High-risk neuropsychiatric copy number variants are associated with convergent transcriptomic changes in human brain cells

Authors: Gabi Dugan[1], Nirmala Akula[1], Stefano Marenco[2], Pavan K Auluck[2], Armin Raznahan[1], Siyuan Liu[1], Alex DeCasien[1], Qing Xu[2], Ningping Feng[2], Bhaskar Kolachana[2], Xueying Jiang[1], Michael D Gregory[3], Karen F Berman[3], Gabriel Hoffman[4], Panos Roussos[4], Francis J McMahon*[1], Anton Schulmann*[1]

Affiliations: [1] Human Genetics Branch, [2] Human Brain Collection Core (HBCC),[3] Clinical and Translational Neuroscience Branch, National Institute of Mental Health (NIMH), [4] Icahn School of Medicine, Mount Sinai

*Denotes equal contribution

Lead Contact: Gabi Dugan, gabi.dugan@nih.gov

Abstract

Large, recurrent copy number variants (CNVs) are among the strongest risk factors for neuropsychiatric conditions, contributing to multiple phenotypes with overlapping psychiatric and cognitive symptoms. However, the molecular basis of this convergent risk remains unknown. We evaluated the human brain transcriptome in carriers of nine high-risk neuropsychiatric CNVs and matched non-carriers using single nucleus RNA-sequencing. Brain tissue from carriers displayed widespread disruptions of gene expression, with deletions showing greater changes than the reciprocal duplications. Functional enrichment analysis revealed changes in mitochondrial energy metabolism and synaptic function that converged across CNVs and cell types. For mirror CNVs, the direction of effects was often reversed between deletions and duplications and showed correlation with CNV gene dosage. These findings suggest that a shared pathophysiology underlies risk for convergent brain phenotypes across CNVs and point toward promising therapeutic targets.

Keywords

Copy number variants, single-cell RNA-sequencing, post-mortem human brain, neuropsychiatric genetics

Introduction

Copy number variants (CNVs) are chromosomal alterations consisting of a deletion or duplication of specific segments on the chromosome [1,2]. This type of variant is widespread across the genome and a significant source of functional genetic variation [1,3,4]. CNVs can range in size from a few kilobases to several megabases, and their effects on gene dosage can lead to various phenotypic outcomes[5]. Some large, rare but recurrent CNVs are associated with substantial risk for various neuropsychiatric conditions, such as autism spectrum disorder, intellectual disability, schizophrenia, and mood disorders [6–10]. Despite this association with a broad range of diagnoses, carriers of neuropsychiatric CNVs exhibit symptom in the domains of cognition, perception, and mood. Overlapping diagnoses and symptoms among carriers of distinct CNVs suggest some shared pathophysiology [9,11], but the molecular basis of this clinical risk remains largely undiscovered.

Highly penetrant CNVs offer a unique opportunity to explore the impact of dosage changes in specific sets of genes within CNV regions. While many neuropsychiatric CNVs have been extensively studied with cell culture, mouse model, and neuroimaging methods [12–14], little is known about the cellular and molecular effects of CNVs within the human brain. Studies in human tissue have so far been limited to bulk tissue analysis [15,16]. Through advances in single-nucleus RNA-sequencing, it is now possible to interrogate the molecular effects of CNVs in individual cell types, revealing cell type-specific changes that may be masked in bulk tissue analyses.

In this study, we used single-nucleus RNA-sequencing of post-mortem human brain from individuals harboring high-risk CNVs to explore the neurobiological mechanisms through

which these CNVs confer risk for neuropsychiatric illnesses. We analyzed human brain tissue samples from donors with pathogenic deletions and duplications on chromosomes 22q11.2, 16p11.2, 15q11.2, or 1q21.1, and deletions on 7q11.23, along with matched non-carriers, derived from three brain banks. This unique collection encompassed some of the most penetrant genetic risk factors for neuropsychiatric illnesses. The inclusion of four "mirror" CNVs allowed us to compare reciprocal deletions and duplications. By leveraging the resolution of our dataset, we uncover cell type specific gene expression changes in the human brain. These findings shed light on biological perturbations associated with these CNVs and reveal convergent mechanisms in mitochondrial energy metabolism and synapse-related function that may underlie overlapping patterns of psychopathology.

Results

Single-Nucleus RNA-Seq Resource for High-Risk CNVs in Human Brain Cell Types

We performed single-nucleus RNA-sequencing of post-mortem human brain tissue from individuals carrying high-risk neuropsychiatric CNVs along with matched controls (Figure S1). Carriers included 13 adults (aged >10 years) and 2 infants (aged <1 year), harboring deletions or duplications on chromosomes 22q11.2, 16p11.2, 15q11.2, 1q21.1, or the deletion on the 7q11.23. CNV carriers were carefully matched with two non-carrier individuals based on brain bank, age, sex, ethnicity, and psychiatric diagnosis, resulting in a total of 28 (24 adult and four infant) non-carrier control individuals. Whenever possible, tissue from dorsolateral prefrontal cortex (dlPFC) and anterior cingulate cortex (ACC) was collected for sequencing (Figure 1A,C).

Starting with the adult samples, strict, iterative quality control was applied to remove poor quality nuclei and samples (Methods; Figure S2,S3). We performed clustering of the remaining 295,299 nuclei after filtering and resolved nine cell types for downstream analysis: CGE- and MGE-derived inhibitory neurons, upper- and lower-layer excitatory neurons, astrocytes, microglia, oligodendrocytes, oligodendrocyte precursor cell (OPCs), and vascular and leptomeningeal cells (VLMC)-endothelial cells (Figure 1C-D and Figure S4A-C).The distribution of cell types was consistent across brain regions, aligning with the high degree of concordance expected for the dlPFC and ACC in broad cell types [17](Figure S5B). The infant dataset underwent a similar pipeline but was processed separately due to differences in cell type composition (Methods; Figure S6A-D). After filtering, the dataset yielded a total of 66 adult and 10 infant samples.

For differential gene expression (DGE), we used dreamlet [18], which can incorporate mixed models and accounts for differences in sequencing depths and cell type counts across samples (Figure S5C and Table S5,S6). Our DGE model included match group, brain region, and subject as random effects and CNV as a fixed effect. "Match group" represents the grouping of carriers and matched non-carrier controls and was correlated with several other demographic variables (Methods; Figure S7A). However, for comparison, an alternative model accounting for additional quality-related variables, such as RNA integrity

number (RIN), gene detection rate, and cell type count, was highly concordant and yielded similar results (Figure S7D-E).

