Single cell transcriptomic profiling of human brain organoids reveals developmental timing- and cell-type-specific vulnerabilities induced by *NRXN1* CNVs in schizophrenia

Rebecca Sebastian[1,2,#], Kang Jin[3,4,#], Narciso Pavon[2], Ruby Bansal[2], Andrew Potter[3], Yoonjae Song[2], Juliana Babu[2], Rafael Gabriel[2], Yubing Sun[5], Bruce Aronow[3,4,6,7], ChangHui Pak[2*]

[1]Graduate Program in Neuroscience & Behavior, UMass Amherst, Amherst, MA 01003 USA
[2]Department of Biochemistry and Molecular Biology, UMass Amherst, Amherst, MA 01003 USA
[3]Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA
[4]Department of Biomedical Informatics, University of Cincinnati, Cincinnati, OH 45229, USA
[5]Department of Mechanical and Industrial Engineering, UMass Amherst, Amherst, MA 01003 USA
[6]Department of Electrical Engineering and Computer Science, University of Cincinnati, Cincinnati, OH 45221, USA
[7]Department of Pediatrics, University of Cincinnati School of Medicine, Cincinnati, OH 45256, USA

#equal contribution
*Correspondence: cpak@umass.edu

Abstract

De novo mutations and copy number variations (CNVs) in *NRXN1* (2p16.3) pose a significant risk for schizophrenia (SCZ). How *NRXN1* CNVs impact cortical development in a cell type-specific manner and how disease genetic background modulates these phenotypes are unclear. Here, we leveraged human pluripotent stem cell-derived brain organoid models carrying *NRXN1* heterozygous deletions in isogenic and SCZ patient genetic backgrounds and conducted single cell transcriptomic analysis over the course of cortical brain organoid development from 3 weeks to 3.5 months. We identified maturing glutamatergic and GABAergic neurons as being consistently impacted due to *NRXN1* CNVs irrespective of genetic background, contributed in part by altered gene modules in ubiquitin-mediated pathways, splicing, and synaptic signaling. Moreover, while isogenic *NRXN1* CNVs impact differentiation and maturation of neurons and astroglia, cell composition and developmental trajectories of early neural progenitors are affected in SCZ-*NRXN1* CNVs. Our study reveals developmental timing dependent *NRXN1* CNV-induced cellular mechanisms in SCZ at single cell resolution and highlights the emergence of disease-specific transcriptomic signatures and cellular vulnerabilities, which can arise from interaction between genetic variants and disease background.

De novo mutations and copy number variations (CNVs) in 2p16.3 have been repeatedly observed in patients with autism spectrum disorders (ASDs), SCZ, and intellectual disability[1–5]. Albeit rare, these CNV losses, usually manifested in heterozygous deletions, present a significant increase in risk for multiple neuropsychiatric disorders[6]. Neurexin-1 (*NRXN1*), the single gene present in this locus, encodes a type I membrane cell adhesion molecule that functions as a synaptic organizer at central synapses[7]. NRXN1, as a presynaptic molecule, associates with multiple soluble and transmembrane molecules, thereby endowing specific synapses with unique synaptic signaling and transmission properties[8–13]. *NRXN1* also undergoes extensive alternative splicing, further enriching the diversity of these interactions[14–17]. Due to this pan-synaptic role throughout the brain, it is not surprising to find strong prevalence of *NRXN1* genetic lesions in multiple neurodevelopmental and psychiatric disorders. Most often, these lesions are large deletions (up to ~1Mb) affecting the long isoform NRXN1α specifically, as well as *NRXN1*α*/*β lesions affecting both the long and short isoforms. However, why or how the same *NRXN1* deletion results in phenotypically distinct disorders in individuals remains unknown. It is often hypothesized that the interaction between common variants (disease genetic background) and *NRXN1* CNVs drives these differences. Yet, experimentally demonstrating this hypothesis has been challenging.

Separate from its canonical function at synapses, which occurs post-neurogenesis, recent evidence suggests possible roles of *NRXN1* in early cortical development. In fact, *NRXN1* mRNAs are abundantly detected in early human embryonic neocortex, as early as gestational week (GW) 14, reaching peak at birth before slowly decreasing with age[18]. Knockdown of *NRXN1* in human neural progenitor cells (NPCs) results in decreased levels of glial progenitor marker GFAP, thereby potentially skewing the ratio of neurons to astrocytes[19]. A bi-allelic *NRXN1*α deletion in human iPSC-derived neural cells has been shown to impair maturation of neurons and shift NPC differentiation potential towards glial rather than neuronal fate[20]. More recently, *in vivo* CRISPR KO of *nrxn1* in *Xenopus tropicalis* embryos led to increased telencephalon size attributed to the increased proliferation of NPCs[21]. Separate validation using human cortical NPCs and iPSC-derived organoids showed increased proliferation of NPCs and an increase in neurogenesis in *NRXN1* mutants[21]. Though these studies provide some clues as to which roles *NRXN1* may play during early corticogenesis, the outcomes from these distinct models are inconsistent due to the differences in genetic lesions and the dosage of *NRXN1* being manipulated at different developmental time points. Therefore, it is worth investigating whether disease-associated *NRXN1* mutations in human cells lead to aberrant cortical development and differentiation of

neuronal populations in the cortex, thereby ultimately impacting cortical circuitry and synaptic function.

Human pluripotent stem cell (hPSC) derived brain organoids have been proven useful to model early developmental processes associated with neuropsychiatric disorders[22–28]. The self-organizing capability of hPSCs under directly guided differentiation produces relatively homogenous brain organoids, which can be maintained under defined conditions over long term[24,29]. By deriving cortical brain organoids from human induced pluripotent stem cells (iPSCs) representing a heterogeneous population of SCZ individuals, studies showed that there exists differentially regulated transcriptomic profiles[30], neuronal synaptic transmission defects[31], and early cortical maldevelopment[32]. More recently, brain organoids derived from idiopathic SCZ iPSCs exhibited reduced capacity to differentiate into neurons from NPCs[33]. Though these studies are promising and provide certain clues to the early brain developmental mechanisms of SCZ, how certain cell types during a continuous developmental time window are affected in such human cellular models of SCZ and how specific disease risk variants affect this process are unclear. More importantly, since genetic backgrounds often contribute to and modulate cellular phenotypes, understanding even how a single disease variant acts is difficult to dissect unless an isogenic mutant model is analyzed side by side with the patient model.

Here, we utilized a panel of hPSC lines, where *NRXN1* CNVs have been either artificially engineered (isogenic) or deleted genetically in individuals with SCZ (patient iPSCs) paired with controls. These cell lines, when differentiated into NGN2-induced cortical excitatory neurons, elicit reproducible synaptic impairment, i.e. decreased synaptic strength and probability of neurotransmitter release[34,35]. Using these extensively characterized hPSCs, we generated dorsal forebrain organoids with the goal of creating a comprehensive single cell atlas across developmental time and *NRXN1* mutation status. We generated a total of 141,039 high-quality single cell transcriptomes and performed an in-depth analysis on the neurodevelopmental impact of *NRXN1* CNVs in both isogenic and SCZ genetic backgrounds. We find that maturing glutamatergic and GABAergic neurons as being consistently impacted due to *NRXN1* CNVs irrespective of genetic background, contributed in part by altered gene expression programs in ubiquitin-mediated pathways, splicing, and synaptic signaling. In addition, while SCZ-*NRXN1* CNVs affect cell composition and developmental trajectories of early neural progenitors, isogenic *NRXN1* CNVs act at later stages of development influencing neuronal and astroglia differentiation. Ultimately, both isogenic- and SCZ-*NRXN1* CNVs impair neuronal network connectivity in maturing brain organoids. Our study shows developmental timing dependent *NRXN1* CNV-induced cellular mechanisms in SCZ at single cell resolution and highlights the importance of

disease-specific transcriptomic signatures, which can arise from the interaction between genetic variants and disease background.

Generation of forebrain organoids for modeling NRXN1 CNVs in neocortical development

We generated dorsal forebrain organoids as previously described, chosen for its reported homogeneity of the cellular constituents and reproducibility in disease modeling[31,36]. With minor modifications, including the use of Aggrewell plates to control the size of generated embryoid bodies and gentle agitation starting at day 6-8 to reduce spontaneous fusion (see Methods), patterned forebrain organoids showed expected developmental milestones as previously reported (Figs. 1, S1-2). In early time points, actively dividing proliferative ventricular zones (MKI67, SOX2) appeared, which decreased in abundance over the course of maturation (Figs. 1, S1-2). Starting at day 50 and well into day 100, the spatial organization of HOPX+ outer radial glia (oRGs), EOMES+ intermediate progenitor cells (IPCs), and BCL11B+ deeper layer and SATB2+ upper layer cortical neurons were detected (Figs. 1, S1-2). Moreover, the presence of S100B+ developing astrocytes and NEUN+ mature neurons were reliably detected at day 100 (Figs. 1, S1-2), indicative of active neurogenesis and the start of astrogenesis. At this time point and beyond, presynaptic markers (SYNAPSIN and SYNAPTOPHYSIN) and postsynaptic marker (HOMER) were also detected along MAP2+ dendrites (Figs. 1, S1-2), suggesting that the developing neurons are actively forming synapses in this organoid model.

Having established a reliable protocol, we then subjected a collection of hPSC lines, which have been extensively characterized[34,35], for brain organoid differentiation. Previously, we have generated iPSC lines from peripheral blood mononuclear cells (PBMCs) of SCZ patients carrying *NRXN1* heterozygous deletions (CNVs) and from unaffected, healthy control individuals, and showed that cortical excitatory induced neuronal cells (iN) differentiated from SCZ *NRXN1*^{del} iPSCs compared to controls displayed significant defects in synaptic strength and neurotransmitter release probability[35]. In addition, we have previously engineered a heterozygous *NRXN1* cKO allele in a control hESC (H1) genetic background[34], which reproduced the same neuronal phenotypes as the *NRXN1* cKO iPSC line and SCZ *NRXN1*^{del} patient iPSCs[35]. Thus, we chose to work with the engineered *NRXN1* cKO hESC line, which can conditionally create control and *NRXN1* cKO by the expression of Flp and Cre recombinases, respectively[34], and two sets of SCZ patient and matched control donor iPSC pairs (2 SCZ *NRXN1*^{del} lines and 2 control lines). In this experimental design, we aimed to investigate the cellular and molecular contributions of isogenic *NRXN1* CNVs (hereafter referred to as ‘engineered’) and SCZ-associated *NRXN1* CNVs

(hereafter referred to as 'donor') in the context of a developing human neocortex using a brain organoid model system.

Creating a single cell transcriptomic atlas from developing organoids with or without NRXN1 CNVs

All hPSC lines (engineered - 1 *NRXN1* cKO hESC line producing control (Flp) and cKO (Cre); donor - 2 control iPSC lines, and 2 SCZ *NRXN1*^{del} lines) normally developed into forebrain organoids, transitioning reproducibly through the major developmental cell types over time as visualized by the presence of key molecular markers (Figs. S1-2). We chose 3 distinct collection time points for 10X single-cell RNA sequencing (scRNAseq) - day 22/23, day 50, and day 101/112 - which captures the different pools of cell identities undergoing fate specification and maturation. First and foremost, we established a protocol that would allow us to reliably dissociate brain organoids into live single cells with >75% viability across time points (see Methods). After dissociation, approximately ~10,000 cells per sample were subjected to droplet based sequencing and rigorous QC data processing steps were performed (see Methods; Fig. S3). Subsequently, scRNAseq data was processed, normalized, and clustered to generate distinct cell clusters, which were further annotated manually using canonical markers (see Methods). A total of 29 cell clusters were annotated, which consisted of both cycling and non-cycling neural progenitors (NECs), outer radial glial cells (oRGs), intermediate cells and intermediate progenitor cells (IPCs) that give rise to distinct subpopulations of glutamatergic excitatory neurons (CNs) and GABAergic inhibitory neurons (INs; Fig. S6A). Non-neuronal cells, which mostly consisted of glial cells and astroglia, were also annotated (Fig. 1). Violin plots showing the expression of various canonical markers and their distributions on the UMAPs are shown in Supplementary information (Figs. S4-5, S6A). Here, we also provide an interactive visualization of 3D UMAP for further exploration (see Supplementary .html files).

To further validate our cell annotations, we compared gene expression signatures of our scRNAseq cell clusters to published brain organoid datasets as reference (Fig. S6B)[24,37–41]. Remarkably, we saw a high correlation between published cell clusters and ours, suggesting that the specific cell clusters in our scRNAseq dataset share similar gene expression patterns with other brain organoids that were generated with slightly different protocols, thus, showing overall reproducibility of 3D culturing across protocols. Altogether, we established a scRNAseq dataset consisting of high-quality 141,039 single cells from engineered and donor-derived organoids with or without *NRXN1* CNVs at 3 distinct time points for downstream analysis.

