene annotations[37]. Bioconductor package tximport was used to compute raw counts by reversing the coverage formula used by Stringtie with the input of read length[36]. The output was then imported to the Bioconductor package, DESeq2, for differential gene expression analysis[38]. Benjamini-Hochberg adjustment[39] was used to calculate adjusted p-values (padj). Before

expression analysis, principal component (PC) analysis (**Figure S6; Figure S7**) was performed on the adjusted read counts and the first PC was included in further analysis as a covariate.

2.7 Bioinformatics Analysis

To explore which genes were differentially expressed (DE) between edited and unedited iPSC-derived neurons (rs4129585 non-risk and risk alleles), we used the Panther bioinformatics platform[40] (https://www.pantherdb.org/), which reports enrichments across the Gene ontology (GO) resource terms (https://geneontology.org/) corrected for multiple testing. We used all genes with detectable transcripts in our dataset as a background. To look for interactions between the DE genes, we used the STRING database[41,42] (string-db.org). As suggested by the curators of the STRING database website, we used the medium-level confidence score (0.4) for interactions and did not include any additional level of interactors, to ensure the validity of the observed network enrichment p-values. In addition to enrichment for interactions, STRING reports significant functional enrichments, similar to Panther but tested only among interacting DE genes, thus pointing to the function of the identified network.

3. RESULTS

3.1. A transcription factor binding site containing rs4129585 drives reporter gene expression in transgenic zebrafish neuroanatomy.

To test rs4129585 and the surrounding sequence for regulatory function, we utilized a transgenic zebrafish reporter assay in which the presence of an upstream regulatory element is required for a minimal promoter to induce GFP reporter gene expression. The SCZ-associated variant rs4129585 lies within a ChIP-seq REST/NRSF TF binding site (TFBS), and the entire sequence representing the ChIP-seq peak was integrated into the zebrafish genome upstream from a *cfos* minimal promoter/GFP reporter construct. REST is an epigenetic remodeler that is widely expressed during human embryogenesis and late-stage

289 neuronal differentiation. REST acts as a transcriptional repressor of neural-specific genes involved in
290 synaptogenesis, axonal pathfinding, synaptic plasticity, and structural remodeling in humans[43,44];
291 therefore, it is unsurprising that this sequence drives GFP expression in the zebrafish brain (**Figure 1)**.

292 Representative GFP expression pattern from F_2 zebrafish embryos suggest that this sequence is
293 functional in three major neuroanatomical regions (**Figure 1 & Figure S1**): (1) the habenula, which is
294 involved in aversive social responses in zebrafish[45], including anxiety[46], fear[47], and aggression[48]; (2) the
295 caudal hypothalamus, which is involved in appetite control[49], but in which dopaminergic (DA) neurons are
296 also involved in audio-sensory functions, including locomotion and sensation of acoustic/vibrational
297 stimuli[50]; and (3) the inferior olive, which receives DA projections from the caudal hypothalamus[50], but
298 also provides climbing fibres to Purkinje cells of the cerebellar cortex that are required for motor
299 learning[51,52]. The more widespread GPF signal in these fish is consistent with expression in glutamatergic
300 neurons, as well as a few regions with GABAergic neurons (**Figure 1**). These findings suggest that the SCZ-
301 associated variant, rs4129585, likely lies within a TFBS that elicits its regulatory function in glutamatergic
302 neurons and regions of the brain that have been implicated in neurodevelopment and SCZ pathology[53–62].
303
304 *3.2. iPSC-derived NPCs exhibit transcriptomic changes associated with a SCZ risk allele.*

305 With evidence that rs4129585 lies within a functional regulatory element implicated in
306 neurobiology, we obtained human-derived iPSCs, used CRISPR-mediated genome editing to engineer the
307 rs4129585 risk and non-risk alleles into otherwise isogenic cell lines, and differentiated these lines into
308 NPCs. NPCs were selected as the differentiation endpoint because they are at an early stage in neuronal
309 differentiation, and SCZ is considered a neurodevelopmental disorder[59,60,63,64]. Immunostaining and RT-
310 qPCR with NPC- and iPSC-specific markers showed the former to be upregulated and the later
311 downregulated (**Figure S2, Figure S3, Figure S4, Figure S5**); thus, confirming successful differentiation from
312 iPSCs to NPCs. We then performed RNA-seq to identify the transcriptional changes elicited by the altered

variants. PC analysis of the RNAseq data identified one reference clone (REF_6) to be an outlier across PC1 (**Figure S6**). NPC and iPSC marker gene expression (**Table S4**) suggest that this cell line might not be as far along in the differentiation process. To correct for this difference going forward, we used PC1 as a covariate in the analysis. Differential expression analysis revealed differences in the expression of 109 genes at FDR <0.05 and 259 genes at FDR <0.1. Of these, 116 were upregulated and 143 were downregulated (**Table S5** shows the complete results at the gene level, while **Table S6** shows the results at the transcript level). We then proceeded to test these genes for functional enrichments using the Panther Bioinformatics database[40], using the FDR <0.1 threshold to obtain higher statistical power. This revealed strong enrichments in many functional categories with direct relevance to schizophrenia, such as neuron development, differentiation, axonogenesis, and axon guidance, among many others (**Figure 2**).