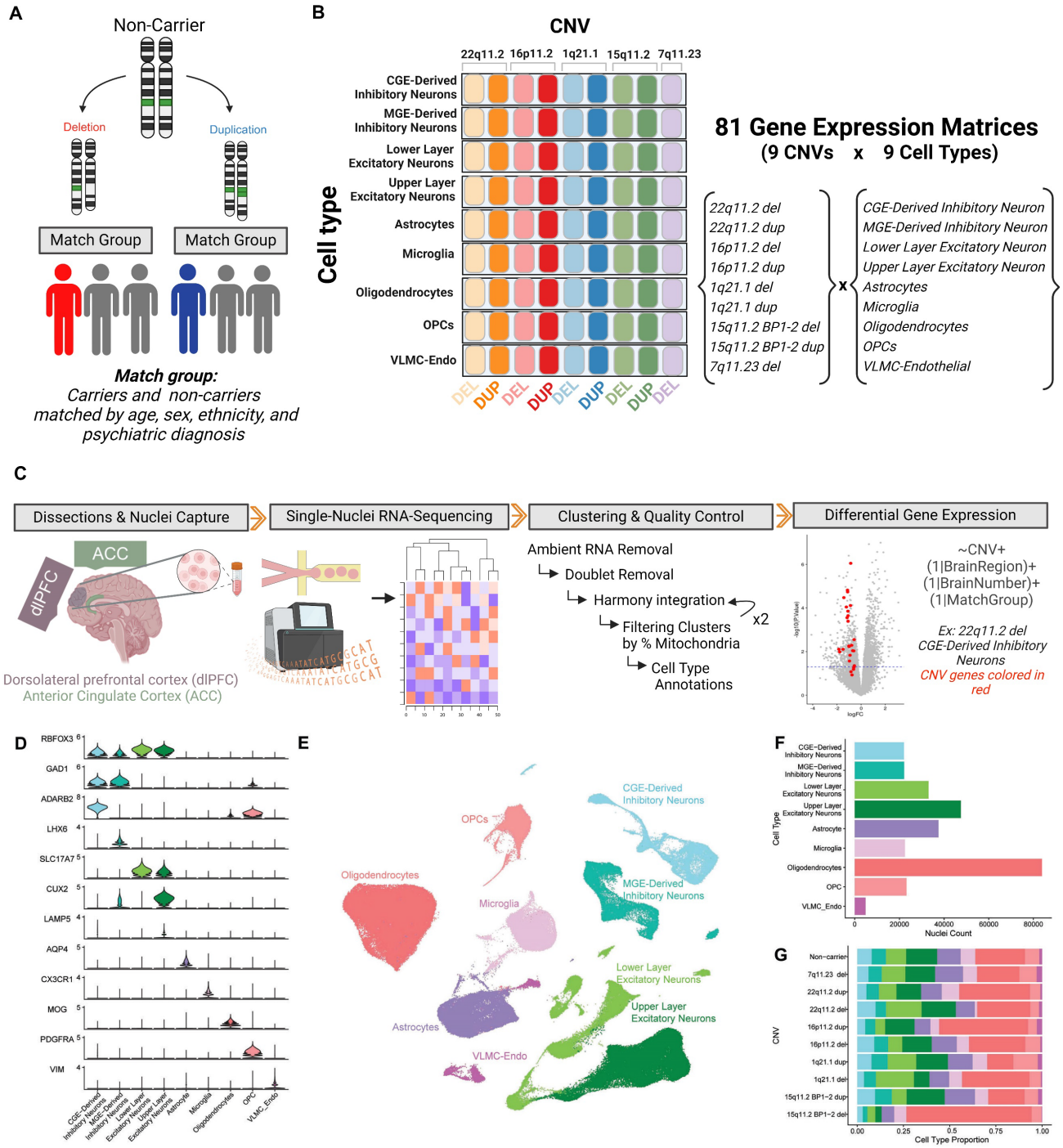

***Figure 1**: Overview of single-nucleus transcriptomic dataset to explore the effects of high-risk CNVs in human brain cells. (A) Outline of study design. Each of the 15 carriers (13 adults and 2 infants) across nine CNVs was matched with two non-carrier individuals, which comprised the match groups. (B) Schematic of dataset, 9 CNVs and 9 cell types, leading to 81 differential gene expression (DGE) tests. (C) Overview of data analysis*

pipeline. Dorsolateral prefrontal cortex (dlPFC) and anterior cingulate cortex (ACC) were dissected, nuclei were captured and underwent single-nucleus RNA-sequencing with the 10X genomics platform. Sequenced data were analyzed with iterative clustering and quality control. Resolved clusters were annotated via Azimuth[19,20] and expression of marker genes into nine well-resolved cell types. Differential gene expression analysis was conducted between carriers and non-carriers accounting for brain region, brain sample number, and match group as random effects using dreamlet [18]. (D) Stacked violin plot showing normalized expression of selected marker genes for each major cell type. (E) UMAP visualization of single nuclei colored by annotated cell types. Rigorous quality control and filtering were used to refine cluster resolution. Decisions about resolution of cell types aimed to retain resolution of biologically meaningful cell types with sufficient coverage across samples. (F) Barplot depicting total number of nuclei for each of the nine cell types across the dataset (G) Barplot depicting cell-type proportions across CNV carrier and non-carrier groups. [Created with Biorender.com].

DGE model captures expected changes in CNV regions

We first examined DGE within each CNV region. This step served as an important checkpoint to test if gene expression changes reflected CNV gene dosage, as expected. We first extracted all genes within each CNV region (Methods; Table S1,S2) and evaluated the DGE summary statistics from each model. Mean log fold-change (logFC) across cell types confirmed reduced expression of most genes within deletions and increased expression of most genes within duplications (Figure 2 and Figure S8). Thus, this model captured the anticipated direction of gene dosage change for each deletion and duplication studied.

To test if the DGE model captured the expected CNV gene dosage effects, we compared the difference in mean t-statistic between expression of genes within a CNV and those elsewhere in the transcriptome. This difference was significant for all CNVs (using a two-sided Mann-Whitney U-test, p-value <0.05) (Figure S9). This finding further affirms reliability of DGE model results.

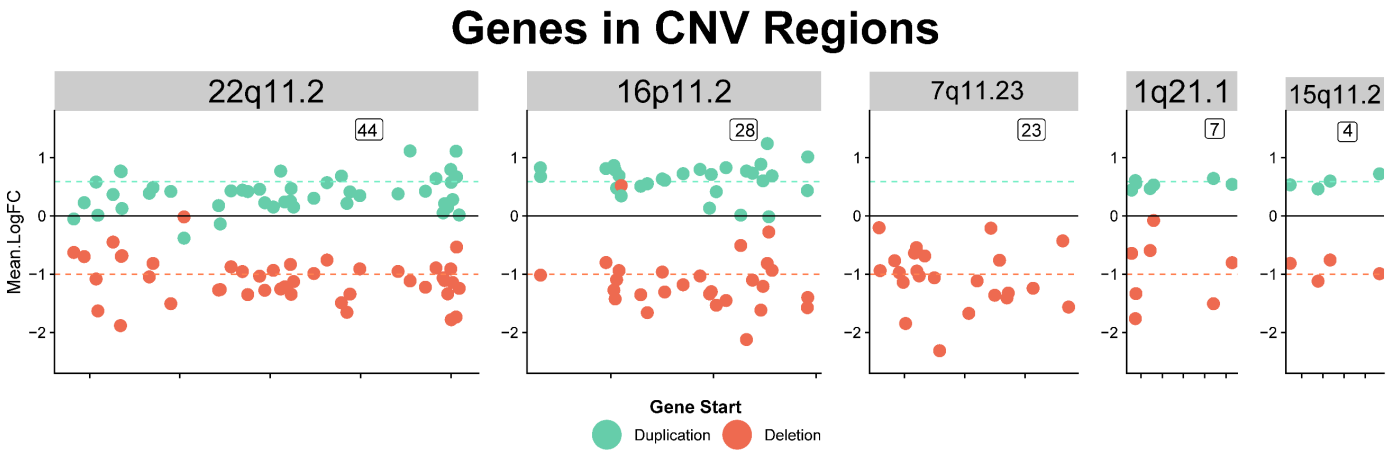

Figure 2: *Differential gene expression results within CNV regions. CNV regions were taken from ClinGen[21]. Genomic coordinates were taken from GRCH38, respective gene starts are shown on the x-axis. The mean logFC (y-axis) was calculated for each CNV gene across nine cell types from the respective DGE results. Genes are colored to indicate those within deletions (orange-red) and duplications (green), showing that with few exceptions genes from deletion model are downregulated (negative logFC) and genes from duplication models are upregulated (positive logFC). Dotted lines indicate expected logFC based on copy number.*

Previous work has also suggested that large CNVs can affect expression of nearby genes, on the same chromosome arm as the CNV [22]. To explore this question in our data set, we examined the DGE t-statistics of genes in proximity to each CNV region: those on the same chromosome, on the same chromosome arm, within 5 megabases (Mb) of the CNV, and within 1 Mb of the CNV (Figure S9). We observed some evidence of disrupted gene expression near CNV regions. These effects were more prominent in duplications than in deletions. However, most of the DEGs lay far outside of the CNV region and surrounding areas. Thus, DGE results captured widespread effects beyond the CNV and surrounding genomic regions.

Deletions have greater impact on the transcriptome than reciprocal duplications

Having confirmed that the DGE model captured expected differences in gene copy number, we proceeded to investigate changes across the entire transcriptome. Our DGE analysis revealed thousands of differentially expressed genes (DEGs) spanning cell types and CNVs. Across these dimensions, we compared the number of DEGs, at the FDR threshold of 0.1, as a measure of transcriptomic impact for each CNV. We consistently observed higher numbers of DEGs in deletions compared to the corresponding duplications. The largest numbers of DEGs were detected in carriers of deletions on 22q11.2 (4692 genes), 7q11.23 (3876 genes), 16p11.2 (1028 genes), and 1q21.1 (9990 genes). In contrast, the lowest numbers of DEGs were detected in samples carrying duplications on: 15q11.2 (15 genes), 1q21.1 (8 genes), 22q11.2 (58 genes), and 16p11.2 (23 genes) (Figure 3A). Comparing deletions and reciprocal duplications, we found a significantly greater numbers of DEGs in deletions at FDR <0.1 across all cell types (Figure 3B). A greater impact of deletions aligns with phenotypic expectations given the typically higher penetrance of deletions compared to duplications [9,23,24] and with more extensive structural changes reported in neuroimaging studies of individuals for deletion CNVs[25].