Single-cell NRXN1 mRNA expression patterns in the developing forebrain organoids and human neocortex

Using this scRNAseq dataset, we first wanted to analyze the cell type-specific expression of *NRXN1* in the control donor samples across organoid development, which would allow identification of specific cell types enriched for *NRXN1* function. We quantified the percentage of *NRXN1* mRNA expressing cells and found that CNs and INs reproducibly showed the highest expression of *NRXN1* across the three time points (Fig. 2A,B). Though not to the same degree as these cell types, IPCs, oRGs, and astroglia did express *NRXN1* at low levels to start (day 22) and progressively increased in expression over time (day 101). Lastly, NEC subtypes showed the lowest abundance of cells expressing *NRXN1* with <15 % of total cells reliably expressing the gene (Fig. 2A). We also used developmental trajectory analysis (monocle3[42]; see Methods) to quantify *NRXN1* expressing cell types across pseudotime (Fig. 2B). Similar to fixed time point analysis, cells with highest *NRXN1* expression included most mature CN and IN subtypes followed by IPC subtypes, oRGs, and astroglia (Fig. 2B). Based on this result, we concluded that the function of human *NRXN1* gene can be most reliably studied in differentiated neurons and astroglia, as well as in cortical progenitors, such as IPCs and oRGs, in the brain organoid model.

While this result highlights the important cell types for *NRXN1* function in organoid models, we were curious about *NRXN1* expression patterns in human primary tissue and how similar they are to brain organoids. To this end, we leveraged a published human fetal brain scRNAseq dataset, which reported single cell transcriptomes from second trimester microdissected tissues, representing a time period associated with peak of neurogenesis and early gliogenesis (14 GW to 25 GW)[40] and performed similar analysis. Comparable to brain organoids, at the youngest fetal age (14 GW), glutamatergic neurons showed the highest % of *NRXN1* expressing cells, and at 16 GW, both glutamatergic and GABAergic neurons consisted of the highest % of *NRXN1* expressing cells (Fig. 2C). These neuronal cells showed varying degrees of expression at older time points (20 and 25 GWs). Due to the low number of astrocytes represented in the dataset, astrocytes were not quantified here. In contrast, progenitor cells and forebrain radial glial cells showed a relatively low number of *NRXN1* expressing cells at 14 GW, which increased dramatically at 16 GW and caught up to similar levels to glutamatergic neurons and GABAergic neurons at 25 GW (Fig. 2C). Thus, based on both organoid and human tissue data, major cell types, which may be most affected by *NRXN1* haploinsufficiency in the developing forebrain, include glutamatergic neurons and GABAergic neurons in addition to progenitor cells and astroglia.

NRXN1 isogenic CNVs induce moderate changes in brain organoid maturation and gene expression

We first investigated the developmental timing- and cell type-specific effects of *NRXN1* CNVs in an isogenic control background. By using relative abundance visualization, we compared differential cell proportion effects in control vs. *NRXN1* cKO brain organoids over development (days 23, 50, and 112; Fig. S7). We found that, in the engineered brain organoids, there were no major changes in cell composition at early time points (days 23, 50). At all time points, there were no changes in NEC subtypes. At a later time point (day 112), the proportions of astroglia, IPC subtypes (IPC1-4), and differentiated neuronal subtypes (CN1/2/4/5, IN1-5) were altered (Fig. S7). We next investigated the overall developmental trajectory of these organoids by constructing single-cell trajectories in monocle3. By calculating the densities of cells across pseudo-time values, we drew density plots for cells in multiple time points, genotypes and brain organoid types, representing the dynamic cell abundance changes and cellular transitions throughout their developmental trajectory (Fig. S8, see Methods). As expected, using data from the three time points, we saw that the brain organoids underwent a progressive developmental trajectory that mirrored the corresponding maturity across pseudotime: early stage of trajectory corresponding to proliferating cells while later stage of trajectory corresponding to differentiated and mature neuronal subtypes (Fig. S8A,B). The developmental trajectories across time points mirrored this effect in cell composition where *NRXN1* cKO brain organoids followed similar developmental trajectories as controls at days 23 and 50 until reaching day 112, when there was a noticeable difference between the control and cKO (Fig. S8B). *NRXN1* cKO brain organoids displayed abnormal developmental trajectories during mature developmental stages (at longer pseudotime lengths), suggesting that the timing of cellular differentiation and maturation in brain organoids may be affected by isogenic *NRXN1* engineered CNVs. Importantly, at ~day 100-120, brain organoids reach the peak of neuronal diversity and amplification and the beginnings of astrogenesis[23,37,43], indicating that isogenic *NRXN1* CNVs may impact gene expression programs that regulate active neurogenesis, gliogenesis, and synapse development.

To understand differential gene expression patterns in *NRXN1* cKO organoids vs. controls, we performed analysis of differentially expressed genes (DEGs) (see Methods for DEG criterion) in each cell type associated with each time point. We identified 43, 538, and 486 DEGs at day 23, 50, and 112 respectively (Fig. 3, Table S1), which showed a modest perturbation effect at the transcriptional level overall. In day 112 engineered organoids, astroglia, glial cells, oRGs, CN3, CN4, and IN7 had a number of DEGs (Fig. 3B, Table S1), which correlated with the results from cell composition and trajectory analysis. There was minimal DEG overlapping patterns

among these DEG sets across cell types at each time point, as assessed by hypergeometric test, which was used to measure gene set associations (see Methods, Fig. 3C, Table S2). Using day 112 cell type-specific DEGs, we performed gene set enrichment analysis (GSEA, ToppGene[44]) to examine whether specific molecular functions, biological processes and/or biochemical pathways were significantly enriched (Fig. S9). Collectively, DEG sets from astroglia, glial cells, and oRGs were involved in neuron and glia development (Fig. S9, Table S3). All of these cell types also showed an enrichment of genes involved in cell cycle and programmed cell death (Fig. S9, Table S3). Interestingly, oRG-DEGs were enriched in RNA splicing, Ubiquitin (Ub)-dependent proteolysis, and WNT pathways (Fig. S9, Table S3). Notably, DEGs representing both glutamatergic excitatory neurons (CN3/4) and GABAergic neurons (IN7) were enriched for synaptic genes, as well as components of the Ub-mediated proteosome degradation (Fig. S9, Table S3). In GABAergic neurons, neuronal splicing factors were particularly enriched, including NOVA1, RBFOX2, SRRM1, and KHDRBS1 (Table S3). Altogether, GSEA suggests that regulators of RNA splicing and Ub-mediated proteasome pathway are consistently perturbed from oRGs to differentiated neuronal subtypes in the engineered brain organoids.

Composition and developmental trajectories of various cell types are affected in SCZ-NRXN1 ^{del} donor-derived brain organoids

We next compared differential cell proportion effects in control vs. SCZ *NRXN1* ^{del} donor brain organoids over development using scCODA (days 22, 50, and 101; Fig. S10), which allows quantification of cell composition changes using a Bayesian model[45]. In day 22 organoids, we did not observe major changes in the composition of various cell classes. However, in day 50 organoids, we captured a significant and uniform decrease in the ratios of NEC subtypes in *NRXN1* del samples, including cycling dorsal, cycling ventral, cycling and non-cycling NECs (Fig. S10). In parallel, we observed an increase in the number of astroglia, CN1 and IN2 neuronal subtypes (Fig. S10). Moreover, at day 101, *NRXN1* del donor organoids showed a selective decrease in the cycling NECs in addition to a decrease in the proportions of IPC1, IPC2, and CN2, while displaying an increase in various progenitors (oRG, IPC3, IPC4) and neuronal subtypes (CN3-5, IN2, IN6). Collectively, these results highlight that, in the developing *NRXN1*^{del} donor brain organoids, specific changes in cell composition are induced starting at day 50 until day 101, resembling precocious development of neural progenitors into differentiated neuronal subtypes. Using a relative abundance visualization, we saw an agreement with scCODA results (Fig. S10B), further validating changes in cell proportions in the donor-derived organoids.

We then sought to explore the effects of SCZ *NRXN1* CNVs on the overall developmental trajectory of various cell types in these brain organoids using monocle3 (Fig. S8C). Notably, we found a significant difference between *NRXN1*^{del} donor brain organoids and controls in their developmental trajectories at days 50 and 101, mirroring the cell composition changes (Fig. S8C). Moreover, the observed abnormal developmental trajectories are concentrated in the cell populations that occupy more mature developmental stages (later stage of trajectory), suggesting that SCZ-associated *NRXN1* CNVs induce most dominant effects on the developmental trajectories affecting neuronal differentiation and maturation.

Perturbations in transcriptional profiles across cell types in SCZ-NRXN1^{del} donor-derived brain organoids show greatest effects in mature time points

To investigate the developmental timing- and cell type-specific gene expression signatures that may actively contribute to abnormal developmental dynamics and cell lineage trajectories, we independently performed DEG analyses between genotypes across cell types at each fixed time point. Consistent with the findings where day 101 organoids showed the greatest perturbation effects in terms of cell composition and developmental trajectory, we found the largest number of DEGs (adjusted p-value (FDR)< 0.05, see Methods) across multiple cell types in the day 101 organoids (3105) compared to day 22 and 50 (1399 and 1094 respectively) (Fig 4; Table S4). Additionally, to better understand the contribution of specific gene modules associated with cell type-specific DEGs across each developmental time point, we explored whether there exist any overlaps between DEGs in different cell types using hypergeometric test (see Methods). We found that, at early time points (days 22 and 50), NECs, glial cells, and intermediate cells shared the most overlapping DEGs while IPCs and differentiated neurons showed minimal DEG overlaps (Fig. 4C, Table S2). However, at a later time point (day 101), we observed an overall increase in the number of overlapping DEGs shared between developmentally distinct and similar cell types across cell lineages (Fig. 4C, Table S2, S5), in agreement with this time point having most significant perturbations in cell composition and transcriptional changes (Fig. S10). For example, overlapping DEG patterns in CN and IN neuronal subtypes emerged as well as between differentiated neurons and progenitor populations (NECs, glial cells, and intermediate cells), indicating that specific changes in gene expression programs are being shared across neuronal lineages and cell types during active neurogenesis and gliogenesis.

Intersection of DEG overlaps points to Ub biology and RNA splicing as commonly disrupted molecular programs across cell types in SCZ-NRXN1 ^{del} donor-derived brain organoids

We further explored which DEGs and biological pathways make up the overlapping DEG patterns in day 101 *NRXN1*^{del} donor-derived brain organoids by performing GSEA (ToppGene). First, we separately analyzed the following overlapping DEG hotspots, organized into specific cell clusters: 1) neuronal cluster (CN1/2/3, IN4/5), 2) non-neuronal cluster (NECs/astroglia/glial cells), and 3) all cluster (neuronal and non-neuronal) (Figs. 4C, S11). Intriguingly, while each cluster showed specific enrichment of biological processes relevant to each cell type (Table S6), we identified two distinct biological processes that were repeatedly observed across clusters - Ub-mediated proteolysis and RNA splicing. As mentioned above, these two pathways were also disrupted in engineered brain organoids (Table S3). Specifically, factors involved in protein turnover and Ub-mediated proteolysis were identified as DEGs in the 'all cluster' (*NDFIP1, SKP1, SUMO1/2, UBB,* and *GABARAPL2*) and in the 'neuronal cluster' (*UBE2M, UBE2V2, RNF7, STUB1*, and *UCHL1*) (see Table S5 for a complete list). These genes encode for proteins that are either direct binding partners to or are themselves E3 Ub ligases (*NDFIP1, SKP1, RNF, STUB1*), E2 conjugating enzymes (*UBE2M, UBE2V2*), a deubiquitinase (*UCHL1*), and are associated with Ub processing (*UBB*) and autophagy *(GABARPL2*). Though it is not clear from the list of DEGs whether or not these molecules are actively participating in the protein quality control or in the regulation of protein components involved in signal transduction, it has been hypothesized that alterations in proteostasis and Ub-mediated regulation of synaptic signaling contribute to SCZ pathogenesis[46–50]. Additionally, protein truncating variants of Ub ligases (*CUL1* and *HERC1*) were recently found to be associated with SCZ at exome-wide scale[51], further highlighting the importance of this molecular pathway in SCZ.

Interestingly, cluster-specific enrichment of DEGs encoding splicing factors was observed in the 'non-neuronal cluster' (*SRSF6, SAP18, U2SURP, HNRNPA2B1, PSIP1,* and *SNRPG*) and 'neuronal cluster' (*SRSF3, YBX1, HNRNPA1, RBM39,* and *SF1*) (see Table S5 for a complete list). These genes are components of the catalytic spliceosome, splicing factors, or regulators of alternative splicing. HNRNPU was the only overlapping splicing factor identified in 'all cluster' which implies that the effects seen here could be reflective of cell type-specific regulation of alternative splicing[52,53]. In fact, there are distinct splicing factors that are differentially expressed in glutamatergic neurons (CN1/2/3) vs. GABAergic neurons (IN4/5) (Tables S5, S6). These results correlate with previous findings which report global changes in alternatively spliced

transcriptomes in post-mortem brains of individuals with neuropsychiatric disorders including SCZ[54,55].