We next used the STRING database of protein-protein interactions[41,42] to test whether the protein products of the dysregulated genes are functionally related as evidenced by reported interactions between them, an indication of validity of the results which also allowed further functional exploration. This analysis showed that these genes are highly interconnected (STRING enrichment-p=1.2x10^{-11}) with 1.66-fold more interconnections than expected by chance for the same number of proteins. Further, functional enrichment analysis for the network proteins as performed by STRING showed many of the same enrichments observed with Panther Bioinformatics, now including synapse organization and assembly, (FDR<0.02), central nervous system development, regulation of synapse assembly and axon guidance. **Figure 3** shows all enrichments and **Figure 4** shows as an example the 11 axon guidance genes in the STRING network, representing a more than 4-fold enrichment from the expected.

Results from the Genotype-Tissue Expression (GTEx) project (www.gtexportal.org/home/) show that rs4129585 is an eQTL for its' two nearest genes, *TSNARE1* and *ADGRB1,* regulating both the genes and specific transcripts' abundances in a variety of tissues and at varying directions (including cerebellum but not brain). We looked at our results for the effects of the variant on these two genes and found no

337 effect on *TSNARE1* and a suggestive effect on *ADGRB1*(*p*=0.06), with the risk allele increasing expression
338 in our induced NPCs. Interestingly, *ADGRB1*, is a regulator of synaptogenesis[65]. None of the transcripts of
339 *TSNARE1* (9 present in our data) showed significant differences. Of the 5 *ADGRB1* transcripts in our data,
340 two showed significant differences (ENST00000517894, nominal *p*=0.005 and ENST00000518812, nominal
341 *p*=0.03), both of which were upregulated with the risk allele. Note that ENST00000518812 is not a protein
342 coding transcript.
343
344 4. DISCUSSION
345 In the last decade, GWASs have successfully identified relatively common genetic variants that
346 drive disease associations, which has inspired a new chapter in genetic disease research that aims to
347 identify the biological mechanisms through which genetic variation influences disease risk. Cellular
348 modeling is quickly becoming a powerful tool for the study of genetic risk variants, particularly those
349 within non-coding regions of the genome, as technological advances are being made in genome editing,
350 iPSC generation from somatic cells, and directed differentiation methods[22]. Now, we are able to generate,
351 study, and compare human neuronal cells with only a single nucleotide difference across the entire
352 genome, as we and others have previously reported[21,29,66–70]. This approach provides the opportunity to
353 elucidate the individual contributions of risk variants in a human- and disease-relevant system. Here, we
354 apply this strategy to study a GWAS-identified, non-coding variant implicated in SCZ risk, rs4129585.
355 We first tested the regulatory potential of this variant in zebrafish and found that it drives reporter
356 gene expression exclusively in the brain, and in regions relevant to SCZ, including areas involved in fear,
357 anxiety, audio-sensory functions, and sensory-motor functions. Further, according to ENCODE data, this
358 non-coding variant lies within a REST binding site, a TF already implicated in SCZ[44]. We next introduced the
359 risk and non-risk alleles into iPSCs, differentiated them into biologically relevant NPCs, and produced
360 clones homozygous for each allele on otherwise identical genomic backgrounds. Comparing the risk and

reference alleles, we observed changes in the abundance of an alternatively spliced nearby gene, *ADGRB1,* as well as transcriptome changes affecting genes involved in axon guidance, synaptogenesis, and other functional pathways. The identified DE genes encode proteins that interact with each other more than would be expected by chance; further validating our results and supporting the observed functional enrichments.

Although we are not the first group to link synaptogenesis, axon guidance, and neuronal development with SCZ, we do so here for a specific SCZ risk variant, which we implicate as a regulator of a specific gene, *ADGRB1*, an adhesion G protein-coupled receptor[71]. Consistent with our transcriptomic observations, *ADGRB1* has been linked to brain development and developmental disorders[72,73], dendritic arborization[74], synapse development and synaptic plasticity[65,75,76]. As such, it is not surprising to find suggestive evidence that it drives the risk in genetic associations with SCZ. The neighboring *TSNARE1* is also a strong candidate for SCZ risk, and it was originally reported as the driver of the association due to the localization of the lead variants within the gene[77] in addition to its role in regulating endosomal trafficking in cortical neurons[78].

The SNV rs4129585 has been previously identified as influencing the expression of nearby genes; in fact, GTEx shows that this variant is an eQTL for both *TSNARE1* and *ADGRB1*. However, since GTEx data is purely statistical – not functional – and comes from multiple tissues with bulk cells containing a variety of cell types, it only reflects the cumulative effects of all variants in LD at the locus where rs4129585 resides. There is no reason to assume that rs4129585 is the only functional variant at this locus, especially since there are other variants in LD with functional potential, such as rs13262595, as discussed in our methods. Our data shows that *ADGRB1* transcription is influenced by rs4129585 and by the sets of DE genes across the transcriptome, supporting the involvement of *ADGRB1* rather than *TSNARE1*, it remains possible that *TSNARE1* is also involved, but is regulated by another SNV in LD, such as rs13262595. Study of all such variants individually will shed additional light onto how these alleles and corresponding

385 haplotypes influence SCZ risk and will provide valuable insights for the study of GWAS-identified variation,
386 where the identification of multiple variants in strong LD is the rule, rather than the exception. Such future
387 studies will eventually solve the puzzle of the complex interplay between multiple variants that can
388 influence disease risk.