Our single-nucleus RNA-seq approach also allowed us to delineate the impact of CNVs on distinct cell types. For each CNV, we investigated the number of DEGs detected within each of the nine cell types we studied. Notably, we observed the greatest differences across cell types in the 7q11.23 and 22q11.2 deletions. Astrocytes from the 22q11.2 deletion showed a DEG count up to 9.08 times greater than that of the other eight cell types, 18.59% of genes that passed expression filtering level were differentially expressed (Table S7). This was the greatest proportion of genes differentially expressed across all models studied, with the exception of the 1q21.1 deletion. Upper- and lower-layer excitatory

neurons from the 7q11.23 deletion showed 3.23-22.05 times more DEGs than the other seven cell types, 12.40% of genes that passed quality control were differentially expressed (Table S7). The pronounced variation in DGE across cell types in the 22q11.2 and 7q11.23 deletion carriers suggests cell type-specific disruptions in these high-risk deletions, implicating astrocytes and excitatory neurons, respectively.

Investigating the intersection of DEGs across the dataset, we found that most DEGs were unique to each cell type (Figure 3C). However, neurons exhibited a higher intersection of DEGs with other neuron types. Across all CNVs, upper- and lower-layer excitatory neurons shared the largest intersection of DEGs as well as the largest total number of DEGs. Astrocytes showed the greatest number of unique DEGs (n = 1246), but still shared some DEGs with other glial populations, such as oligodendrocytes, microglia, and OPCs (Figure 3C). The large number of cell type-specific DEGs highlights the relevance of using single-cell genomics to elucidate the complex impact of these high-risk variants.

Disruption of the transcriptome correlates with deletion size

In order to quantify the relationship between the genes within each CNV region and the disruption throughout the rest of the transcriptome, we scored CNVs for gene content and loss-of-function intolerance. When exploring deletions and duplications separately, deletions showed a correlation between the number of genes in the CNV region and the number of DEGs (Figure S10A). Duplication size did not correlate to number of DEGs (Figure S10B). We excluded the 1q21.1 CNV from this analysis due to variable involvement of the proximal TAR region, leading to an inflated number of DEGs in the 1q21.1 deletion carriers we studied.

To refine this test, we utilized a metric known as the loss-of-function observed over expected upper bound fraction (LOEUF), which is a continuous measure of genetic tolerance to loss-of-function mutations [26]. A higher inverse-LOEUF score indicates intolerance to loss of function and suggests that the gene is essential for normal function and survival [27,28]. For each of our DGE models, we took the sum of the inverse-LOEUF scores of all genes retained from the CNV region after expression filtering, thereby ensuring that the genes included in the score were well expressed in that cell type. We then correlated this sum with the number of DEGs, separately for deletions and duplications. Among deletions (again excluding 1q21.1), we observed a positive correlation between the number of DEGs and the summed inverse LOEUF scores (Figure S10C). In contrast, no correlation was detected among duplications. Thus deletions with greater aggregated constraint metrics are associated with larger transcriptomic disruptions.

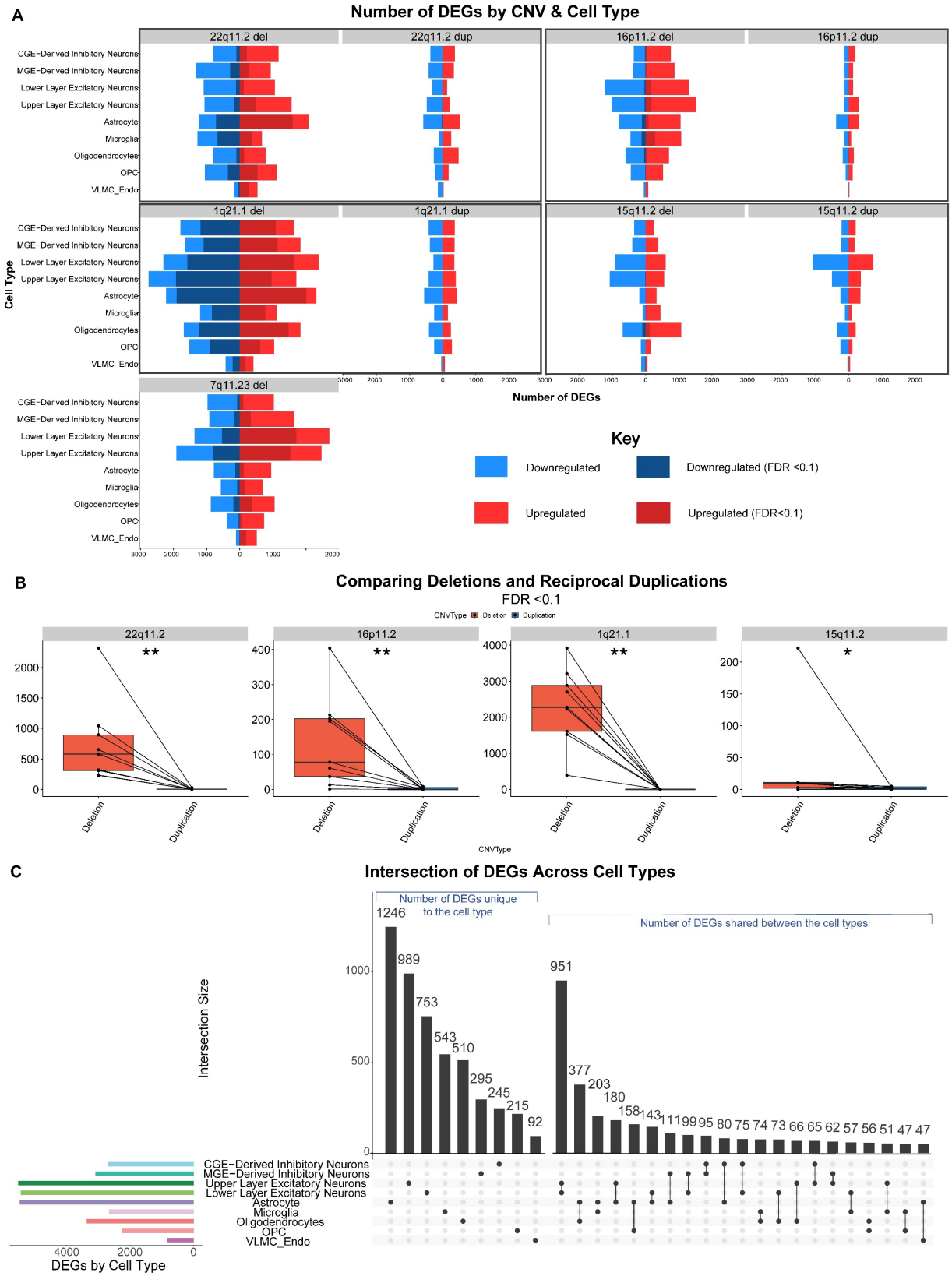

***Figure 3**: Number of differentially expressed genes (DEGs) across CNVs and cell types. (A) Number of DEGs for each CNV in all nine cell types, at p-value<0.05 (nominal) and FDR <0.1.*

Upregulated DEGs are shown as positive bars and downregulated DEGs as negative bars, with color indicating directionality and significance level. (B) Number of DEGs for each deletion and duplication across the 9 cell type models. Paired Mann-Whitney U-test on genes with FDR <0.1 shows significant differences between deletions and reciprocal duplications. (C) Intersection of DEGs across cell types shown by Upset plot. Bars show the number of unique DEGs for each cell type first and the magnitude of the intersection of DEGs between cell type pairs second, ordered by intersection size. The top 30 intersections are shown. Bar plot on bottom left shows the union of all DEGs by cell type.

Functional enrichments reveal convergence on mitochondrial energetics and synaptic terms across CNVs

A key objective of our study was to assess the functional impact of changes in gene dosage across CNVs and cell types and to explore functional convergences. To accomplish this task, we performed rank-based functional enrichment testing via GSEA for each of the 81 DGE models (9 CNVs by 9 cell types). To more comprehensively interrogate convergent signals, we evaluated significant gene ontology (GO) terms in three ways: across all 81 models (Figure 4A), for each CNV across cell types (Figure S11A), and for each cell type across CNVs (Figure 5A).

To uncover major convergent functional themes, we first explored the most frequently implicated terms with FDR<0.1 from GSEA across all 81 models. The most frequently implicated terms included “mitochondrial ATP synthesis coupled electron transport”, “cellular respiration”, “oxidative phosphorylation” (OXPHOS), “regulation of synaptic plasticity”, and “synapse organization” (Figure 4A). Nine of the top ten most frequent terms were related to two major themes: mitochondrial energy metabolism and synaptic function.