Dysregulated NMDAR signaling as a common neuronal mechanism across genetic backgrounds

We separately examined neuronal-specific DEG overlaps that were exclusive to cortical excitatory and GABAergic inhibitory neuronal types (CN1/2/3, IN4/5; neuronal cluster) and absent in the non-neuronal (NEC/astroglia/glial) cluster in the donor-derived brain organoids (Table S5). Importantly, we identified the N-methyl-D-aspartate receptor 2B subunit (*GRIN2B*) and fatty acid binding protein 7 (*FABP7*), among others, both of which have been linked to SCZ pathogenesis[56–60]. It is well documented that NMDAR hypofunction underlies SCZ pathology[61,62] and recent exome sequencing and GWAS studies identified the NMDAR subunit *GRIN2A* as a significant SCZ risk allele[63]. Genetic variants in *FABP7* have been identified in SCZ and ASD patients and its function has been linked to NMDAR signaling regulation[56–58,64]. Furthermore, while searching for neuronal-specific DEGs that were consistently perturbed between engineered and donor derived organoid types, we found *GRIN2B* as a commonly perturbed gene in GABAergic neuronal subtypes across genetic backgrounds (engineered IN7 vs. donor IN4/5; Table S7, Fig. S12). These findings suggest that misregulated NMDAR signaling in GABAergic neurons could potentially impact synaptic connectivity and signaling in these brain organoid models.

Differential enrichment of disease associations in NRXN1^{del} donor vs. NRXN1 cKO engineered brain organoid DEGs

To test whether up- and down-regulated DEGs identified from donor-derived and engineered brain organoids at mature time points (days 101, 112) were associated with specific neuropsychiatric disease gene signatures, we computed ‘disease enrichment’ score (-log10 (FDR-adjusted *p* values)) based on a previously established curated list of SCZ, bipolar disorder (BD), major depression disorder (MDD), and ASD-associated genes[65]. Excitingly, in day 101 donor-derived organoids, we observed the strongest enrichment of up-regulated DEGs from CN and IN subtypes (CN1/2, IN4/5) in ASD and SCZ-related gene sets (Fig. 5A). Next, dividing and non-dividing NECs, as well as glial cells showed significant enrichment in SCZ-related gene sets, all of which were up-regulated (Fig. 5A). Interestingly, there was no significant enrichment of the DEGs in MDD and BD-related gene sets across the cell types, suggesting that the DEG pool from SCZ *NRXN1* del organoids most closely resembles dysregulated transcriptional signatures related to SCZ and ASD, similar to what has been reported regarding shared genetic signals

between SCZ and ASD[51]. On the other hand, in isogenic engineered brain organoids, down-regulated DEGs from day 112 glial cells and oRGs showed a modest but significant enrichment in ASD and SCZ-associated gene sets (Fig. 5A).

Next, we independently examined the enrichment score of rare and common variants of SCZ in the DEG sets, by comparing them to a list of risk genes recently reported by SCZ-GWAS PGC wave3 and SCHEMA consortium[51,63]. Remarkably, several of the SCZ risk genes were represented across cell types in the donor derived brain organoids (Fig. 5B), which highlights SCZ-specific transcriptional signatures present in the patient genetic backgrounds. Engineered brain organoid DEGs showed minimal to no overlap with SCZ-associated risk variants, clearly demonstrating the absence of disease gene signatures in an isogenic background (Fig. 5B).

Perturbations in neuronal network connectivity in NRXN1^{del} donor-derived and NRXN1 cKO brain engineered brain organoids

To test whether the observed developmental abnormalities and gene expression programs translate to functional and sustained differences in neuronal activity, we performed live Ca^{2+} imaging in both donor derived brain organoids (*NRXN1*^{del} brain organoids compared to controls) and engineered brain organoids (*NRXN1* cKO brain organoids compared to controls) at day 130-160 (Fig. 6). Under normal conditions, without stimulation, we measured the frequency and amplitude of spontaneous Ca^{2+} transients, which are indicative of spontaneous neuronal network activities. In addition, we quantified the frequency of synchronous firing, which indicates how often neurons fire together, thereby producing synchronized bursts of activities. Compared to controls, *NRXN1*^{del} brain organoids showed a significant decrease in the frequency of spontaneous Ca^{2+} transients without a change in the amplitude of the responses, as measured by dF/F_0 intensity (Fig. 6A,B). In addition, there was an overall decrease in the synchronous firing rate in these brain organoids, demonstrating a significant decrease in the neuronal network bursts (Fig. 6A,B). Interestingly, *NRXN1* cKO brain organoids produced a slightly different phenotype in which the frequency of spontaneous Ca^{2+} transients was increased without any changes in the amplitude of the responses as well as the synchronicity of spontaneous firing events (Fig. 6C,D). These data suggest that although spontaneous neuronal activities are uniformly altered in the brain organoids carrying *NRXN1* CNVs, depending on the genetic background, different phenotypic outcomes manifest, reflective of the differences in transcriptomic landscape of these brain organoids.

Discussion

Here we provide a systematic analysis of the cell-type- and developmental timing-dependent perturbations induced by *NRXN1* CNVs in the developing human brain organoids using single cell transcriptomics. We initially had two specific goals in mind – 1) to understand the developmental effects of *NRXN1* heterozygous deletions in an isogenic background to uncover which time points and cell types are important for *NRXN1* function, and 2) to utilize SCZ-*NRXN1* del patient iPSC derived organoids as a model to study the molecular and cellular biology of SCZ. By profiling the transcriptomes of *NRXN1* cKO brain organoids, we found that cellular phenotypes associated with *NRXN1* haploinsufficiency manifests at a developmental window of brain organoids at the peak of neurogenesis and start of astrogenesis. Moreover, developmental trajectories and gene expression profiles of maturing glutamatergic and GABAergic neurons are impacted by *NRXN1* CNVs.

By comparing engineered and donor-derived organoids side by side, we found both commonalities and differences, reflective of the contribution of genetic background effects. We unbiasedly found shared molecular programs that are perturbed across organoid types. First, the NMDAR subunit *GRIN2B* was differentially expressed in GABAergic neurons across genetic backgrounds. This finding was significant as it confirms previous results showing that *Nrxn1* signals through NMDARs[9,10] and *NRXN1* haploinsufficient human induced neurons carry upregulated levels of the endogenous NMDAR antagonist *KYAT3*[35]. Moreover, genetic variants in the NMDAR subunits, *GRIN2B* and *GRIN2A*, are both observed in SCZ populations[63,66,67]. Lastly, NMDAR hypofunction in SCZ has been a longstanding hypothesis supported by multiple post-mortem studies and brain imaging studies from SCZ patients as well as mouse models of NMDAR blockade through ketamine and phencyclidine[61,62].

Second, through the interrogation of DEG overlaps and GSEA, we discovered two distinct biological pathways that are enriched across genetic backgrounds – splicing and Ub-proteasome system (UPS) regulation – both of which have been previously implicated in SCZ. Alternative splicing is highly regulated in the brain[68,69] and is influenced by development-specific splicing factors like NOVAs, PTBPs, RBFOXs, and SSRMs[70–73]. It has been shown that global splicing changes and alternative transcript usage are overrepresented in SCZ brains, more so than in ASDs and BD[54]. Differential splicing of various genes has been observed in the brain samples of SCZ patients compared to controls, including *DRD2, NRG1, ERBB4, GRM3,* and *GRIN1*[54,74–77]. More recently, differential splicing effects of *NRXN1* has been appreciated in SCZ iPSC-derived neurons[78] and in postmortem brains of SCZ and BD patients[18,54], further highlighting the

importance of splicing regulation in SCZ as a potential molecular mechanism. It may be possible that there exists a systematic problem of the splicing machinery, which results in global changes in alternative splice usage in SCZ.

In addition to splicing, DEGs responsible for UPS regulation have been identified in our study. Importantly, post-mortem brain tissues from SCZ individuals compared to controls showed increased ubiquitin immunoreactivity[47], an increase in Lys63-linked Ub species, increased polyubiquitinated protein levels, and increased brain protein insolubility[49,50]. These observations are further strengthened by the recent discovery of protein truncating variants in the Ub ligases, *HERC1* and *CUL1*, in SCZ exomes[51]. Possibly more UPS genes are to be discovered for both rare and common variants in the future. Though it remains to be determined whether these changes are indeed causal or merely reporting a consequence of the disease, altered UPS does exist and this, in turn, could affect protein homeostasis in the brains of SCZ patients. Importantly, proteasome function at synapses is tightly regulated by NMDAR activity, as NMDAR activation regulates 26S proteosome assembly and catalytic activity[79,80] and stability of proteasomes in the post-synaptic density[81]. Moreover, E3 Ub-ligases and deubiquitinases act in concert to regulate the ubiquitination, internalization and localization of NMDARs, AMPARs, and mGluRs, and therefore, actively participate in Hebbian and homeostatic plasticity[82–84]. Further investigations on the interplay between NMDAR signaling and UPS regulation during synaptic development would enhance our understanding of how these distinct biological pathways converge in the context of SCZ pathogenesis.

Interestingly, there were two major differences between donor vs. engineered organoids that we observed. First, unlike the donor derived organoids, engineered organoids did not exhibit any changes in the cell proportion or gene expression in the NEC subtypes. Moreover, the magnitude of gene expression changes in various cell types in the engineered organoids was minimal compared to donor derived organoids, which showed a greater number of DEGs overall. These findings indicate that brain organoids derived from patient genetic background induce a greater degree of transcriptional perturbations and uncover NECs as a vulnerable cell type during cortical development in addition to neuronal subtypes and glial/astroglial cells, which are commonly affected to varying degrees in both patient and engineered genetic backgrounds carrying *NRXN1* CNVs. The specificity of NEC phenotypes in the SCZ *NRXN1* del genetic background is further supported by previous studies reporting alterations in the morphology, differentiation potential, and gene expression profiles from SCZ iPSC derived NECs and brain organoid models, all of which are in support of the neurodevelopmental hypothesis of SCZ[85–87].

Second, the enrichment of disease associated DEGs was minimal in the engineered organoids compared to donor derived organoids. This finding makes sense since donor derived organoids carry SCZ-relevant genetic background. Due to this effect, differences in the magnitude of gene expression changes and the directionality of those changes were observed in these organoid types. This is also apparent in the differences in the specific neuronal firing patterns observed in the isogenic engineered vs. donor derived organoids.

There are two main limitations to this study. First, despite obtaining high-quality dataset, the overall study is underpowered due to the small sample size of patient/control cohort with limited genetic backgrounds being represented. Changes in the cell proportion, for example, could be further analyzed using larger sample size and multiple technical replicates. It is unclear how cell proportion could be initiated earlier on as answering this question would require granular analysis of multiple time points across development and a greater number of single cells. Second, while we focused on early developmental time points leading up to days 101/112, which allows investigation of the molecular programs underlying peak of neuronal diversity and amplification, older time points could reveal postnatal gene signatures that are being missed here. For example, astrocytes/glial cells are prominent cell types that are shared among engineered and donor derived organoids and produce changes in gene expression and cell proportions. This finding could be further explored using older organoid samples, as astrocyte development is initiating at ~day 100 and requires long term cultures to study their biology[88]. In addition, the developmental switch from *GRIN2B*- to *GRIN2A*-containing NMDARs occurs at ~300 day old brain organoids[29], which could potentially allow one to study postnatal human brain biology.

In the future, it will be important to expand upon this work by comparing this dataset with single cell transcriptomes obtained from *NRXN1* del individuals with other neuropsychiatric disorders like ASDs as well as those from healthy, unaffected individuals who also carry *NRXN1* deletions. This type of experimental design would allow dissection of the contribution of disease-specific effects at a greater scale – common vs. distinct molecular features across neuropsychiatric disorders which uniformly affect brain development and synaptic function.

Acknowledgments

We thank Kelly Rangel (CCHMC gene expression core) and Dr. Jim Chambers (IALS Light Microscopy core) for assistance with 10X scRNAseq and Ca^{2+} imaging set up as well as members of the Pak lab for experimental assistance and helpful discussions. We also thank Dr. Zhiping Pang for sharing the psychiatric risk summary gene list for disease enrichment analysis.

Funding

This work was supported by NIMH (R01 MH122519 to C.P., R21 MH130843 to Y.S. and C.P.), UMass IALS/BMB faculty start up fund (to C.P.), Tourette Association of America (Young investigator award to C.P.), and NIGMS T32 BTP training program (T32 GM135096 to N.P.).

Author contributions

R.S., R.B., Y.J.S., and J.B. cultured brain organoids and performed experiments. K.J. carried out all scRNAseq data analysis. R.S. and N.P. conducted live Ca^{2+} imaging and analysis. A.P. and R.B. optimized single cell dissociation protocol. R.G. optimized image analysis. R.S., K.J., Y.B.S., B.A., and C.P. designed the experiments and R.S., K.J., and C.P. wrote the manuscript.