389 Our work suggests that, despite having a small effect on risk, a common variant can have a
390 measurable effect on the transcriptome of NPCs, which points to specific functional pathways that, in turn,
391 indicate disease mechanisms. There are three main reasons why this could be the case. First, studying risk
392 and non-risk alleles in homozygosity, in otherwise isogenic cells, and under well-controlled conditions in
393 vitro, may reduce noise and increase power to detect even small effects. Second, the study of cells in vitro
394 not only reduces noise, but may also avoid a feedback buffering mechanism that acts at the organism level
395 to dampen the effect of the variant[79]. Finally, studying an allele in isolation from other potentially
396 functional alleles in LD might allow for the observation of effects that would otherwise be counteracted
397 by the other alleles on the same haplotype. Natural selection could be making such a phenomenon more
398 frequent than we recognize, if deleterious haplotypes are being selected against. This is particularly likely
399 in SCZ, where negative selection appears to be strong[80–83].

400 While studying NPC cultures in isolation has advantages, this isolation is also a limitation, since
401 only cell-autonomous effects can be revealed. To bypass this limitation, 3D cerebral organoid culture
402 systems could be leveraged, and accompanied by single cell sequencing. Despite being a more time
403 consuming and expensive approach, it would be a valuable addition to monocellular cultures. While we do
404 not currently have such results in organoid-based cultures or involving other GWAS-identified risk variants,
405 we consider this, along with the study of multiple variants per locus, as an important addition to future
406 projects.

407

AUTHOR CONTRIBUTIONS: D.A. conceptualized the study design. R.J.B. designed and performed transgenic zebrafish reporter assays. M.H.W., C.Y., and B.R. performed cell culturing. M.H.W. performed genome editing of iPSCs, differentiation, and validation of NPCs. R.J.B. analyzed RT-qPCR data. D.A. performed bioinformatics analyses with contributions by R.J.B. R.J.B, D.A., and M.H.W. wrote the paper. M.H.W, R.J.B., C.Y., A.S.M., and D.A. contributed to the scientific discussion, data interpretation, and paper revision. All authors have read and agreed to the published version of the manuscript.

FUNDING: This research was supported in part by awards from the National Institutes of Health (R01MH113215 and R21MH122936) to D.A., (R01HG010480 and R21NS128604) to A.S.M., and (T32GM007814-40) to R.J.B.; and by the Canadian Institutes of Health Research (DFD-181599) to R.J.B.

DATA AVAILABILITY: RNA-sequencing data will be available at the Gene Expression Omnibus (GEO) upon article acceptance.

ACKNOWLEDGEMENTS: The authors would like to acknowledge Anna Vakhnovetsky and Akul Umamageswaran for their technical support, as well as the FInZ Zebrafish Core Center for assistance with zebrafish husbandry.

CONFLICTS OF INTEREST: The authors declare no competing interests.

REFERENCES

1. Rahman, T. & Lauriello, J. Schizophrenia: An Overview. *Focus J. Life Long Learn. Psychiatry* **14**, 300 (2016).

2. Solmi, M. *et al.* Incidence, prevalence, and global burden of schizophrenia - data, with critical appraisal, from the Global Burden of Disease (GBD) 2019. *Mol. Psychiatry 2023* **6**, 1–9 (2023).

3. Crump, C., Winkleby, M. A., Sundquist, K. & Sundquist, J. Comorbidities and mortality in persons with schizophrenia: a Swedish national cohort study. *Am. J. Psychiatry* **170**, 324–333 (2013).

4. Hjorthøj, C., Stürup, A. E., McGrath, J. J. & Nordentoft, M. Years of potential life lost and life expectancy in schizophrenia: a systematic review and meta-analysis. *The Lancet Psychiatry* **4**, 295–301 (2017).

5. Lawrence, D., Hancock, K. J. & Kisely, S. The gap in life expectancy from preventable physical illness in psychiatric patients in Western Australia: retrospective analysis of population based registers. *BMJ* **346**, (2013).

6. Lichtenstein, P. *et al.* Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet* **373**, 234–239 (2009).

7. Wahbeh, M. H. & Avramopoulos, D. Gene-Environment Interactions in Schizophrenia: A Literature Review. *Genes 2021, Vol. 12, Page 1850* **12**, 1850 (2021).

8. Stilo, S. A. & Murray, R. M. Non-Genetic Factors in Schizophrenia. *Curr. Psychiatry Rep.* **21**, (2019).

9. Tsuang, M. T., Gilbertson, M. W. & Faraone, S. V. The genetics of schizophrenia. Current knowledge and future directions. *Schizophr. Res.* **4**, 157–171 (1991).

10. Rabinowitz, J. *et al.* Determinants of antipsychotic response in schizophrenia: implications for practice and future clinical trials. *J. Clin. Psychiatry* **75**, (2014).

11. Tsuang, M. T. & Faraone, S. V. The case for heterogeneity in the etiology of schizophrenia.

Schizophr. Res. **17**, 161–175 (1995).

12. Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502–508 (2022).

13. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).

14. Visel, A., Rubin, E. M. & Pennacchio, L. A. Genomic views of distant-acting enhancers. *Nature* **461**, 199–205 (2009).