Within these themes, we found that the direction of enrichment often diverged between deletions and reciprocal duplications, suggesting a CNV gene dosage effect. Among synapse-related terms, CNVs on 22q11.2 and 1q21.1 showed the most prominent directional divergence between the respective deletion and duplication: synaptic terms were upregulated in the 22q11.2 deletion and 1q21.1 duplication and downregulated in the 22q11.2 duplication and 1q21.1 deletion (Figure 4B). Mitochondrial energy metabolism terms showed the most distinct reversal of direction in the 16p11.2: the deletion showed downregulation and the duplication showed upregulation of genes enriched for these terms. The 7q11.23 deletion showed the strongest upregulation of energy metabolism-related terms across cell types of any individual CNV.

Notably, OXPHOS was the most frequently implicated term, with FDR<0.1 in 47 of the 81 models across CNVs and cell types. OXPHOS was significantly enriched in at least one cell type for all nine CNVs, and the direction of enrichment stayed mostly consistent across cell types for each CNV. The strongest OXPHOS enrichments were associated with the 7q11.23 deletion, with consistent upregulation across cell types, and the 16p11.2 deletion, with consistent downregulation across most cell types. These two deletions showed OXPHOS enrichment with an FDR<0.001 for most cell types (Figure 4C). Enrichment of genes within four of the classical complexes of the electron transfer chain (complex II was

removed due to consisting of only four genes and therefore limited power to test for enrichment), showed that OXPHOS enrichments extended across all four complexes tested. Specifically, the 7q11.23 and 16p11.2 deletions showed respective up- and downregulation for complexes I-V, excluding complex II, across all neuronal and many glial cell types (Figure S12A-B).

To ensure that these and other finding were not a consequence of confounding or quality-related factors, we compared these results to our quality-covariate adjusted model, which also included RNA integrity number (RIN), mean gene detection rate, and cell count of each cell type by sample as fixed effects (Methods; Figure S7D). We observed highly correlated transcriptome-wide t-statistics between the original and the quality-covariate-adjusted model (Figure S7E). Importantly, OXPHOS enrichments remained highly significant and consistent across CNVs and cell types (Figure S13B), demonstrating that the OXPHOS signal was not driven by quality-related variables.

To capture the most pervasive disruptions across cell types for individual CNVs, we next examined the most consistent functional enrichments for each individual CNV (Figure S11A). Themes of synaptic disruption and OXPHOS emerged for most individual CNVs. Some CNVs, specifically in the 22q11.2 duplication and 15q11.2 deletion, showed synaptic terms as the most frequently implicated terms across cell types. On the other hand, OXPHOS was one of the most frequently implicated terms across cell types for six of the nine CNVs. For example, the most frequently implicated terms across cell types for the 16p11.2 and 7q11.23 deletions were related to mitochondrial energy metabolism, consistent with the strong enrichment signal for OXPHOS observed for these deletions (Figure 4C). We also noted that a related theme of hypoxia-related terms emerged for 22q11.2 across cell types with the most frequent terms including “response to oxygen levels” and “cellular response to hypoxia” (Figure S11A). Lastly, although OXPHOS was not significantly enriched for the 22q11.2 in most cell types, another term related to energy metabolism, “glycolytic processes”, was frequently implicated across cell types (Figure S11B).

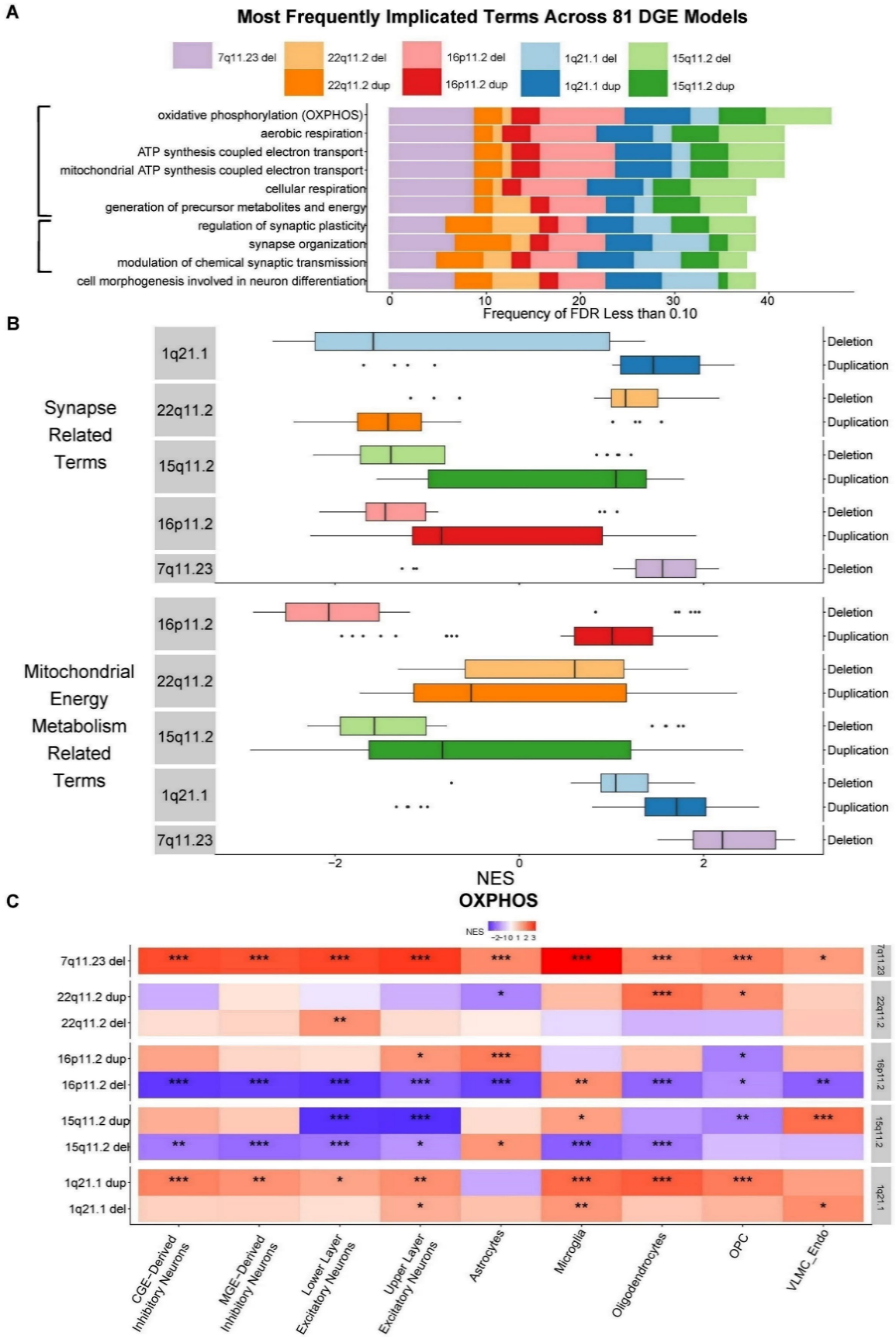
A
Most Frequently Implicated Terms Across 81 DGE Models
7q11.23 del
22q11.2 del
22q11.2 dup
16p11.2 del
16p11.2 dup
1q21.1 del
1q21.1 dup
15q11.2 del
15q11.2 dup
oxidative phosphorylation (OXPHOS)
aerobic respiration
ATP synthesis coupled electron transport
mitochondrial ATP synthesis coupled electron transport
cellular respiration
generation of precursor metabolites and energy
regulation of synaptic plasticity
synapse organization
modulation of chemical synaptic transmission
cell morphogenesis involved in neuron differentiation
0
10
20
30
40
Frequency of FDR Less than 0.10
B
Synapse Related Terms
Mitochondrial Energy Metabolism Related Terms
1q21.1
22q11.2
15q11.2
16p11.2
7q11.23
Deletion
Duplication
−2
0
2
NES
C
OXPHOS
NES
−2 −1 0 1 2 3
7q11.23 del
22q11.2 dup
22q11.2 del
16p11.2 dup
16p11.2 del
15q11.2 dup
15q11.2 del
1q21.1 dup
1q21.1 del
CGE-Derived Inhibitory Neurons
MGE-Derived Inhibitory Neurons
Lower Layer Excitatory Neurons
Upper Layer Excitatory Neurons
Astrocytes
Microglia
Oligodendrocytes
OPC
VLMC_Endo