Declaration of interests

Nothing to declare.

Online methods

hPSC culture and forebrain organoid generation

hESC and iPSCs were cultured on feeder-free conditions as previously described[34,35]. In order to form serum-free floating embryoid body (EB) aggregates, hPSCs were dissociated into single cells using Accutase (Innovative Cell Technologies). Dissociated cells were then reaggregated in low adhesion microwell culture plates (AggreWell-800, Stem cell technologies). 3 million cells were plated per 2mL well in mTesR plus Y-27632 (1 μM, Axon Medchem). 24 hours after plating, aggregated EBs were transferred to ultra-low attachment 10cm petri dishes and cultured as previously described[89]. Briefly, EBs in 10 cm dishes are cultured in E6 medium (ThermoFisher) for up to 6 days with SB431542 (10 μM, Peprotech) and dorsomorphin (5 μM, Peprotech) for dual SMAD inhibition, promoting neural stem cell differentiation. Following day 6, EBs were cultured in Neurobasal (ThermoFisher) containing B27 without vitamin A supplement, Glutamax (Life Technologies), Penicillin-Streptomycin (ThermoFisher) and the following morphogens at 20 ng/mL (Peprotech): human EGF, human FGF, human BDNF, and human NT3. At day 6-8, forebrain organoids were placed on an orbital shaker for gentle agitation to reduce spontaneous fusion. Starting at day 43, all morphogens were removed and brain organoids were cultured solely in B27 containing Neurobasal media.

Lentivirus generation

Lentiviral plasmid constructs used in this study are Cre-recombinase and Flp-recombinase fused to EGFP driven by the ubiquitin-C promoter as previously described[35]. For all lentiviral vectors, viruses were produced in HEK293T cells (ATCC, VA) by co-transfection with three helper plasmids (3.25 µg of pRSV-REV, 8.1 µg of pMDLg/pRRE and 10 µg of lentiviral vector DNA per 75 cm^2 culture area using calcium phosphate transfection method[90]. Lentiviruses were harvested from the medium 48 hrs after transfection. Viral supernatants were then centrifuged at a high speed of 49,000 x g for 90 min and aliquoted for storage in -80C. Viral preparations that yielded 90% EGFP expression were assessed to be efficiently infected and used for experiments.

Cryopreservation and sectioning

Organoid samples were collected at day 21, 50 and 100. Samples were fixed in 4% paraformaldehyde at 4°C overnight then submerged in 30% sucrose/PBS solution for 24-48hrs in 4°C. Organoids were flash frozen in gelatin solution (gelatin in 10% sucrose/PBS) using dry ice/ethanol slurry and were stored in -80°C for long term storage or until cryosectioning. Cryosections were between 12 to 25 micron section thickness. Organoid sections were directly adhered to microscope slides and subsequently used for immunohistochemistry or stored for long term storage in -20°C.

Immunostaining

Organoid sections were washed three times in 0.2% Triton-X in PBS (0.2%PBS/T) and then blocked in 10% normal goat serum diluted in 0.2%PBS/T (blocking solution) for 1hr at room temperature. Sections were incubated in primary antibodies diluted in blocking solution overnight at 4°C and were subsequently washed three times with 0.2%PBS/T, followed by incubation with secondary antibodies and DAPI diluted in PBS/T at room temperature for 2 hours. Finally,

sections were washed three times (20 minutes per wash), and then mounted using Fluoromount mounting media (Southern Biotech). Primary antibodies used are as follows: mouse anti-Ki67 (1:250, BD Biosciences BDB550609), rabbit anti-SOX2 (1:500 Cell Signaling 3697S), rabbit anti-HOPX (1:500, Proteintech 11419-1-H), rat anti-CTIP2 (1:2000, Abcam ab18465), rabbit anti-TBR2 (1:1000, Abcam ab23345), mouse anti-SATB2 (1:1000 Abcam ab51502), mouse anti-NEUN (1:500, EMD Millipore MAB377), rabbit anti-NEUN (1:1000, EMD Milipore ABN78), rabbit anti-S100B (1:1000, Sigma S2644), chicken anti-MAP2 (1:5000, Abcam ab5392), rabbit anti-SYNAPTOPHYSIN (1:1000, Abcam ab14692), rabbit anti-HOMER (1:1000 Synaptic System 160003), and mouse anti-SYNAPSIN (1:500, Synaptic System 111011). Secondary antibodies conjugated with Alexa 488, 594, 647 (Invitrogen) and DAPI (1:1000, Sigma MBD0015) were used.

Calcium imaging and analysis

Organoids were incubated in 1 μM of X-Rhod1 AM dye (Invitrogen) diluted in a modified HEPES buffer (130mM NaCl, 5mM KCl, 2mM CaCl2, 1mM MgCl2, 10mM HEPES, 10mM Glucose, pH 7.4 adjusted with NaOH) for 15 minutes at room temperature. Excess dye was washed with modified HEPES buffer once, then imaged using a confocal microscope (Nikon, A1R25). Imaging was carried out in glass bottom petri dishes (MatTek). Temperature was maintained at 37C using the Ibidi stage heater. Time lapse images were acquired at 250ms intervals for a period of 5 mins. Images were processed using ImageJ software to produce binary images. Analysis was then carried out using a stimulation-free Matlab protocol as demonstrated previously[91]. Using the MATLAB protocol, we first 'stacked' the time lapse images captured by the confocal microscope to produce a Maximum Intensity Projection (MIP). Guided by the MIP, we then selected 4-5 regions of interest (ROI) indicating the most active regions of the organoid, with each ROI measuring 50μm in diameter. Next, we selected the time interval which is determined by the recording duration (in seconds)/frames; in our recordings we used 300/109 for a time interval of 2.75. Changes in image intensity within ROI's are then quantified and plotted as raw calcium traces. Using the average calcium intensity across all ROI's in one field of view, synchronous spikes were plotted and a synchronous firing rate was determined using the number of detected synchronous spikes every minute. Frequency was determined by the total number of detected peaks every minute across all traces. Amplitude was established using the mean value of □F/F0 from individual peaks.

Quantification and statistical analysis

Data wrangling was performed in Microsoft Excel, and all raw data points were transferred to Prism (9.3.0) for basic statistics, outlier detection, significance tests, and graph generation. To identify outliers from pooled replicates, the ROUT outlier test was used to identify outliers by fitting data with nonlinear regression and using a false discovery rate of Q=1%. An unpaired parametric two-tailed Student's t test was performed to compare the two genotypes (CTRL vs. cKO or CTRL vs. NRXN1del) for statistical significance.

Live single cell dissociation

Organoids were rinsed 3 times with HBSS (10X HBS salt, 1M HEPES, 0.004M NaHCO3 diluted to 1X), then minced into small pieces and transferred to a 15mL conical tube for incubation in digestion solution (consisting of HBSS, 1 mg/mL Papain, 0.5mM EDTA, and 1mM L-cysteine) for 15 minutes at 37°C. Upon incubation, digestion mixture containing organoids were gently triturated with DNase I (25 µg/mL, Worthington-Biochem) and subsequently incubated again for another 10 minutes at 37°C followed by filtration with 70uM and 30uM filters (Miltenyi Biotech). Cell mixture was then centrifuged and pelleted. Cell pellet was resuspended in Neurobasal media with B27. This step was done to help dilute any remaining enzyme, EDTA, and other components of the digest mix. After the final re-suspension in Neurobasal media (without supplements), single cell mix was filtered again using 40 µM FlowMi pipet (Milipore Sigma) to help remove debris.

10X scRNASeq Protocol

Following isolation of single cells from brain organoids, cells were centrifuged at 300 g for 5 min and then re-suspended in 1 mL ice-cold Neurobasal media. Cell concentration and viability were determined using a hemocytometer with trypan blue dye exclusion and cell concentrations were adjusted to 700-1200 cells/µL for 10X single cell sequencing. For each sample, 9,600 cells were loaded into the 10X Chromium controller to target recovery of 6,000 cells and a Gel Beads in Emulsion (GEM) was generated. 10X Genomics 3'v3.1 chemistry was used. The samples were processed according to the protocol from 10X Genomics, using 14 cycles for cDNA amplification. Single cell libraries were sequenced using the Illumina NovaSeq 6000.

Single cell data alignment

10x single-cell RNA-sequencing data in Fastq files were aligned to transcripts using Cell Ranger 3.1.0 (https://www.10xgenomics.com/support/single-cell-gene-expression). Reference genome GRCh38 (Ensembl 93) was used as the reference genome. In the CellRanger *count* command, parameters *chemistry* and *expected-cells* were set as SC3Pv3 and 6000, respectively.

Single cell preprocessing and normalization

Cell Ranger output h5 files were loaded using Seurat 4[92] as the raw data. To reduce the impact of low-quality cells, we first removed cells with less than 1200 or more than 25,000 unique molecular identifiers (UMI). In addition, we removed cells with less than 600 or more than 6000 unique genes. Since low-quality or dying cells often exhibit extensive mitochondrial contamination, we removed cells with more than 10% mitochondrial transcripts.

Furthermore, we removed several clusters (details of the clustering will be mentioned later) of cells with low-sequencing depth to avoid the influence of poorly sequenced cells. Clusters with lower-than-normal distributions of the number of UMIs or unique genes were manually removed from the data. In the end, 9 lowly-sequenced clusters were removed from both donor-derived and engineered organoids single cell data (Figure S3). In addition, we utilized Scrublet[93] investigate the doublets in the data. Only a small number of cells reached the threshold of doublets, indicating a low prevalence of doublets.

After the quality control, we finally harvested 33,538 genes and 141,039 high-quality cells, including 88,623 cells from 16 donor-derived organoid samples and 52,416 cells from 10

engineered organoid samples. Both original and processed data can be found in Data Availability. We normalized the total UMI counts per gene to 1 million (CPM) and applied $\log_2$(CPM+1) transformation for heatmap visualization and downstream differential gene expression analysis, which were conducted in Scanpy[94]. In the following Seurat integration procedure, we applied the default normalization approach of Seurat.

Single cell integration

To reduce the influence of batch effects from multiple samples in single-cell data analysis, we applied the Seurat integration procedure to the data. We first loaded the raw data of each sample separately and created a list of Seurat objects after the quality control. Then we normalized each Seurat object and found the top 2,000 highly variable genes using the "vst" method in *FindVariableFeatures* function. 2000 integration features were selected from the series of Seurat objects using *SelectIntegrationFeatures*. Then integration features in each dataset were scaled and centered using the *ScaleData* function, based on which we ran the Principal Component Analysis (PCA) to reduce the high dimensions of features into 50 principal components.

In this study, we used the Reciprocal PCA (RPCA) procedure as the default method of integrating our large-scale data due to its high computational performance. We first identified integration anchors with previously identified integration features and top 30 reciprocal principal components, after which we ran the integration using the function IntegrateData with the top 30 dimensions for the anchor weight procedure. After integration, the data was scaled and PCA was conducted using *ScaleData* and *RunPCA* functions, respectively. Then the nearest neighbor graph was constructed using 20 k-nearest neighbors and 30 principal components. Louvain clustering was applied on the neighbor graph using the function *FindClusters* and multiple resolutions (0.5, 1.0, 2.0) were used to find clusters in both coarse and fine resolutions for comprehensive downstream analysis. Additionally, 2-dimensional and 3-dimensional embeddings of cells were generated using Uniform Manifold Approximation and Projection (UMAP) based on top 30 principal components. These integration, clustering, and dimensionality reduction procedures were applied to cells from donor-derived organoids, engineered organoids, as well as cells from both types of organoids.

Cell Annotations

After the quality control and integration procedure, we got high-quality cells and clusters in multiple resolutions. A total of 49 clusters in a fine resolution (2.0) were generated from aforementioned procedures for cells from both donor-derived and engineered organoid samples. Canonical markers from previous studies were collected and used for manual annotations of each cluster, such as VIM for neural progenitor cells, STMN2 for neurons, and AQP4 for astrocytes. In addition, enrichment results of ToppCell-derived gene modules and prediction labels from reference datasets were used as supplementary evidence of annotations as well. A total of 29 cell classes were derived eventually, including subpopulations from NEC, glia cells, intermediate cells, neurons and supportive cells. Cluster 9 and 24 were labeled as unknown cells since there were no clear associations with known cell types based on marker genes or predicted cell types. Two clusters, including cluster 24 and 34, were labeled as low-sequencing-depth cells since their lower-than-normal transcript abundance levels. To focus on neuron differentiation, low-

sequencing depth unknown cells, microglia cells, and mesenchymal cells were not included in the downstream analysis.