15. Hindorff, L. A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. USA* **106**, 9362–9367 (2009).

16. Schaub, M. A., Boyle, A. P., Kundaje, A., Batzoglou, S. & Snyder, M. Linking disease associations with regulatory information in the human genome. *Genome Res.* **22**, 1748–1759 (2012).

17. Majewski, J. & Pastinen, T. The study of eQTL variations by RNA-seq: from SNPs to phenotypes. *Trends Genet.* **27**, 72–79 (2011).

18. Gilad, Y., Rifkin, S. A. & Pritchard, J. K. Revealing the architecture of gene regulation: the promise of eQTL studies. *Trends Genet.* **24**, 408–415 (2008).

19. Cookson, W., Liang, L., Abecasis, G., Moffatt, M. & Lathrop, M. Mapping complex disease traits with global gene expression. *Nat. Rev. Genet.* **10**, 184–194 (2009).

20. Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).

21. Feuer, K. L. *et al.* CRISPR Del/Rei: a simple, flexible, and efficient pipeline for scarless genome editing. *Sci. Reports 2022 121* **12**, 1–4 (2022).

22. Das, D., Feuer, K., Wahbeh, M. & Avramopoulos, D. Modeling Psychiatric Disorder Biology with Stem Cells. *Curr. Psychiatry Rep.* **22**, (2020).

23. Fisher, S. *et al.* Evaluating the biological relevance of putative enhancers using Tol2 transposon-

475 mediated transgenesis in zebrafish. *Nat. Protoc.* **1**, 1297–1305 (2006).

476 24. Fisher, S., Grice, E. A., Vinton, R. M., Bessling, S. L. & McCallion, A. S. Conservation of RET
477 regulatory function from human to zebrafish without sequence similarity. *Science* **312**, 276–279
478 (2006).

479 25. McGaughey, D. M. *et al.* Metrics of sequence constraint overlook regulatory sequences in an
480 exhaustive analysis at phox2b. *Genome Res.* **18**, 252–260 (2008).

481 26. Ripke, S. *et al.* Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**,
482 421–427 (2014).

483 27. Myint, L. *et al.* A screen of 1,049 schizophrenia and 30 Alzheimer's-associated variants for
484 regulatory potential. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* **183**, 61 (2020).

485 28. Karlsson, J., Von Hofsten, J. & Olsson, P. E. Generating transparent zebrafish: a refined method to
486 improve detection of gene expression during embryonic development. *Mar. Biotechnol. (NY).* **3**,
487 522–527 (2001).

488 29. Wahbeh, M. H. *et al.* A Missense Variant in CASKIN1's Proline-Rich Region Segregates with
489 Psychosis in a Three-Generation Family. *Genes (Basel).* **14**, 177 (2023).

490 30. Yang, L. *et al.* Optimization of scarless human stem cell genome editing. *Nucleic Acids Res.* **41**,
491 9049–9061 (2013).

492 31. Chambers, S. M. *et al.* Highly efficient neural conversion of human ES and iPS cells by dual
493 inhibition of SMAD signaling. *Nat. Biotechnol.* **27**, 275–280 (2009).

494 32. Wen, Z. *et al.* Synaptic dysregulation in a human iPS cell model of mental disorders. *Nature* **515**,
495 414–418 (2014).

496 33. Taylor, S. C. *et al.* The ultimate qPCR experiment: Producing publication quality, reproducible
497 data the first time. *Trends Biotechnol.* **37**, 761–774 (2019).

498 34. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-Based Genome Alignment and

Genotyping with HISAT2 andHISAT-genotype. *Nat. Biotechnol.* **37**, 907 (2019).

35. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078 (2009).

36. Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290 (2015).

37. Pertea, M., Kim, D., Pertea, G. M., Leek, J. T. & Salzberg, S. L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie, and Ballgown. *Nat. Protoc.* **11**, 1650 (2016).

38. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, (2014).

39. Hochberg, Y. & Benjamini, Y. More powerful procedures for multiple significance testing. *Stat. Med.* **9**, 811–818 (1990).

40. Mi, H. *et al.* PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Res.* **49**, D394–D403 (2021).

41. Szklarczyk, D. *et al.* The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **49**, D605–D612 (2021).

42. Szklarczyk, D. *et al.* The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **51**, D638–D646 (2023).

43. Hwang, J. Y. & Zukin, R. S. REST, a master transcriptional regulator in neurodegenerative disease. *Curr. Opin. Neurobiol.* **48**, 193 (2018).

44. Loe-Mie, Y. *et al.* SMARCA2 and other genome-wide supported schizophrenia-associated genes: regulation by REST/NRSF, network organization and primate-specific evolution. *Hum. Mol. Genet.* **19**, 2841–2857 (2010).

45. Ogawa, S. & Parhar, I. S. Role of Habenula in Social and Reproductive Behaviors in Fish:

Comparison With Mammals. *Front. Behav. Neurosci.* **15**, 818782 (2022).

46. Mathuru, A. S. & Jesuthasan, S. The medial habenula as a regulator of anxiety in adult zebrafish. *Front. Neural Circuits* **7**, 99 (2013).