Figure 4: *Convergence of functional enrichments across CNVs on synapse and mitochondrial energy metabolism. (A) Barplot showing the intersection of the top ten most frequently implicated terms (y-axis) across the 81 tests (nine cell types and nine CNVs). Nine of the top ten terms were categorized into two key themes: synapse related terms and mitochondrial energy metabolism related terms. One term did not fit into these categories, so it was left out of groupings. (B) Boxplot depicting normalized enrichment scores (NES) (x-axis) of synapse-related terms (aggregating three GO terms) and those related to mitochondrial energy metabolism (aggregating five GO terms). NES values are arranged on the x-axis and colored by CNV. There were observable CNV gene dosage effects with several deletions and reciprocal duplications having opposite directions of enrichments, e.g. synapse related terms are positive in 22q11.2 deletion and negative in 22q11.2 duplication. (C) Heatmap showing the OXPHOS (GO:0006119) functional enrichment results, which was the most frequent term across the nine cell types (x-axis) and nine CNVs (y-axis). The grid is colored by NES and annotated by FDR significance thresholds (FDR <0.1 = *, FDR <0.01 = **, FDR <0.001 = ***). OXPHOS is significant in all CNVs, often across cell types.*

Disruption in mitochondrial energetics extends to infant CNV carriers

We also analyzed a smaller infant dataset, which included samples from the dlPFC and ACC from infant carriers of the 7q11.23 (WBS) or 15q11.2 (BP1-BP2) deletion, along with matched infant non-carriers. A parallel approach for cell type annotations and clustering (Methods; Figure S6A-D), as well as DGE via dreamlet and functional enrichments via GSEA was conducted on the infant samples (Figure S6,S15). DGE models in infants showed lower residual variance, suggesting that infant brain may be less influenced by external factors than adult brain samples (Figure S14B). Genes with the deletions were downregulated, as expected (Figure S15A). In the 7q11.23 deletion, we observed that GO terms implicated in the adult carriers were also implicated in infants, including OXPHOS (Figure S15C-D). Genes involved in OXPHOS were highly significantly (FDR <0.001) downregulated across all four neuron types in the 7q11.23 deletion in the infant brain and upregulated across several cell types in the 15q11.2 deletion (Figure S15D). Thus, while the direction of enrichment was reversed between the infant and adult samples, gene expression changes implicated OXPHOS changes in both adult and infant carriers of the 7q11.23 and 15q11.2 deletions.

Cell type-specific disruptions shared across multiple CNVs

Having explored the most consistent terms across the entire dataset and for each individual CNV across cell types, our focus next turned to identifying common signals among cell types across distinct CNVs. This analysis contextualizes alterations specific to each cell type, which is masked in bulk-tissue studies, while also revealing convergent patterns across distinct high-risk variants. As before, we examined the most frequent terms with FDR<0.1 within each cell type.

For neurons, we observed that the top three most frequently implicated terms were distinct for each neuronal cell type, but largely related to regulation of the synapse, protein quality

control, and cellular communication. For CGE-derived inhibitory neurons, enrichments involved cellular communication and signaling, e.g. “signal release”, “negative regulation of transport”, and “cell-cell adhesion via plasma membrane adhesion molecules”. MGE-derived inhibitory neurons showed enrichments in “proteasomal protein catabolic processes” and “vesicle-mediated transport in synapse”. Both upper- and lower- layer excitatory neurons displayed functional enrichments associated with synaptic transmission modulation, synapse organization, and mitochondrial function (Figure 5A).

For glial cell types, many of the convergent terms also involved energy metabolism. The most frequently implicated enrichments for astrocytes included “actin cytoskeleton organization”, “autophagy”, and ATP metabolic processes. Microglia-specific enrichments involved cytokine responses and terms related to OXPHOS, suggesting aberrant energy metabolism and alterations in gene expression involved in inflammation and cell survival. Oligodendrocytes and OPCs exhibited enrichments in neurotransmitter transport, cell projection morphogenesis, synapse organization, and OXPHOS, which demonstrates that the pervasive disruption of synaptic and metabolic terms across CNVs extended to major glial populations (Figure 5A). Interestingly, we observed that the directions of functional enrichments again often diverged between deletions and duplications. This was most pronounced in excitatory neurons, especially in the “synapse organization” and “signal release” terms for upper- and lower-layer excitatory neurons respectively.

Given the widespread disruptions in OXPHOS and synaptic functions across CNVs, we hypothesized that changes in gene expression for OXPHOS may coincide with changes in gene expression for the synapse in neurons. Using the SynGO database to select synapse-related terms, we found that presynapse was the most convergent term across the dataset, with an FDR<0.1 in 42 of the 81 models (Figure S16A,S16C). We observed a significant, linear relationship between the enrichment scores for OXPHOS and presynapse (Figure 5B). Postsynaptic density was also among the most convergent SynGO terms, with FDR<0.1 in 37 of the models (Figure S16A). Similarly, we observed a significant, linear relationship between the enrichment scores for OXPHOS and postsynaptic density, although this relationship was less pronounced than for the presynapse (Figure 5C).

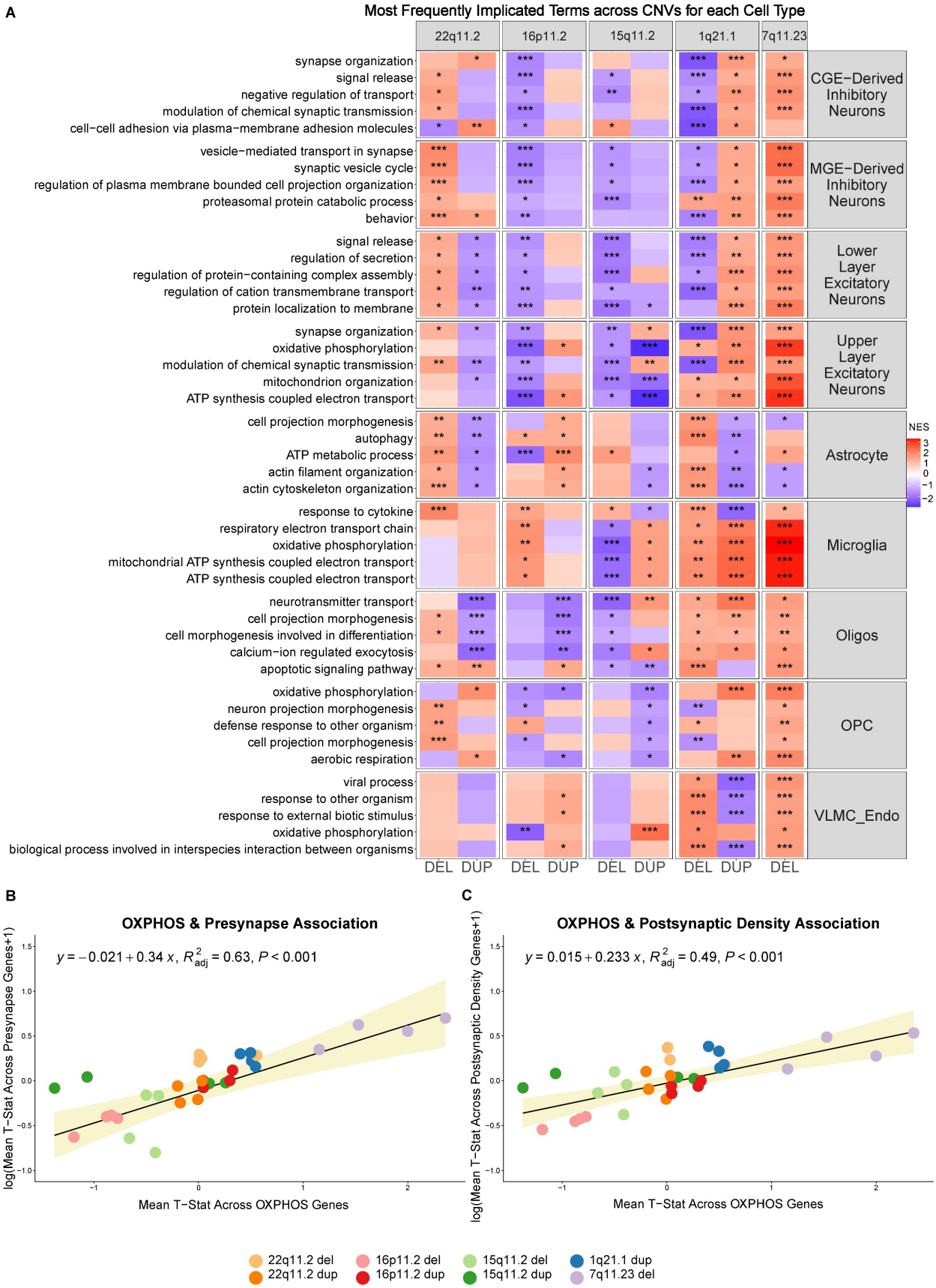
A
Most Frequently Implicated Terms across CNVs for each Cell Type
22q11.2
16p11.2
15q11.2
1q21.1
7q11.23
synapse organization
signal release
negative regulation of transport
modulation of chemical synaptic transmission
cell−cell adhesion via plasma−membrane adhesion molecules
CGE−Derived Inhibitory Neurons
vesicle−mediated transport in synapse
synaptic vesicle cycle
regulation of plasma membrane bounded cell projection organization
proteasomal protein catabolic process
behavior
MGE−Derived Inhibitory Neurons
signal release
regulation of secretion
regulation of protein−containing complex assembly
regulation of cation transmembrane transport
protein localization to membrane
Lower Layer Excitatory Neurons
synapse organization
oxidative phosphorylation
modulation of chemical synaptic transmission
mitochondrion organization
ATP synthesis coupled electron transport
Upper Layer Excitatory Neurons
cell projection morphogenesis
autophagy
ATP metabolic process
actin filament organization
actin cytoskeleton organization
Astrocyte
NES
response to cytokine
respiratory electron transport chain
oxidative phosphorylation
mitochondrial ATP synthesis coupled electron transport
ATP synthesis coupled electron transport
Microglia
neurotransmitter transport
cell projection morphogenesis
cell morphogenesis involved in differentiation
calcium−ion regulated exocytosis
apoptotic signaling pathway
Oligos
oxidative phosphorylation
neuron projection morphogenesis
defense response to other organism
cell projection morphogenesis
aerobic respiration
OPC
viral process
response to other organism
response to external biotic stimulus
oxidative phosphorylation
biological process involved in interspecies interaction between organisms
VLMC_Endo
DEL DUP DEL DUP DEL DUP DEL DUP DEL
B
OXPHOS & Presynapse Association
y = −0.021 + 0.34 x, R²adj = 0.63, P < 0.001
log(Mean T−Stat Across Presynapse Genes+1)
Mean T−Stat Across OXPHOS Genes
C
OXPHOS & Postsynaptic Density Association
y = 0.015 + 0.233 x, R²adj = 0.49, P < 0.001
log(Mean T−Stat Across Postsynaptic Density Genes+1)
Mean T−Stat Across OXPHOS Genes
22q11.2 del
22q11.2 dup
16p11.2 del
16p11.2 dup
15q11.2 del
15q11.2 dup
1q21.1 dup
7q11.23 del