Logistic regression for label prediction

To better understand the cell identities of clusters, we built up simple logistic regression models in the reference single cell data to predict cell type annotations in our own data. Such models were previously used in by Young et al.[95] to infer the similarity between kidney tumor cell populations and known normal kidney cell types. In our study, we established logistic regression models as classifiers for each cell type in 4 public brain organoid single cell datasets and 1 fetal brain single cell dataset. The prediction scores from the models were used to classify whether one query single cell belongs to a specific cell type. We applied models of all cell types from reference data to each cell in our single cell data and calculated the average prediction scores of cell types or clusters. The results represent the association or similarity between reference and query cell types (Figure S6B).

Differential expression analysis

In our study, we used the Wilcoxon test in the function *rank_genes_group* of Scanpy to calculate gene differential expression statistics. We applied the DE tests for comparisons between NRXN1 del cells and control cells in all cell classes and time points. Normalized expression values were used as the input data. FDR adjusted p values were used to control the type I error. Genes with FDR-adjusted p values lower than 0.05 in DE tests were defined as significant DE genes. $-log_{10}(FDR\ adjusted\ p\ values)$ were defined as significance scores for the differential expression analysis. In order to highlight DEGs relevant to our analysis ('filtered' list), we extracted and integrated a list of gene sets from Gene Ontology, including neurogenesis (GO:0022008), generation of neurons (GO:0048699), neuron differentiation (GO:0030182), neuron projection development (GO:0031175), neuron development (GO:0048666), neuron projection morphogenesis (GO:0048812), neuron projection (GO:0043005), somatodendritic compartment (GO:0036477), neuronal cell body (GO:0043025), myelin sheath (GO:0043209), axonal growth cone (GO:0044295). Additionally, we excluded genes associated with translational initiation (GO:0006413), ATP metabolic process (GO:0046034), and mitochondrion organization (GO:0007005). Both 'filtered' and 'unfiltered' DEG lists are shown in the Supplementary Tables. Volcano plots were generated for the visualization of DE genes using the *EnhancedVolcano* package[96]. In Figures 3 and 4, we conducted a hypergeometric test for each comparison of two gene lists to infer the significance of the number of overlapping genes in those two lists. p value were corrected using FDR-adjusted p values.

Gene modules from ToppCell

We used ToppCell toolkit to generate gene modules of cell types and clusters in our single cell data (Figure S6B)[97]. We applied ToppCell to user-provided cell annotations and derived well-organized gene modules for all cell classes. Each gene module contains the top 200 DEGs from ToppCell, representing the most prominent transcriptomic profile of this cell class. ToppCell-derived gene modules were seamlessly enriched using ToppGene[44] and ToppCluster[98].

Gene enrichment analysis

Gene set enrichment analysis (GSEA) was conducted using ToppGene for gene sets from either ToppCell output or differential expression analysis. Gene ontologies were used to annotate

molecular functions, biological processes and cellular components. In addition, we used the *prerank* function in GSEAPY package for the customized GSEA analysis. We used the manually curated neurological-disorder-associated gene sets[65] as the reference, such as genes of autism spectrum disorder and schizophrenia. We calculated FDR adjusted p values of enrichment for differentially expressed genes to infer their associations with neurological diseases.

Trajectory inference and pseudotime analysis

We used Monocle3[42] to infer the pseudotime and trajectories of cell differentiations in the brain organoid single-cell data. We took advantage of the Seurat integration procedure and transferred Seurat objects into Monocle3 *cell_data_set* objects. Then we learned trajectories on the UMAP using the *learn_graph* function to get the pseudotime ordering of cells using the *order_cells* function. Cells with the highest expression levels of cell cycle genes in cycling NECs were selected as the start point of trajectories. In the end, every cell was assigned a pseudotime value, representing the estimated differentiation stages along the trajectory. Ridge plots were drawn based on the density of cells across pseudotime values.

Cell abundance changes inferred by scCODA

We investigated cell abundance changes in NRXN1 del samples using the scCODA model , which was used to perform compositional data analysis and determine abundance changes of cell populations using Bayesian modeling in the single cell data[45]. For cells in each cell type and each time point, compositional models were constructed using 'genotype' as the covariate in the formula. Cell types with around equal cell abundance in control and NRXN1 del samples were selected as the reference cell types. Hamiltonian Monte Carlo sampling was then initiated by calling *model.sample_hmc*. Since the lack of number of samples, we used 0.4 as the threshold of false discovery rate for significant cell abundance changes. The signs of final parameters were used to show the direction of cell abundance changes.

NRXN1 expression analysis

We collected human fetal cortex single-cell data from a large-scale single-cell dataset[40]. The ratio of NRXN1-expressing cells of each developmental stage and each cell type was defined as the number of cells with NRXN1 UMI counts greater than 1, divided by the total number of cells.

Reference Datasets

Several datasets were used for the cell type prediction in this study, including:

Kanton et al. (2019)[41]: This is a single cell dataset of human cerebral organoids derived from iPSC- and embryonic stem cell (ESC)- derived cells (43,498 cells) at different time points (day 0 ~ day 120) during the differentiation.

Paulsen et al. (2022)[37]: This is a single cell dataset of human cerebral cortex organoids with haploinsufficiency in three autism spectrum disorder (ASD) risk genes in multiple cell lines from different donors of more than 745,000 cells.

Tanaka et al. (2020)[38]: This is synthetic analysis of single cell data from multiple brain organoid and fetal brain datasets. Data of 8 different protocols were collected and 190,022 cells were selected for the reannotation, where they classified 24 distinct clusters and 13 cell types.

Velasco et al. (2019)[24]: This is a study to validate the reproducibility of brain organoids with single cell sequencing. They collected 166,242 cells from 21 individual organoids and identified indistinguishable compendiums of cell types and similar developmental trajectories.
Zhong et al. (2018)[39]: This is a single cell dataset with more than 2,300 cells in developing human prefrontal cortex from gestational weeks 8 to 26.
Bhaduri et al. (2021)[40]: This is a large-scale single cell data of developing human brain from gestation week (GW) 14 to GW 25. Multiple brain regions and neocortical areas were sampled for the data.

Data availability

Single cell-RNAseq data will be deposited and available on the NCBI Gene Expression Omnibus.

References

1. Dabell, M. P. *et al.* Investigation of NRXN1 deletions: clinical and molecular characterization. *American Journal of Medical Genetics. Part A* **161A**, 717–731 (2013).
2. Béna, F. *et al.* Molecular and clinical characterization of 25 individuals with exonic deletions of *NRXN1* and comprehensive review of the literature. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* **162**, 388–403 (2013).
3. Castronovo, P. *et al.* Phenotypic spectrum of NRXN1 mono- and bi-allelic deficiency: A systematic review. *Clin Genet* **97**, 125–137 (2020).
4. Marshall, C. R. *et al.* Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nature Genetics* **49**, 27–35 (2017).
5. Lowther, C. *et al.* Molecular characterization of NRXN1 deletions from 19,263 clinical microarray cases identifies exons important for neurodevelopmental disease expression. *Genet Med* **19**, 53–61 (2017).
6. Fuccillo, M. V. & Pak, C. Copy number variants in neurexin genes: phenotypes and mechanisms. *Current Opinion in Genetics & Development* **68**, 64–70 (2021).
7. Sudhof, T. C. Synaptic Neurexin Complexes: A Molecular Code for the Logic of Neural Circuits. *Cell* **171**, 745–769 (2017).
8. Etherton, M. R., Blaiss, C. A., Powell, C. M. & Südhof, T. C. Mouse neurexin-1alpha deletion causes correlated electrophysiological and behavioral changes consistent with cognitive impairments. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 17998–18003 (2009).
9. Dai, J., Aoto, J. & Südhof, T. C. Alternative Splicing of Presynaptic Neurexins Differentially Controls Postsynaptic NMDA and AMPA Receptor Responses. *Neuron* **102**, 993-1008.e5 (2019).
10. Davatolhagh, M. F. & Fuccillo, M. V. Neurexin1α differentially regulates synaptic efficacy within striatal circuits. *Cell Reports* **34**, 108773 (2021).
11. Trotter, J. H. *et al.* Synaptic neurexin-1 assembles into dynamically regulated active zone nanoclusters. *J Cell Biol* **218**, 2677–2698 (2019).
12. Luo, F., Sclip, A., Jiang, M. & Südhof, T. C. Neurexins cluster Ca2+ channels within the presynaptic active zone. *EMBO J* **39**, e103208 (2020).
13. Chen, L. Y., Jiang, M., Zhang, B., Gokce, O. & Südhof, T. C. Conditional Deletion of All Neurexins Defines Diversity of Essential Synaptic Organizer Functions for Neurexins. *Neuron* **94**, 611-625.e4 (2017).
14. Ullrich, B., Ushkaryov, Y. A. & Südhof, T. C. Cartography of neurexins: more than 1000 isoforms generated by alternative splicing and expressed in distinct subsets of neurons. *Neuron* **14**, 497–507 (1995).
15. Tabuchi, K. & Südhof, T. C. Structure and evolution of neurexin genes: insight into the mechanism of alternative splicing. *Genomics* **79**, 849–859 (2002).
16. Treutlein, B., Gokce, O., Quake, S. R. & Südhof, T. C. Cartography of neurexin alternative splicing mapped by single-molecule long-read mRNA sequencing. *Proceedings of the National Academy of Sciences of the United States of America* **111**, E1291-1299 (2014).
17. Schreiner, D. *et al.* Targeted combinatorial alternative splicing generates brain region-specific repertoires of neurexins. *Neuron* **84**, 386–398 (2014).
18. Jenkins, A. K. *et al.* Neurexin 1 (NRXN1) splice isoform expression during human neocortical development and aging. *Molecular Psychiatry* **21**, 701–706 (2016).
19. Zeng, L. *et al.* Functional impacts of NRXN1 knockdown on neurodevelopment in stem cell models. *PloS One* **8**, e59685 (2013).
20. Lam, M. *et al.* Single cell analysis of autism patient with bi-allelic NRXN1-alpha deletion reveals skewed fate choice in neural progenitors and impaired neuronal functionality. *Exp Cell Res* **383**, 111469 (2019).

21. Willsey, H. R. *et al.* Parallel in vivo analysis of large-effect autism genes implicates cortical neurogenesis and estrogen in risk and resilience. *Neuron* **109**, 788-804.e8 (2021).
22. Lancaster, M. A. *et al.* Cerebral organoids model human brain development and microcephaly. *Nature* **501**, 373–379 (2013).
23. Paşca, A. M. *et al.* Functional cortical neurons and astrocytes from human pluripotent stem cells in 3D culture. *Nature Methods* **12**, 671–678 (2015).
24. Velasco, S. *et al.* Individual brain organoids reproducibly form cell diversity of the human cerebral cortex. *Nature* **570**, 523–527 (2019).
25. Qian, X. *et al.* Brain-region-specific organoids using mini-bioreactors for modeling zikv exposure. *Cell* **165**, 1238–1254 (2016).
26. Bershteyn, M. *et al.* Human iPSC-Derived Cerebral Organoids Model Cellular Features of Lissencephaly and Reveal Prolonged Mitosis of Outer Radial Glia. *Cell Stem Cell* **20**, 435-449.e4 (2017).
27. Xiang, Y. *et al.* Fusion of Regionally Specified hPSC-Derived Organoids Models Human Brain Development and Interneuron Migration. *Cell Stem Cell* **21**, 383-398.e7 (2017).
28. Sebastian, R., Song, Y. & Pak, C. Probing the molecular and cellular pathological mechanisms of schizophrenia using human induced pluripotent stem cell models. *Schizophr Res* S0920-9964(22)00263–8 (2022) doi:10.1016/j.schres.2022.06.028.
29. Gordon, A. *et al.* Long-term maturation of human cortical organoids matches key early postnatal transitions. *Nat Neurosci* **24**, 331–342 (2021).
30. Kathuria, A. *et al.* Transcriptomic Landscape and Functional Characterization of Induced Pluripotent Stem Cell-Derived Cerebral Organoids in Schizophrenia. *JAMA Psychiatry* **77**, 745–754 (2020).
31. Khan, T. A. *et al.* Neuronal defects in a human cellular model of 22q11.2 deletion syndrome. *Nat Med* **26**, 1888–1898 (2020).
32. Stachowiak, E. K. *et al.* Cerebral organoids reveal early cortical maldevelopment in schizophrenia-computational anatomy and genomics, role of FGFR1. *Transl Psychiatry* **7**, 6 (2017).
33. Notaras, M. *et al.* Schizophrenia is defined by cell-specific neuropathology and multiple neurodevelopmental mechanisms in patient-derived cerebral organoids. *Mol Psychiatry* **27**, 1416–1434 (2022).
34. Pak, C. *et al.* Human Neuropsychiatric Disease Modeling using Conditional Deletion Reveals Synaptic Transmission Defects Caused by Heterozygous Mutations in NRXN1. *Cell Stem Cell* **17**, 316–328 (2015).
35. Pak, C. *et al.* Cross-platform validation of neurotransmitter release impairments in schizophrenia patient-derived NRXN1-mutant neurons. *Proceedings of the National Academy of Sciences of the United States of America* **118**, e2025598118 (2021).
36. Yoon, S.-J. *et al.* Reliability of human cortical organoid generation. *Nature Methods* **16**, 75–78 (2019).
37. Paulsen, B. *et al.* Autism genes converge on asynchronous development of shared neuron classes. *Nature* **602**, 268–273 (2022).
38. Tanaka, Y., Cakir, B., Xiang, Y., Sullivan, G. J. & Park, I.-H. Synthetic Analyses of Single-Cell Transcriptomes from Multiple Brain Organoids and Fetal Brain. *Cell Rep* **30**, 1682-1689.e3 (2020).
39. Zhong, S. *et al.* A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. *Nature* **555**, 524–528 (2018).
40. Bhaduri, A. *et al.* An atlas of cortical arealization identifies dynamic molecular signatures. *Nature* **598**, 200–204 (2021).
41. Kanton, S. *et al.* Organoid single-cell genomic atlas uncovers human-specific features of brain development. *Nature* **574**, 418–422 (2019).