47. Agetsuma, M. *et al.* The habenula is crucial for experience-dependent modification of fear responses in zebrafish. *Nat. Neurosci. 2010 1311* **13**, 1354–1356 (2010).

48. Chou, M. Y. *et al.* Social conflict resolution regulated by two dorsal habenular subregions in zebrafish. *Science* **352**, 87–90 (2016).

49. Wee, C. L. *et al.* A bidirectional network for appetite control in larval zebrafish. *Elife* **8**, (2019).

50. Barrios, J. P., Wang, W. C., England, R., Reifenberg, E. & Douglass, A. D. Hypothalamic Dopamine Neurons Control Sensorimotor Behavior by Modulating Brainstem Premotor Nuclei in Zebrafish. *Curr. Biol.* **30**, 4606 (2020).

51. Pose-Méndez, S., Schramm, P., Valishetti, K. & Köster, R. W. Development, circuitry, and function of the zebrafish cerebellum. *Cell. Mol. Life Sci.* **80**, 227 (2023).

52. Hiyoshi, K. *et al.* Two-photon laser ablation and in vivo wide-field imaging of inferior olive neurons revealed the recovery of olivocerebellar circuits in zebrafish. *Int. J. Environ. Res. Public Health* **18**, 8357 (2021).

53. Zhang, L. *et al.* Altered Volume and Functional Connectivity of the Habenula in Schizophrenia. *Front. Hum. Neurosci.* **11**, 636 (2017).

54. Xue, K. *et al.* Altered static and dynamic functional connectivity of habenula in first-episode, drug-naïve schizophrenia patients, and their association with symptoms including hallucination and anxiety. *Front. Psychiatry* **14**, 1078779 (2023).

55. Bernstein, H. G., Keilhoff, G. & Steiner, J. The implications of hypothalamic abnormalities for schizophrenia. *Handb. Clin. Neurol.* **182**, 107–120 (2021).

56. Li, J. *et al.* Hypoactivity of the lateral habenula contributes to negative symptoms and cognitive

dysfunction of schizophrenia in rats. *Exp. Neurol.* **318**, 165–173 (2019).

57. Walker, E., Mittal, V. & Tessner, K. Stress and the Hypothalamic Pituitary Adrenal Axis in the Developmental Course of Schizophrenia. *Annu. Rev. Clin. Psychol.* **4**, 189–216 (2008).

58. Andreasen, N. C. & Pierson, R. The Role of the Cerebellum in Schizophrenia. *Biol. Psychiatry* **64**, 81 (2008).

59. Kruse, A. O. & Bustillo, J. R. Glutamatergic dysfunction in Schizophrenia. *Transl. Psychiatry 2022 121* **12**, 1–13 (2022).

60. McCutcheon, R. A., Krystal, J. H. & Howes, O. D. Dopamine and glutamate in schizophrenia: biology, symptoms and treatment. *World Psychiatry* **19**, 15 (2020).

61. Väisänen, J., Ihalainen, J., Tanila, H. & Castrén, E. Effects of NMDA-receptor antagonist treatment on c-fos expression in rat brain areas implicated in schizophrenia. *Cell. Mol. Neurobiol.* **24**, 769–780 (2004).

62. Jahangir, M., Zhou, J. S., Lang, B. & Wang, X. P. GABAergic System Dysfunction and Challenges in Schizophrenia Research. *Front. Cell Dev. Biol.* **9**, 663854 (2021).

63. Rund, B. R. The research evidence for schizophrenia as a neurodevelopmental disorder. *Scand. J. Psychol.* **59**, 49–58 (2018).

64. Birnbaum, R. & Weinberger, D. R. Genetic insights into the neurodevelopmental origins of schizophrenia. *Nat. Rev. Neurosci.* **18**, 727–740 (2017).

65. Duman, J. G. *et al.* The adhesion-GPCR BAI1 regulates synaptogenesis by controlling the recruitment of the Par3/Tiam1 polarity complex to synaptic sites. *J. Neurosci.* **33**, 6974–6978 (2013).

66. Feuer, K. L., Peng, X., Yovo, C. K. & Avramopoulos, D. DPYSL2/CRMP2 isoform B knockout in human iPSC-derived glutamatergic neurons confirms its role in mTOR signaling and neurodevelopmental disorders. *Mol. Psychiatry* (2023) doi:10.1038/S41380-023-02186-W.

67. Zhang, S. *et al.* Multiple genes in a single GWAS risk locus synergistically mediate aberrant synaptic development and function in human neurons. *Cell genomics* **3**, (2023).

68. Broadway, B. J. *et al.* Systematic Functional Analysis of PINK1 and PRKN Coding Variants. *Cells* **11**, (2022).

69. Li, Y. *et al.* Regulatory Variant rs2535629 in ITIH3 Intron Confers Schizophrenia Risk By Regulating CTCF Binding and SFMBT1 Expression. *Adv. Sci. (Weinheim, Baden-Wurttemberg, Ger.* **9**, (2022).

70. Hoedjes, K. M., Kostic, H., Flatt, T. & Keller, L. A Single Nucleotide Variant in the PPARγ-homolog Eip75B Affects Fecundity in Drosophila. *Mol. Biol. Evol.* **40**, (2023).

71. Weng, Z. *et al.* Structure of BAI1/ELMO2 complex reveals an action mechanism of adhesion GPCRs via ELMO family scaffolds. *Nat. Commun.* **10**, (2019).