***Figure 5**: Cell type-specific enrichments across CNVs. (A) GSEA results of the top 5 most frequently implicated terms enriched at FDR <0.1 across CNVs were selected for each of the nine cell types, ordered in reverse alphabetical order. Heatmap is colored by NES, with red indicating positive NES and blue indicating negative NES. FDR significance is marked with asterisks (FDR <0.1 = *, FDR <0.01 = **, FDR < 0.001 = ***). (B) Association of OXPHOS and presynapse DGE statistics. Mean t-statistics across all genes in the OXPHOS term for each neuronal CNV model (n=36) (x-axis) and across all genes in presynapse term (y-axis), which was log-transformed for normality. (C) Association of OXPHOS and postsynaptic density DGE statistics. Same x-axis as described above and mean t-statistic across genes in postsynaptic density term for CNV models for four neuronal cell types.*

Mirror CNVs show gene dosage effects in synaptic and energy metabolism terms

Our collection contains four 'mirror' CNVs, allowing us to test gene dosage effects between deletions and reciprocal duplications. Evaluating each CNV independently, we found that several significant functional enrichments were shared, but diverged in their direction for deletions and the reciprocal duplications. Previous work has demonstrated reciprocal changes in clinical phenotypes, brain morphology, and gene expression between carriers of deletions and duplications of the same genomic segments [28–31]. These "mirror" phenotypes suggest gene-dosage dependent effects for these mirror CNVs across the transcriptome.

To directly examine CNV gene dosage effects, we constructed gene dosage models for each mirror CNV, assigning a value of 1 for deletion carriers, 2 for non-carriers, and 3 for duplication carriers (Figure 6A). These modified DGE models test for genes that are positively or negatively correlated with CNV gene dosage for each CNV, followed by functional enrichments to evaluate terms most frequently associated with CNV gene dosage across cell types. The 22q11.2 CNV showed the largest number of CNV dosage-sensitive genes with FDR<0.1. (Figure 6B). In the 22q11.2 gene dosage model, similar to the individual CNV models, astrocytes showed the largest number of DEGs, with 1245 genes at FDR <0.1, which was almost 10 times greater than the cell type with the next greatest number of DEGs (CGE-derived inhibitory neurons with 146 DEGs).

Functional enrichments of DEGs detected under the gene dosage models largely recapitulated our previous results for individual CNVs. The 22q11.2 gene dosage model was enriched for terms related to hypoxia response and glycolytic processes, which were both negatively correlated with gene dosage. Additionally, "regulation of synaptic plasticity" was enriched in many cell types, all negatively correlated with copy number (Figure 6C). This aligns with our previous analysis of convergent synapse-related terms for the 22q11.2 deletion and duplication (Figure 4B). Among carriers of 16p11.2 CNV, almost all of the most frequently implicated term across cell types were related to energy metabolism: "aerobic respiration", "OXPHOS", "mitochondrial ATP synthesis", etc. (Figure 6C). This aligns with our previous analysis of convergent terms related to in mitochondrial energy metabolism in the 16p11.2 deletion and duplication (Figure 4A). Our gene dosage model thus confirmed that OXPHOS was significantly positively correlated to 16p11.2 gene

dosage for most cell types. We observed comparable levels of significance, but a negative correlation to 16p11.2 gene dosage for microglia.

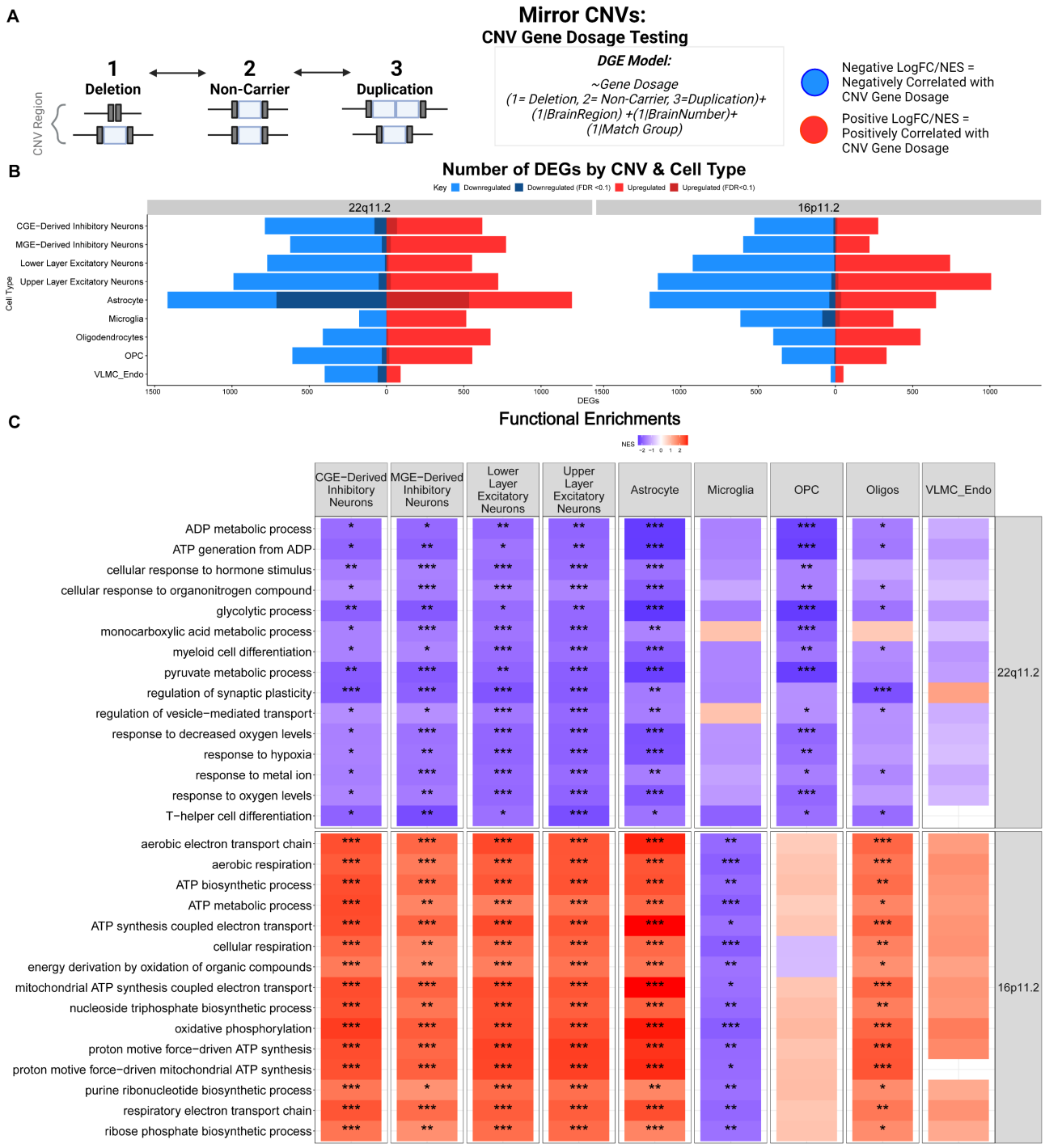

Figure 6: *Differential gene expression testing for association with CNV gene dosage in two 'mirror' CNVs. (A) Schematic showing gene dosage testing. DGE was run separately for each CNV with groups of deletions, reciprocal duplication, and all non-carriers. (B) Number of DEGs (FDR <0.1) within cell types for each gene dosage model. Following the same*