42. Cao, J. *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).
43. Uzquiano, A. *et al.* Single-cell multiomics atlas of organoid development uncovers longitudinal molecular programs of cellular diversification of the human cerebral cortex. 2022.03.17.484798 Preprint at https://doi.org/10.1101/2022.03.17.484798 (2022).
44. Chen, J., Bardes, E. E., Aronow, B. J. & Jegga, A. G. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Research* **37**, W305-311 (2009).
45. Büttner, M., Ostner, J., Müller, C. L., Theis, F. J. & Schubert, B. scCODA is a Bayesian model for compositional single-cell data analysis. *Nat Commun* **12**, 6876 (2021).
46. Luza, S. *et al.* The ubiquitin proteasome system and schizophrenia. *The Lancet Psychiatry* **7**, 528–537 (2020).
47. Nishimura, A. *et al.* The carbohydrate deposits detected by histochemical methods in the molecular layer of the dentate gyrus in the hippocampal formation of patients with schizophrenia, Down's syndrome and dementia, and aged person. *Glycoconj J* **17**, 815–822 (2000).
48. Altar, C. A. *et al.* Deficient hippocampal neuron expression of proteasome, ubiquitin, and mitochondrial genes in multiple schizophrenia cohorts. *Biol Psychiatry* **58**, 85–96 (2005).
49. Rubio, M. D., Wood, K., Haroutunian, V. & Meador-Woodruff, J. H. Dysfunction of the ubiquitin proteasome and ubiquitin-like systems in schizophrenia. *Neuropsychopharmacology* **38**, 1910–1920 (2013).
50. Bousman, C. A. *et al.* Elevated ubiquitinated proteins in brain and blood of individuals with schizophrenia. *Sci Rep* **9**, 2307 (2019).
51. Singh, T. *et al.* Rare coding variants in ten genes confer substantial risk for schizophrenia. *Nature* **604**, 509–516 (2022).
52. Zhang, X. *et al.* Cell-Type-Specific Alternative Splicing Governs Cell Fate in the Developing Cerebral Cortex. *Cell* **166**, 1147-1162.e15 (2016).
53. Johnson, M. B. *et al.* Functional and evolutionary insights into human brain development through global transcriptome analysis. *Neuron* **62**, 494–509 (2009).
54. Gandal, M. J. *et al.* Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* **362**, eaat8127 (2018).
55. Irimia, M. *et al.* A highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell* **159**, 1511–1523 (2014).
56. Watanabe, A. *et al.* Fabp7 maps to a quantitative trait locus for a schizophrenia endophenotype. *PLoS Biol* **5**, e297 (2007).
57. Maekawa, M. *et al.* Polymorphism screening of brain-expressed FABP7, 5 and 3 genes and association studies in autism and schizophrenia in Japanese subjects. *J Hum Genet* **55**, 127–130 (2010).
58. Koga, M. *et al.* Plasma fatty acid-binding protein 7 concentration correlates with depression/anxiety, cognition, and positive symptom in patients with schizophrenia. *J Psychiatr Res* **144**, 304–311 (2021).
59. Hu, C., Chen, W., Myers, S. J., Yuan, H. & Traynelis, S. F. Human GRIN2B variants in neurodevelopmental disorders. *Journal of Pharmacological Sciences* **132**, 115–121 (2016).
60. Takasaki, Y. *et al.* Mutation screening of GRIN2B in schizophrenia and autism spectrum disorder in a Japanese population. *Sci Rep* **6**, 33311 (2016).
61. Coyle, J. T. NMDA receptor and schizophrenia: a brief history. *Schizophr Bull* **38**, 920–926 (2012).
62. Nakazawa, K. & Sapkota, K. The origin of NMDA receptor hypofunction in schizophrenia. *Pharmacol Ther* **205**, 107426 (2020).
63. Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502–508 (2022).

64. Shimamoto, C. *et al.* Functional characterization of FABP3, 5 and 7 gene variants identified in schizophrenia and autism spectrum disorder and mouse behavioral studies. *Hum Mol Genet* **23**, 6495–6511 (2014).
65. Wang, L. *et al.* Analyses of the Autism-associated Neuroligin-3 R451C Mutation in Human Neurons Reveals a Gain-of-Function Synaptic Mechanism. 2021.12.07.471501 Preprint at https://doi.org/10.1101/2021.12.07.471501 (2021).
66. Tarabeux, J. *et al.* Rare mutations in N-methyl-D-aspartate glutamate receptors in autism spectrum disorders and schizophrenia. *Transl Psychiatry* **1**, e55 (2011).
67. Yu, Y. *et al.* Rare loss of function mutations in N-methyl-D-aspartate glutamate receptors and their contributions to schizophrenia susceptibility. *Transl Psychiatry* **8**, 1–9 (2018).
68. Yeo, G., Holste, D., Kreiman, G. & Burge, C. B. Variation in alternative splicing across human tissues. *Genome Biology* **5**, R74 (2004).
69. Raj, B. & Blencowe, B. J. Alternative Splicing in the Mammalian Nervous System: Recent Insights into Mechanisms and Functional Roles. *Neuron* **87**, 14–27 (2015).
70. Ule, J. *et al.* Nova regulates brain-specific splicing to shape the synapse. *Nat Genet* **37**, 844–852 (2005).
71. Spellman, R. *et al.* Regulation of alternative splicing by PTB and associated factors. *Biochem Soc Trans* **33**, 457–460 (2005).
72. Vuong, C. K., Black, D. L. & Zheng, S. The neurogenetics of alternative splicing. *Nat Rev Neurosci* **17**, 265–281 (2016).
73. Quesnel-Vallières, M., Irimia, M., Cordes, S. P. & Blencowe, B. J. Essential roles for the splicing regulator nSR100/SRRM4 during nervous system development. *Genes Dev* **29**, 746–759 (2015).
74. Kaalund, S. S. *et al.* Contrasting changes in DRD1 and DRD2 splice variant expression in schizophrenia and affective disorders, and associations with SNPs in postmortem brain. *Mol Psychiatry* **19**, 1258–1266 (2014).
75. Tan, W. *et al.* Molecular cloning of a brain-specific, developmentally regulated neuregulin 1 (NRG1) isoform and identification of a functional promoter variant associated with schizophrenia. *J Biol Chem* **282**, 24343–24351 (2007).
76. Law, A. J., Kleinman, J. E., Weinberger, D. R. & Weickert, C. S. Disease-associated intronic variants in the ErbB4 gene are related to altered ErbB4 splice-variant expression in the brain in schizophrenia. *Hum Mol Genet* **16**, 129–141 (2007).
77. Sartorius, L. J. *et al.* Expression of a GRM3 Splice Variant is Increased in the Dorsolateral Prefrontal Cortex of Individuals Carrying a Schizophrenia Risk SNP. *Neuropsychopharmacol* **33**, 2626–2634 (2008).
78. Flaherty, E. *et al.* Neuronal impact of patient-specific aberrant NRXN1α splicing. *Nat Genet* **51**, 1679–1690 (2019).
79. Tai, H.-C., Besche, H., Goldberg, A. L. & Schuman, E. M. Characterization of the Brain 26S Proteasome and its Interacting Proteins. *Front Mol Neurosci* **3**, 12 (2010).
80. Bingol, B. *et al.* Autophosphorylated CaMKIIα Acts as a Scaffold to Recruit Proteasomes to Dendritic Spines. *Cell* **140**, 567–578 (2010).
81. Ferreira, J. S. *et al.* GluN2B-Containing NMDA Receptors Regulate AMPA Receptor Traffic through Anchoring of the Synaptic Proteasome. *J. Neurosci.* **35**, 8462–8479 (2015).
82. Scudder, S. L. *et al.* Synaptic Strength Is Bidirectionally Controlled by Opposing Activity-Dependent Regulation of Nedd4-1 and USP8. *J. Neurosci.* **34**, 16637–16649 (2014).
83. Huo, Y. *et al.* The deubiquitinating enzyme USP46 regulates AMPA receptor ubiquitination and trafficking. *J Neurochem* **134**, 1067–1080 (2015).
84. Fu, A. K. Y. *et al.* APCCdh1 mediates EphA4-dependent downregulation of AMPA receptors in homeostatic plasticity. *Nat Neurosci* **14**, 181–189 (2011).
85. Weinberger, D. R. Future of Days Past: Neurodevelopment and Schizophrenia. *Schizophr Bull* **43**, 1164–1168 (2017).

86. Murray, R. M. & Lewis, S. W. Is schizophrenia a neurodevelopmental disorder? *Br Med J (Clin Res Ed)* **295**, 681–682 (1987).
87. Raedler, T. J., Knable, M. B. & Weinberger, D. R. Schizophrenia as a developmental disorder of the cerebral cortex. *Curr Opin Neurobiol* **8**, 157–161 (1998).
88. Sloan, S. A. *et al.* Human Astrocyte Maturation Captured in 3D Cerebral Cortical Spheroids Derived from Pluripotent Stem Cells. *Neuron* **95**, 779-790.e6 (2017).
89. Sloan, S. A., Andersen, J., Pașca, A. M., Birey, F. & Pașca, S. P. Generation and assembly of human brain region-specific three-dimensional cultures. *Nature Protocols* **13**, 2062–2085 (2018).
90. Chen, C. & Okayama, H. High-efficiency transformation of mammalian cells by plasmid DNA. *Mol Cell Biol* **7**, 2745–2752 (1987).
91. Sun, Z. & Südhof, T. C. A simple Ca2+-imaging approach to neural network analyses in cultured neurons. *Journal of Neuroscience Methods* **349**, 109041 (2021).
92. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).
93. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst* **8**, 281-291.e9 (2019).
94. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* **19**, 15 (2018).
95. Young, M. D. *et al.* Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. *Science* **361**, 594–599 (2018).
96. Blighe, K., Rana, S. & Lewis, M. EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling. R package version 1.14.0,.
97. Jin, K. *et al.* An interactive single cell web portal identifies gene and cell networks in COVID-19 host responses. *iScience* **24**, 103115 (2021).
98. Kaimal, V., Bardes, E. E., Tabar, S. C., Jegga, A. G. & Aronow, B. J. ToppCluster: a multiple gene list feature analyzer for comparative enrichment clustering and network-based dissection of biological systems. *Nucleic Acids Res* **38**, W96-102 (2010).