72. Hamanaka, K. *et al.* Large-scale discovery of novel neurodevelopmental disorder-related genes through a unified analysis of single-nucleotide and copy number variants. *Genome Med.* **14**, (2022).

73. Shiu, F. H. *et al.* Mice lacking full length Adgrb1 (Bai1) exhibit social deficits, increased seizure susceptibility, and altered brain development. *Exp. Neurol.* **351**, (2022).

74. Duman, J. G. *et al.* The adhesion-GPCR BAI1 shapes dendritic arbors via Bcr-mediated RhoA activation causing late growth arrest. *Elife* **8**, (2019).

75. Duman, J. G., Tu, Y. K. & Tolias, K. F. Emerging Roles of BAI Adhesion-GPCRs in Synapse Development and Plasticity. *Neural Plast.* **2016**, (2016).

76. Zhu, D. *et al.* BAI1 regulates spatial learning and synaptic plasticity in the hippocampus. *J. Clin. Invest.* **125**, 1497–1508 (2015).

77. Sleiman, P. *et al.* GWAS meta analysis identifies TSNARE1 as a novel Schizophrenia / Bipolar susceptibility locus. *Sci. Rep.* **3**, (2013).

78. Plooster, M. *et al.* Schizophrenia-Linked Protein tSNARE1 Regulates Endosomal Trafficking in

595 Cortical Neurons. *J. Neurosci.* **41**, 9466–9481 (2021).

596 79. Hancock, E. J., Ang, J., Papachristodoulou, A. & Stan, G. B. The Interplay between Feedback and
597 Buffering in Cellular Homeostasis. *Cell Syst.* **5**, 498-508.e23 (2017).

598 80. Liu, C., Everall, I., Pantelis, C. & Bousman, C. Interrogating the Evolutionary Paradox of
599 Schizophrenia: A Novel Framework and Evidence Supporting Recent Negative Selection of
600 Schizophrenia Risk Alleles. *Front. Genet.* **10**, (2019).

601 81. Akingbuwa, W. A., Hammerschlag, A. R., Bartels, M., Nivard, M. G. & Middeldorp, C. M. Ultra-rare
602 and common genetic variant analysis converge to implicate negative selection and neuronal
603 processes in the aetiology of schizophrenia. *Mol. Psychiatry* **27**, 3699–3707 (2022).

604 82. Bassett, A. S., Bury, A., Hodgkinson, K. A. & Honer, W. G. Reproductive fitness in familial
605 schizophrenia. *Schizophr. Res.* **21**, 151–160 (1996).

606 83. Avila, M., Thaker, G. & Adami, H. Genetic epidemiology and schizophrenia: A study of
607 reproductive fitness. *Schizophr. Res.* **47**, 233–241 (2001).