protocol described in Figure 3A. (C) Top 15 most frequently implicated enriched terms across cell types for 22q11.2 and 16p11.2 gene dosage models are shown, ordered alphabetically.

Discussion

We present a unique dataset which, for the first time, investigates the impacts of several high-risk CNVs in human brain with cell type resolution. We systematically delineate major transcriptomic effects of nine recurrent, high-penetrance neuropsychiatric CNVs across cell types, with the overarching goal of shedding light on cellular pathophysiology that underlies shared phenotypes. Leveraging single-nucleus transcriptomics, we attain cellular resolution from human brain tissue masked in previous bulk tissue studies, allowing us to interrogate impacts of each CNV and converging signals within and across cell types. Using this resource, we discovered that differentially expressed genes in high-risk neuropsychiatric CNVs converged on mitochondrial energetics and synaptic terms. We also found cell-type specific changes, with the most pronounced effects in astrocytes and neurons. These findings provide valuable insights into the shared molecular underpinnings of high-risk CNVs, with major implications for future research and therapeutic targets.

Across all DGE models, we observed a consistent upregulation for duplications and downregulation for deletions of genes within the CNV regions, affirming the reliability of the DGE models we employed. We noted some disrupted gene expression in the vicinity of CNV regions, particularly in duplications, which is consistent with previous findings that large CNVs can influence the expression of nearby genes [22]. Despite some localized disruption, the majority of DEGs were located outside the vicinity of the CNVs, indicating that high-risk CNVs have broad, widespread impacts on gene expression beyond their proximal effects. The transcriptome-wide number of DEGs detected in CNV carriers often echoed the expected level of phenotypic risk associated with those CNVs. For instance, deletions exerted a greater influence on the transcriptome than the corresponding duplications of the same genomic region. This aligns with clinical and neuroimaging studies that have shown that individuals with pathogenic deletions tend to exhibit more severe phenotypes and greater penetrance of neuropsychiatric and neurodevelopmental disorders than those with the reciprocal duplications [1,9,13,23,24]. Larger deletions, which encompass more genes and are known to be more penetrant than smaller deletions and duplications[9], exhibited the highest number of DEGs across the dataset, especially evident from the 22q11.2 and 7q11.23 deletions. We observed that the number of DEGs across the transcriptome correlated with the number of genes in the CNV region for deletions. One exception to this was the 1q21.1 deletion, the sole carrier of which showed the highest number of DEGs despite the moderate size of the distal 1q21.1 region. 1q21.1 CNV carriers in our study varied in the involvement of the proximal TAR region, with the deletion spanning both the proximal and distal region, a larger subtype of the 1q21.1 deletion often associated with more severe phenotypes[32]. Additionally, one of the three 1q21.1 distal duplications was also accompanied by a deletion in the proximal region. Due to these irregularities, the 1q21.1 CNV was not included in the CNV size and constraint analyses.

After removing the 1q21.1 CNV, the number of DEGs was significantly correlated with the number of deleted genes as well as their aggregated evolutionary constraint scores (sum of 1/LOEUF), highlighting the extensive impacts of large, rare deletions on the transcriptome of human brain cell types.

A core goal of our study was to elucidate the convergent effects of high-risk CNVs across diverse cell types in the human brain, to provide biological insight into their shared pathophysiology. Through functional enrichment analyses, we found that gene expression changes converged on functions related to energy metabolism and synaptic activity. The most frequently implicated term, both with and without correction for quality-related covariates, was OXPHOS. Gene expression changes in the OXPHOS pathway were widespread across the complexes of the electron transport chain and present across CNVs and cell types. The most pronounced effects were seen in the 16p11.2 and 7q11.23 deletions, and the direction of the changes was often correlated with CNV gene dosage, especially in the case of 16p11.2. These findings are consistent with previous studies that implicated disruptions of mitochondrial function and OXPHOS in cell culture and mouse models of the 16p11.2 deletion [33,34], 7q11.23 deletion [35,36], and 22q11.2 deletion [37,38]. Notably, disruptions in mitochondrial energetics have also been observed in CNVs not covered in our study, but with similar risk profiles, for instance in cell culture models of the 3q29.2 deletion, a significant genetic risk factor for schizophrenia and bipolar disorder[39]. Our findings, taken together with the aforementioned studies of individual CNVs, point to energy metabolism as a common thread among carriers of high-risk neuropsychiatric CNVs. This suggests that, despite potential differences in underlying mechanisms, mitochondrial dysfunction could contribute to the convergent pathophysiology of these CNVs.

The other major theme of convergence arose from synapse-related functions enriched across CNVs. Across all nine CNVs, both neuronal and glial cell types showed functional convergence on the synapse, with terms related to synaptic function and cellular communication for neurons and those related to synaptic support for glia. These results are in line with previous studies that described synaptic disruptions in model systems of the 22q11.2 deletion [60–62], 16p11.2 deletion [34,43,44],and 7q11.23 deletion [35,45]. We found that the direction and strength of DGE signal across neurons for genes encoding components of the pre-synapse and, to a lesser extent, post-synaptic density were correlated with those of OXPHOS. Synaptic activity is highly energy demanding and relies on proper mitochondrial function[46] . This correlation could therefore be explained by the critical role mitochondria play in maintaining synaptic transmission through local ATP production via oxidative phosphorylation [47–50]. We believe that the convergent changes in synaptic gene expression can at least in part be explained by correlated deficits in mitochondrial energy metabolism[51].

Leveraging the power of single-nucleus RNA-seq, we also discovered notable cell-type specific effects of high-risk CNVs, particularly in excitatory neurons and astrocytes. Excitatory neurons showed markedly higher number of DEGs to other cell types, with the

greatest number of total DEGs as well as proportion of the transcriptome differentially expressed. The 7q11.23 deletion exhibited notably larger effects in excitatory neurons compared to other cell types. This result aligns with previous reports that have described marked impacts on excitatory neurons carrying the 7q11.23 deletion in cellular and mouse models[35,52] . Astrocytes exhibited noticeable differences compared to other cell types with the greatest number of unique DEGs compared to the other eight cell types. Across the nine CNVs, functional signals in astrocytes converge on terms indicative of cellular stress responses, such as actin cytoskeleton organization and autophagy. The enrichment in actin cytoskeleton organization suggests that astrocytes may be undergoing structural remodeling[53]. Enrichment in autophagy-related processes in astrocytes could be a result of metabolic stress from changes in mitochondrial energy metabolism. These structural changes in astrocytes may serve to modulate neuroinflammatory responses, protect against oxidative stress, and influence neuronal survival [54–56]. Astrocytes engage in essential cellular communication with neurons and are critical for supporting neuronal survival and synaptic activity [57,58]. A recent study described a linked expression pattern of synaptic genes between astrocytes and neurons in human brain in the context of schizophrenia [59]. The astrocyte-specific changes in our study might represent an adaptive response to the pronounced changes in energy metabolism and synaptic function from neurons.