Figures and figure legends

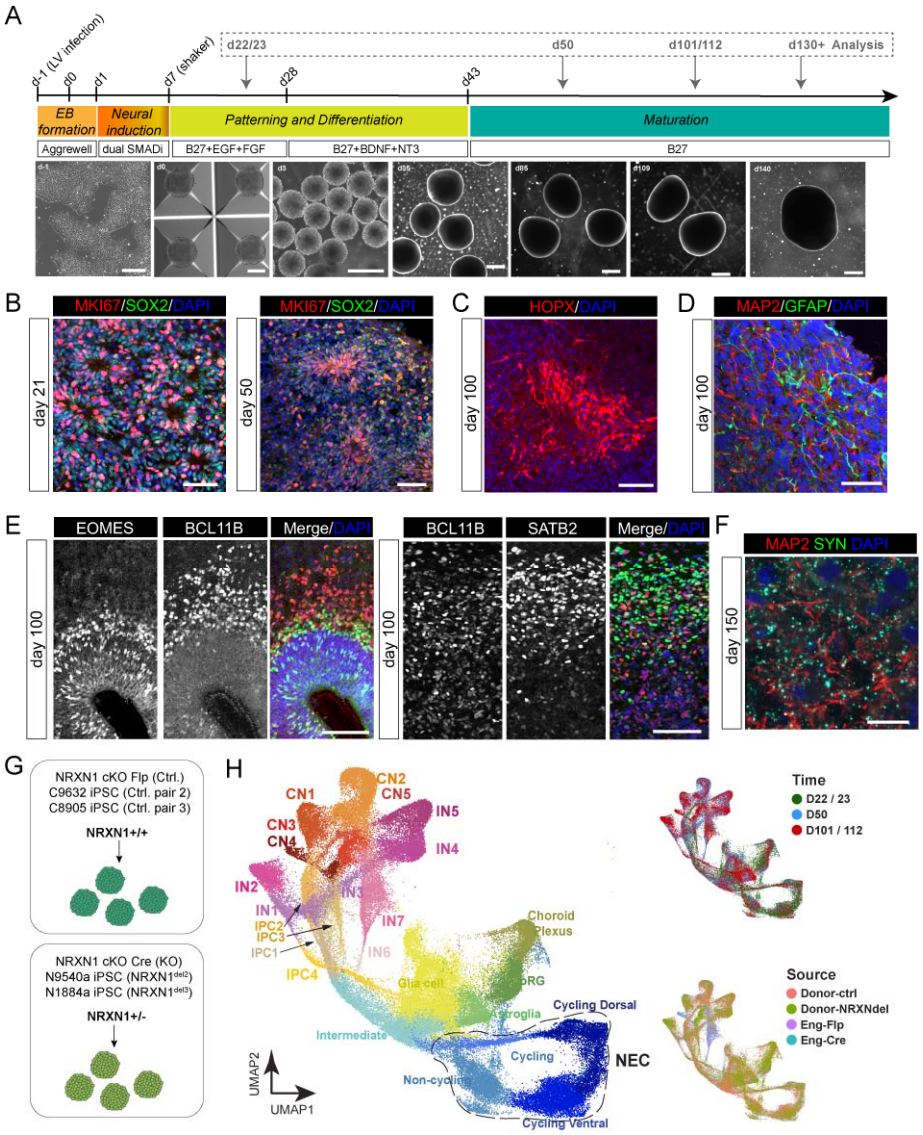

Figure 1. Generation of forebrain organoids from genetically engineered *NRXN1* cKO hESCs and donor derived iPSCs.

(A) Schematic of brain organoid generation protocol and the corresponding representative brightfield images over development. Scale bars – 250 µm for d-1, 0, 3; 100 µm for d35+. (B-F) Representative confocal images of brain organoid sections immunostained with antibodies against key markers across time points. Scale bars – 50 µm (B, D), 100 µm (C, E) 50 µm (F). (G) Schematic showing genotypes used for scRNAseq. (H) Uniform Manifold Approximation and Projection (UMAP) showing distributions of cell classes (left), time points (top right) and genotypes (bottom right) of the integrated single-cell data. Abbreviations: neural precursor cells (NECs); outer radial glial cells (oRG); intermediate precursor cells (IPC); cortical excitatory neurons (CN); and cortical GABAergic inhibitory neurons (IN).

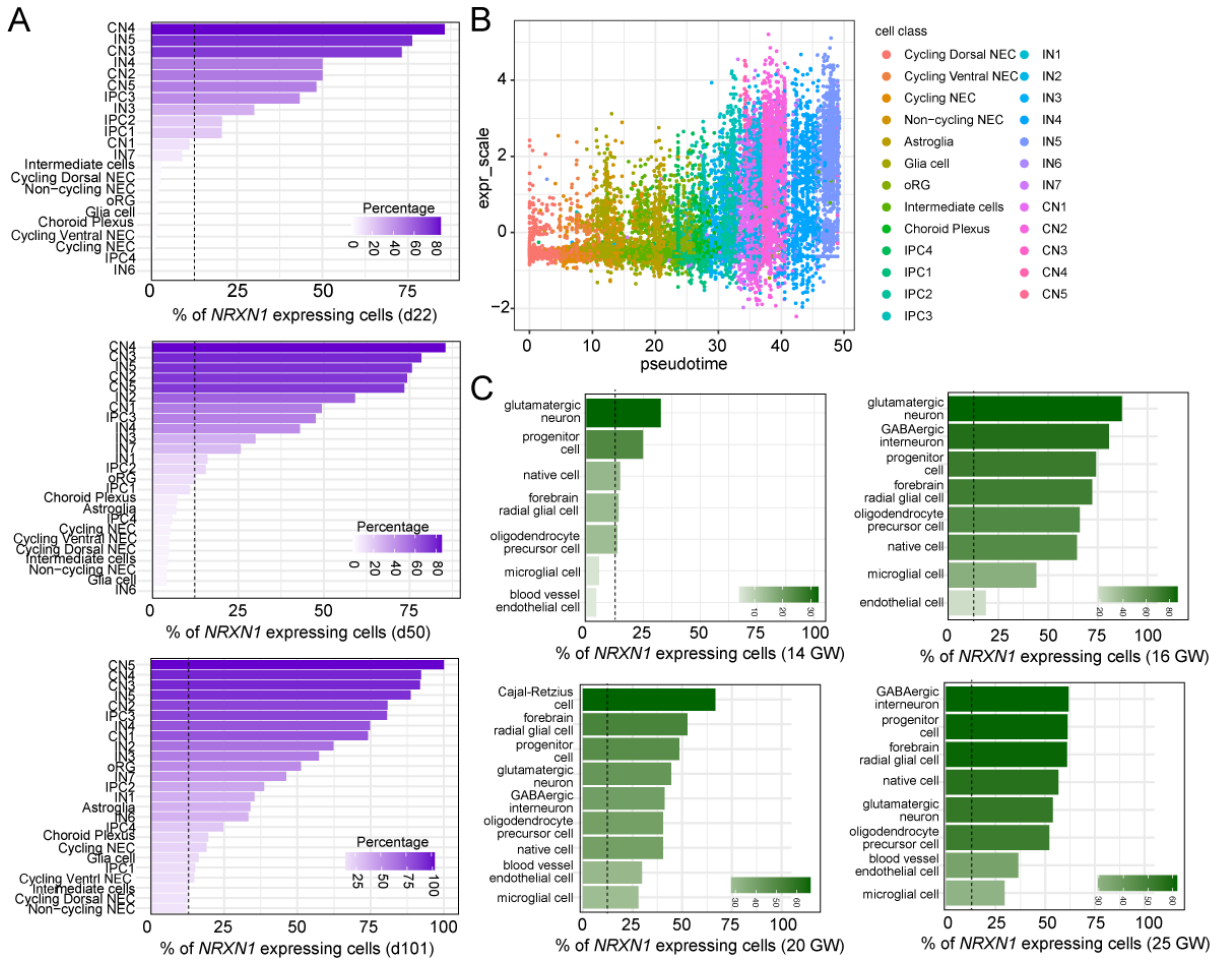

Figure 2. Single cell expression of *NRXN1* transcripts in the developing human fetal cortex and forebrain organoids.

(A) Percentage of *NRXN1*-expressing cells in each cell class from D22, D50, and D101 control donor-derived brain organoids (combined data from 2 control iPSC lines). Quantification of *NRXN1* mRNA expression for each class is shown as a percentage of *NRXN1* expressing cells. The dotted line represents ≥12.5% of cells out of the entire cell population. (B) *NRXN1* mRNA expression (Seurat scaled expression values) across pseudotime in control donor-derived brain organoids. A higher pseudotime value indicates greater maturity as indicated by the number of various neuronal cell classes (Methods). The legend for each class is shown to the right. (C) Bar graphs showing the percentage of *NRXN1* expressing cells in the human fetal tissue from 14, 16, 20, and 25 GWs. Data was mined from the human neocortex single-cell transcriptome study (14-25 GWs; Badhuri et al., 2021). The dotted line represents ≥12.5% of cells out of the entire cell population.

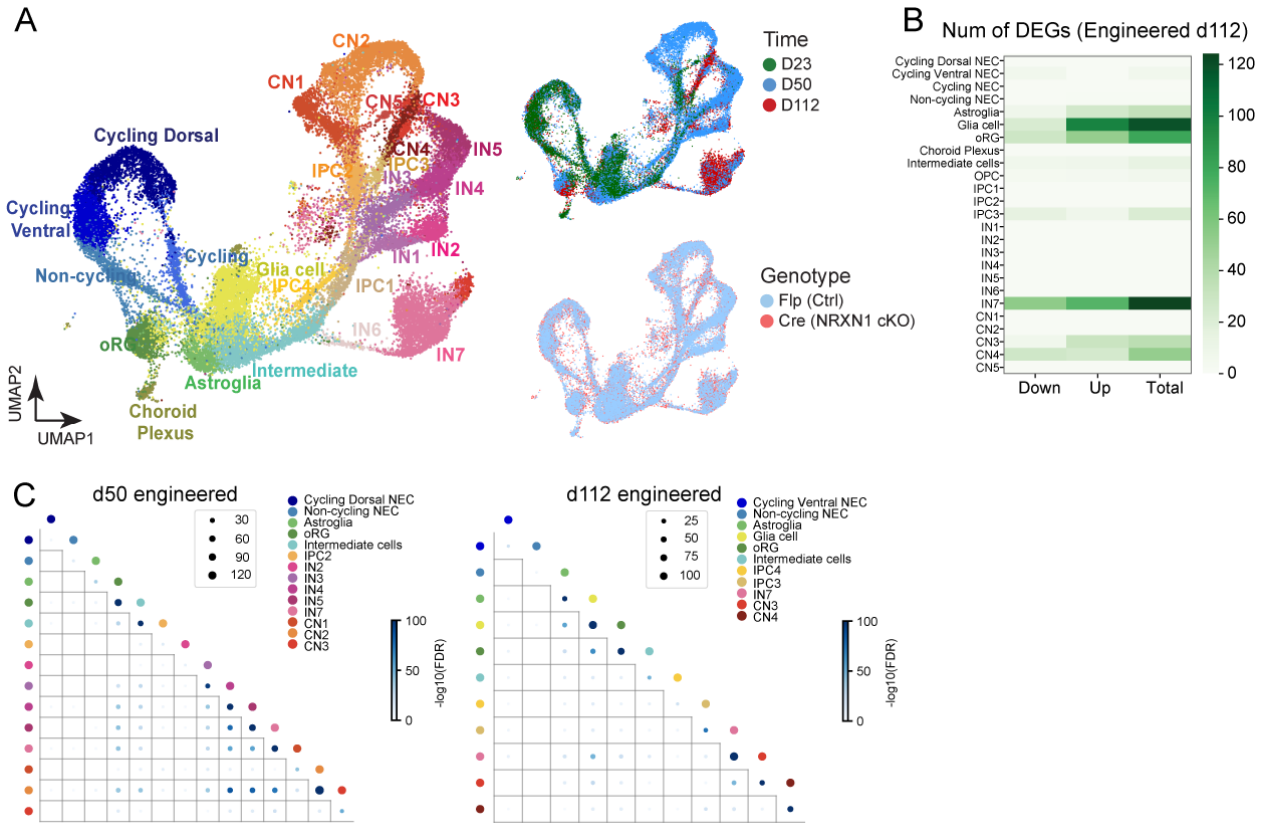

Figure 3. Perturbation effects of *NRXN1* isogenic CNVs.
(A) UMAPs showing distributions of cell classes (left), time points (top right), and genotypes (bottom right) of *NRXN1* cKO engineered brain organoids. (B) Heatmap showing the number of DEGs in each cell class of D112 engineered brain organoids. Down-regulated, up-regulated and total DEGs are shown separately. (C) The size and color of each dot in the dot plots show the number and significance of overlapping DEGs for each comparison of two cell classes in D50 (left) and D112 (right) engineered organoids. The significance was measured by $-log_{10}(FDR\ adjusted\ p\ values)$ of hypergeometric tests (see Methods).