608

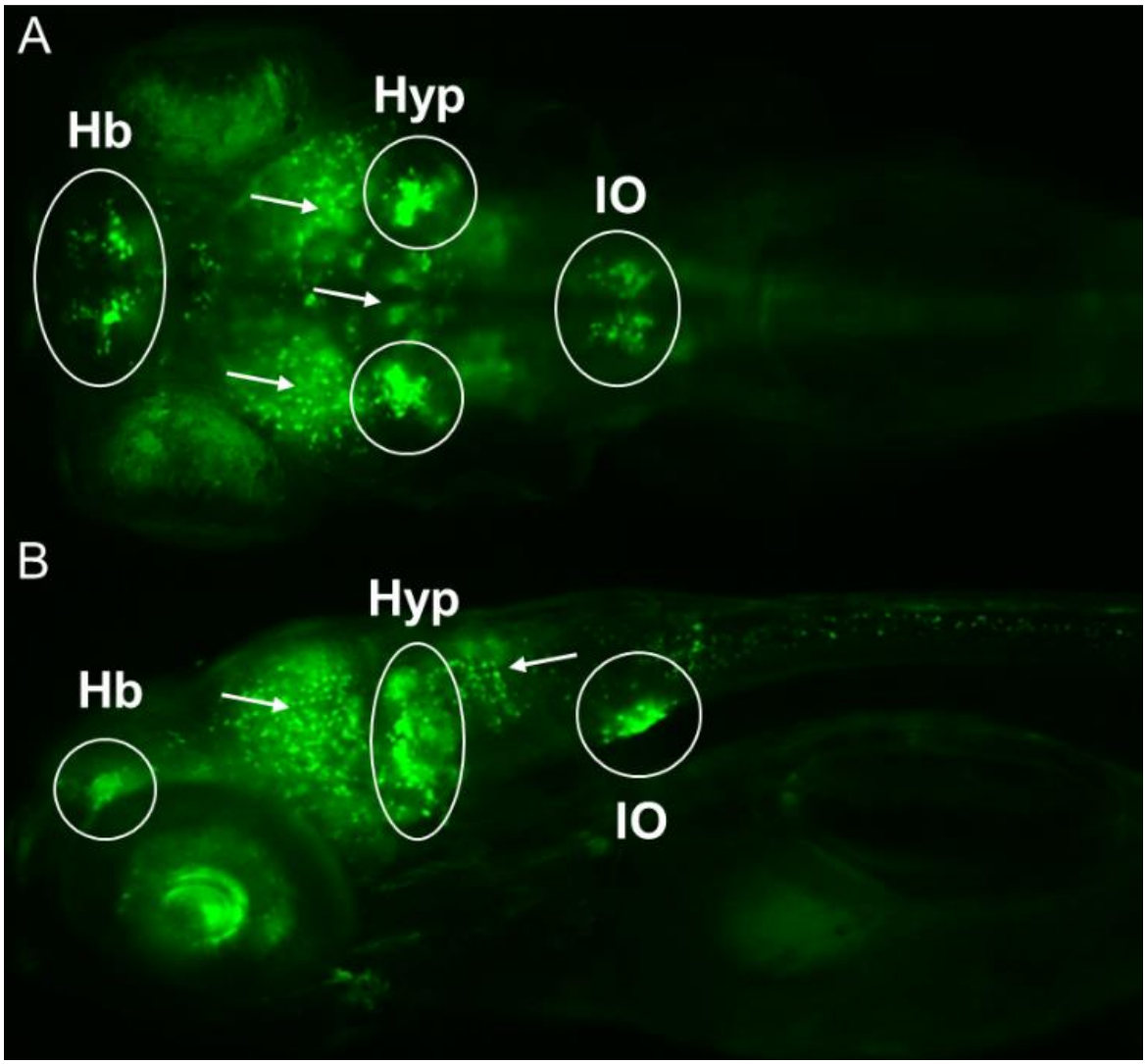

Figure 1. A noncoding REST/NRSF TF binding sequence containing rs4129585 drives reporter expression in F_2 zebrafish embryos 6 d.p.f. **(A)** Dorsal and **(B)** lateral views of a representative zebrafish embryo with GFP expression patterns in the habenula (Hb), hypothalamus (Hyp), and inferior olive (IO), as well as glutamatergic neurotransmitters (arrows).

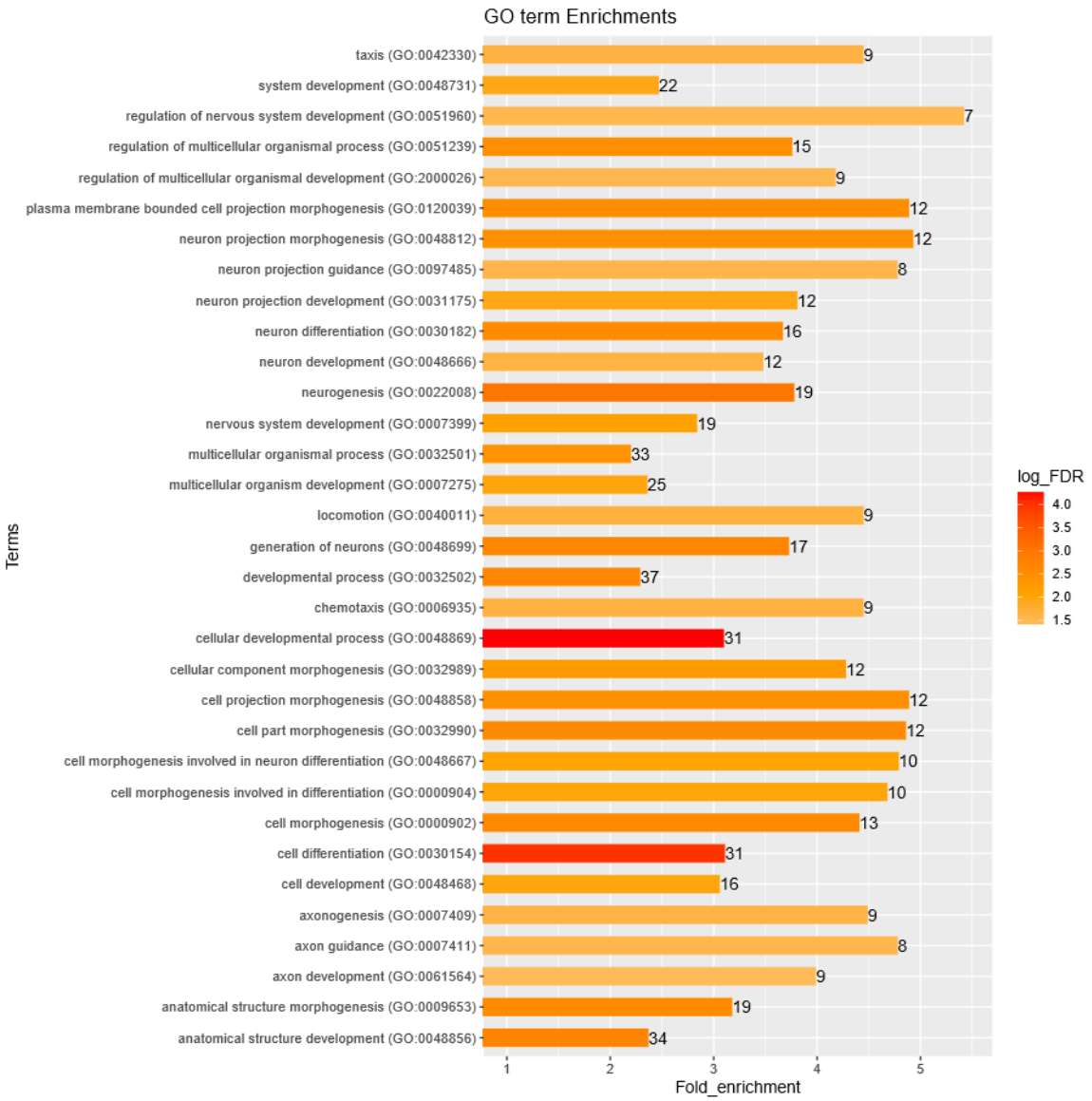

Figure 2. GO terms for functional enrichments among DE genes reported by PANTHER. Bar color indicates level of statistical significance (log_FDR; false discovery rate), numbers indicate the number of genes enriched in the pathway, and bar length indicates the fold enrichment of the genes observed in our uploaded list over the expected representation of genes in each GO category.

Figure 3. GO terms for functional enrichment within the STRING network of interacting proteins identified from our DR genes. Bar color indicates level of statistical significance, numbers indicate the number of genes enriched in the network, and bar length indicates the fold enrichment of the genes observed in our uploaded list over the expected representation of genes in each GO category.

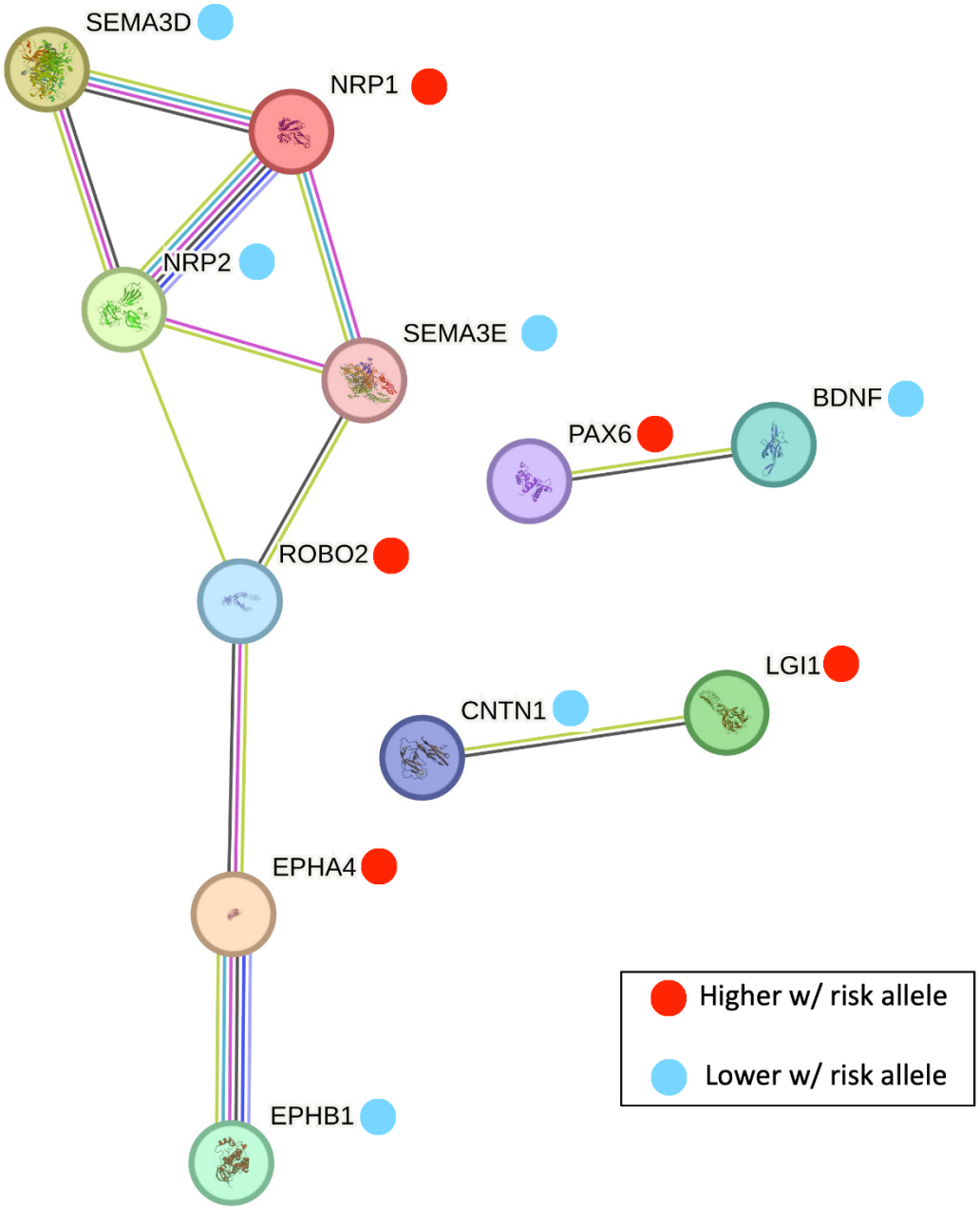

Figure 4. The 11 axon guidance genes in the STRING network of DE genes. Blue dots indicate proteins in the network whose genes are downregulated in the NPCs containing the risk allele, and red dots indicate proteins in the network whose genes are upregulated in the NPCs containing the risk allele. Image generated by the STRING website.