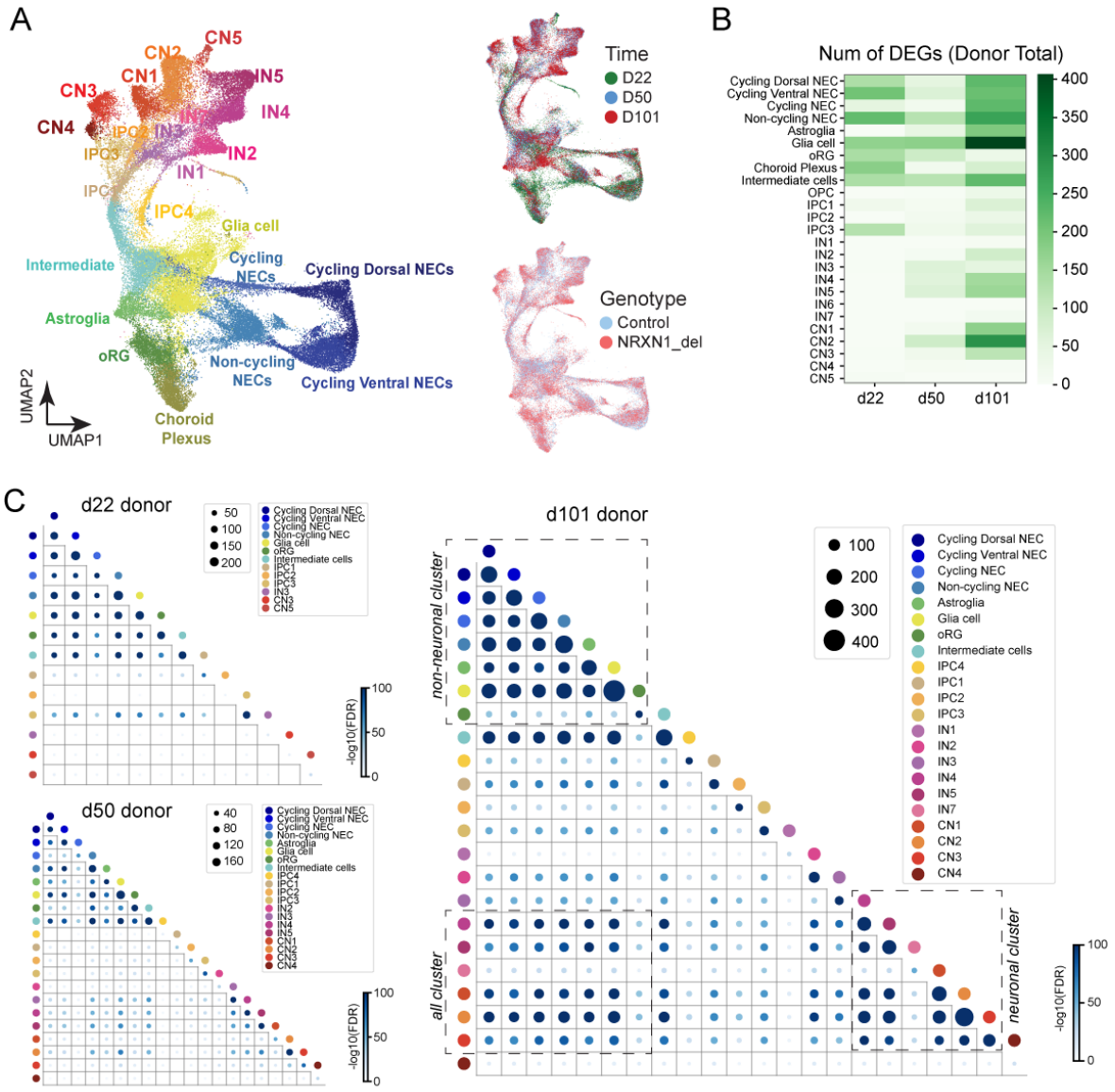

Figure 4. Perturbation effects of SCZ associated *NRXN1* CNVs.
(A) UMAPs showing distributions of cell classes (left), time points (top right), and genotypes (bottom right) of SCZ-*NRXN1*^{del} donor derived brain organoids. (B) Heatmap showing the total number of DEGs in each cell class and each time point of donor brain organoids. (C) The size and color of each dot in the dot plots show the number and significance of overlapping DEGs for each comparison of two cell classes in D22 (top left) and D50 (bottom left) and D101 (right) of donor brain organoids. Three representative gene clusters were highlighted in dotted boxes.

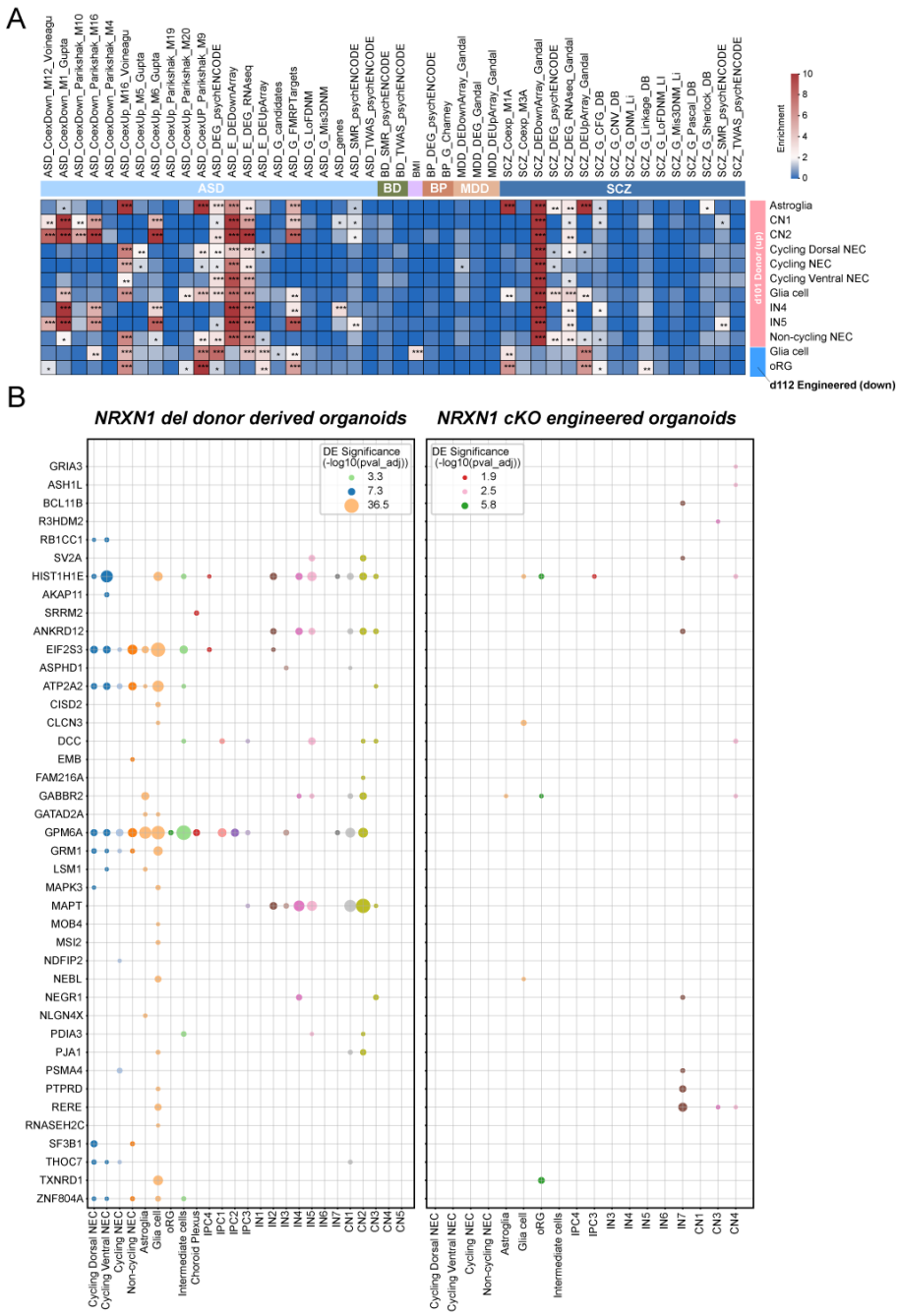

Figure 5. Differential effects of disease enrichment.
(A) Heatmap showing gene enrichment analyses of DEGs from cells of both brain organoid types (engineered and donor derived brain organoids) (rows) using neurological disorder gene sets in several categories (autism spectrum disorders, ASD; bipolar disorder, BP and BD; mood disorder, MDD; schizophrenia, SCZ) (columns). Body Mass Index (BMI) was used as control. Significance scores were defined as $-log_{10}(FDR\ adjusted\ p\ values)$ to represent the associations between DEG sets and neurological disorders. Scores were trimmed to 0~10 (see Methods). Significance levels were represented by numbers of asterisks (*: adjusted p values < 0.05; **: adjusted p values

1112 < 0.01; ***: adjusted p values < 0.001). (B) Significance of differential expression of prioritized
1113 genes obtained from PGC wave 3 and SCHMEA consortium[51,63]. The size of each dot represents
1114 the level of DE significance of each gene in each cell class of D101 donor brain organoids (left)
1115 and D112 engineered brain organoids (right).
1116

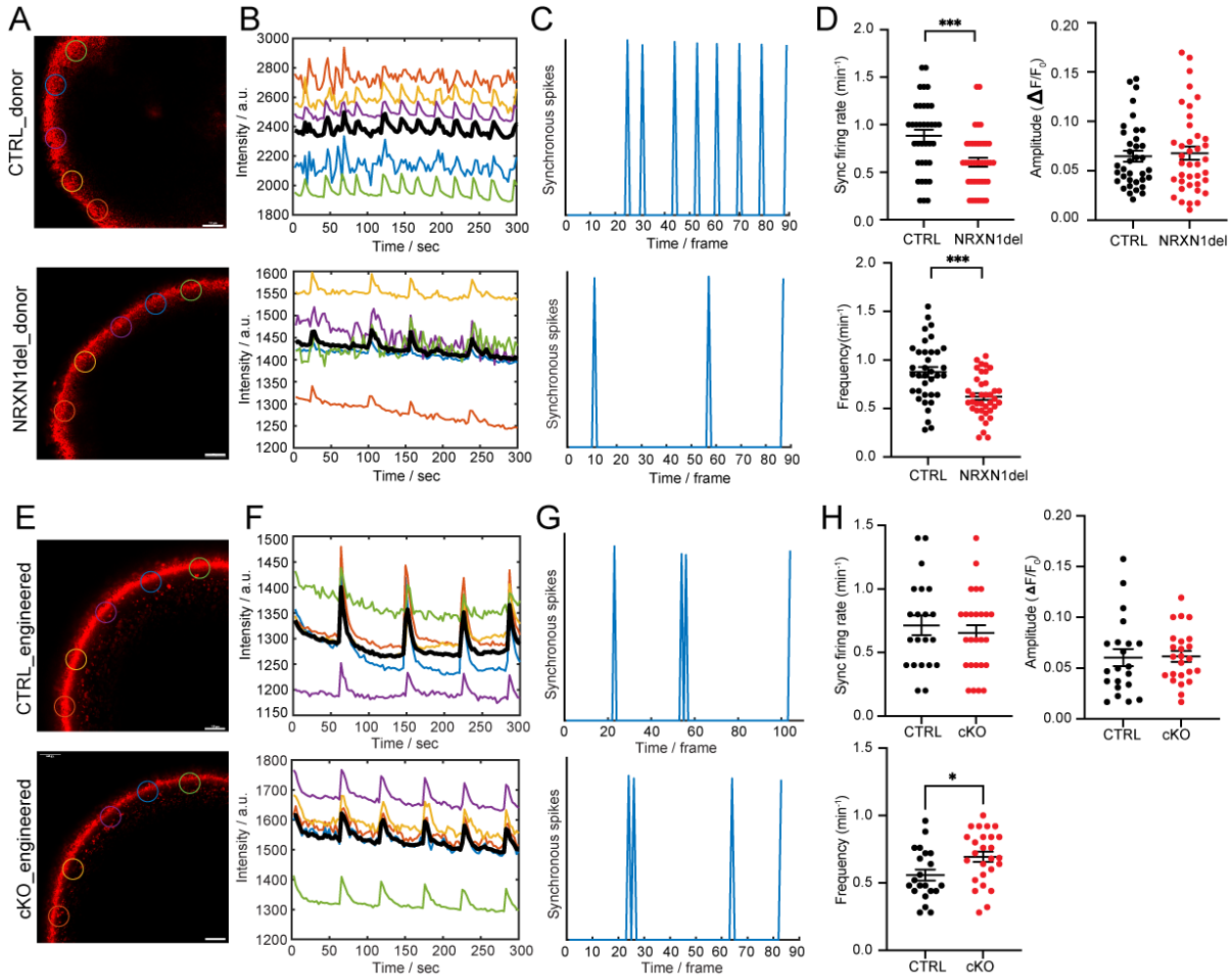

Figure 6. Impaired neuronal network activities in brain organoids carrying *NRXN1* CNVs. Intact isogenic *NRXN1* cKO and SCZ-*NRXN1*^{del} donor derived organoids (*NRXN1*^{del3} (N1884a iPSC), control pair 3 (C8905 iPSC)) at days 130-160 were used for Ca^{2+} imaging using X-Rhod-1 dye. Representative confocal images of brain organoids during live Ca^{2+} imaging (A, E). Colored circles represent regions of interest (ROI) selected for analysis. Corresponding colored raw intensity traces are shown in the boxed graphs with averaged intensities plotted in bolded black (B, F). Representative averaged synchronous spikes for each genotype are shown in C and G. Averaged data for synchronous firing rates (number of detected synchronous spikes/minute) representative of network activity, as well as amplitudes (dF/F_0) and frequencies (total number of detected peaks/minute) of spontaneous spike activity, are shown in scatter plots (D, H). Each data point represents averaged data from a single field of view (FOV) consisting of 4-5 ROIs per FOV. At least 4-6 FOVs were taken from each organoid and 4-7 organoids per genotype were used for experiments. Error bars represent S.E.M. Statistical significance is represented by asterisks: $*p < 0.05$, $***p < 0.001$.