The CNV that showed the greatest cell-type specific effects in astrocytes was the 22q11.2 deletion, which exhibited a substantially higher number of DEGs in astrocytes compared to other cell types for both the individual CNV model and the 22q11.2 gene dosage model. The 22q11.2 deletion showed significant upregulation of terms related to hypoxia-induced stress and glycolysis, enriched not only in astrocytes but also across other cell types. Increased energy demand, e.g. from increased neuronal excitability or reduced oxygen conditions, are known to induce a shift from aerobic (OXPHOS) to anaerobic energy metabolism (glycolysis), particularly in astrocytes [60,61]. The upregulation of genes related to hypoxia stress response and glycolysis indicate that this may be the case in the 22q11.2 deletion. This observed upregulation in glycolysis aligns with previous work that found a higher activity of glycolysis compared to OXPHOS in children carrying the 22q11.2 deletion [62]. Our findings are additionally in line with previous studies that described increased neuronal excitability[63], impaired mitochondria function [37,38,64], and increased glycolysis biomarkers in the 22q11.2 deletion [62]. Taken together with the above-mentioned widespread changes in OXPHOS, these findings further support the notion of convergent energy metabolism changes in high-risk CNV carriers.

This study has several important limitations. While screening more than a thousand individuals across three brain banks, the number of carriers was limited by the relative rarity of neuropsychiatric CNVs in the population. While the results likely capture the most prominent gene expression changes, the sample is underpowered to reliably detect changes unique to individual CNVs. Although carriers were selected based on overlap with known recurrent high-risk CNVs, the exact breakpoints cannot be precisely determined based on SNP array data alone. This does not affect most DGE analyses but may

complicate those that depend on gene location relative to the CNV. Our DGE analysis was limited in resolution to nine major cell classes selected for frequency and even representation across the samples. Rarer cell types were present but were not resolved in differential gene expression analyses. While gene expression changes in post-mortem tissue provide valuable insight, the observed changes may include both those that are caused by the CNV and those that result from the associated illnesses and treatments. By matching of carriers and non-carriers by psychiatric diagnosis and treatment, we attempted to control for those variables, but such matching is never perfect. Studies of post-mortem tissue will require further experimental validation to robustly disentangle gene expression changes caused by the CNV from secondary effects.

Our study provides in-depth analysis of transcriptomic effects of nine recurrent high-risk neuropsychiatric CNVs, utilizing single-cell transcriptomics to achieve cell-type resolution in human brain. These findings offer valuable insights into shared pathophysiological mechanisms underlying these high-risk CNVs and their broader impacts on gene expression beyond proximal effects. Thousands of genes were differentially expressed across CNVs, with most DEGs located outside of the respective CNV regions and greater transcriptomic disruptions in deletions compared to reciprocal duplications. We found that gene expression changes associated with high-risk CNVs converge on mitochondrial energy metabolism, most markedly in the OXPHOS pathway, and in synaptic functions. Within neurons, changes in OXPHOS and synapse-related gene expression were correlated, although our data do not allow for mechanistic explanation. These findings advance the understanding of the neurobiology associated with CNVs and provide insight into potential convergent biology that could explain the phenotypic convergence of these CNVs through shared pathophysiology.

Moving forward, future research endeavors will need to functionally validate these findings of convergent mitochondrial disruptions across cell types and their connection to the synapse. Larger cohorts are essential to further validate and elucidate the complex genetic and molecular landscape of these rare, large CNVs. Additionally, it would be important to expand this work beyond the two cortical brain regions described in this study. Neuroimaging has discovered many changes in subcortical regions tied to CNVs [65,66], but our dataset was limited to dlPFC and ACC. Studying other brain regions could yield further insights and provide a better basis for integration with neuroimaging work, which would seek to answer how molecular changes may change brain physiology and lead to clinical phenotypes. Additionally, an avenue to further interrogate these CNVs would be through incorporating multi-omic approaches, e.g. proteomics or ATAC-seq. Multi-omics could offer promising insights into the landscape of CNV impacts, i.e. pairing gene expression with other genomic assays. This future work will provide a deeper understanding CNV biology and the relationship between mitochondrial and synaptic dysfunction, paving the way for the development of biomarkers, and more effective diagnostic and therapeutic strategies for individuals with high-risk neuropsychiatric CNVs.

Resource Availability

Raw and processed single-nucleus RNA-sequencing data will be deposited through NIH Data Archive (NDA) by the time of publication. Analysis code will be made available via Github.

Acknowledgements

We would like to thank the families who donated the brains of their loved ones for research, Maree Webster, Tom Blanchard, Robert Johnson, and the NeuroBioBank team for brain banking resources, dissection, and distribution of brain tissue from the Stanley Medical Research Institute and the University of Maryland Brain and Tissue Bank, Barbara Lipska, Ajeet Mandal, Bhaskar Kolachana, and Vesna Imamovic for brain banking resources and assistance with genotyping and brain tissue dissection at the NIMH Human Brain Collection Core, Yunlong He, Bao Tran, and other members of the CCR Sequencing Facility for nuclei isolation, library preparation, and sequencing, Asya Khlebodorova and Justin Lack for assistance with bulk RNA-seq data processing, all members of the Human Genetics Branch and Human Brain Collection Core (HBCC) for frequent discussions and ideas, and the NIH Biowulf team for providing computational resources. This project is supported by the NIMH Intramural Research Program: ZICMH002903 (Human Brain Collection Core) and ZIAMH002810 (Human Genetics Branch).

Author contributions

GD analyzed data and wrote the paper. NA performed CNV calling. NA, SM, PKA, FJM, and AS participated in study design and coordinated data acquisition. PKA performed and coordinated dissections. QX extracted RNA and measured RIN. QX, NF, and BK performed genotyping. NA, SM, PKA, AR, SL, AD, XJ, MDG, KFB, GH, PR, FJM, and AS assisted with data analysis and interpretation. FJM and AS assisted with writing the paper. All authors provided feedback.

Declaration of interests

The authors declare no competing interests.

Methods

Human brain tissue samples

> Human brain samples were derived from three brain banks: the NIMH Human Brain Collection Core (HBCC), the University of Maryland Brain and Tissue Bank (UMB), and the Stanley Medical Research Institute Brain Collection (SMRI). All brain samples were obtained using protocols approved by the Institutional Review Boards (IRB) with permission of the next-of-kin. Psychiatric diagnoses were inferred based on next-of-kin interviews and medical records, whenever available, according to DSM-IV and DSM-V criteria.

Genotyping and CNV Calling

Genotyping was performed as previously described [76] using a combination of Human1M-Duo (v3), HumanHap650Y (v3), HumanOmni5-Quad and Global Screening Array BeadChips. To call CNVs, we used PennCNV [77] with standard settings. We examined if any of the CNV calls overlapped with known CNVs previously associated with neuropsychiatric disorders [7]. B-allele frequencies and log R ratios were then plotted for the respective regions to confirm CNV status (Figure S1). We identified reciprocal deletions and duplications on chromosomes 22q11.2 (VCFS, proximal, A-D), 16p11.2 (proximal, BP4-BP5), 15q11.2 (BP1-BP2), 1q21.1 (all covering BP3-BP4, with variable involvement of the proximal BP2-BP3 TAR locus), and 7q11.23 (WBS) deletions. Since the exact CNV breakpoints could not be precisely determined, the ClinGen[21] coordinates of the corresponding CNVs were used (Table S1, S2).

Matching approach and demographics

Each individual harboring a known high-risk neuropsychiatric CNV, as defined above, was matched as close as possible to at least two presumptive non-carrier individuals from the same brain bank with similar demographic characteristics, based on sex, ethnicity, age of death, psychiatric diagnosis, and drug profile in toxicological screening (when available). In cases where two CNV carriers had very similar demographic features, the same non-carrier individuals served as matches for both. This matching approach created groups of three or more individuals with similar characteristics, which we refer to as ‘match groups’. Demographic variables across the dataset are summarized in Table S3.

Regions of interest

Whenever available, dorsolateral prefrontal cortex (dlPFC) and anterior cingulate cortex (ACC) were dissected. 50 mg of frozen tissue were used for nuclei isolation and single-nucleus RNA-sequencing, while another 50 mg were used for RNA isolation, followed by bulk RNA-sequencing and measurement of RNA integrity numbers (RIN). Quality-related variables across the dataset are summarized in Table S4.

Nuclei isolation

50 mg of frozen brain tissue were dissociated using Singulator 100 (S2 Genomics) in a prechilled nuclei isolation cartridge, following the Extended Nuclei Isolation protocol provided by the manufacturer. Nuclei were collected by centrifugation at 500g for 5 min at 4°C in a 15 mL tube, washed with 3 mL nuclei storage reagent (NSR) once, then washed with 3 mL washing buffer (PBS + 1% BSA + 0.2 U/µL RNase inhibitor) and filtered through 30 µm MARC SmartStrainer (Miltenyi Biotec) twice, resuspend in 800 µL washing buffer and filtered through 40 µm Flowmi cell strainer (Bel-Art) before loading.

Nuclei capture and library construction

Nuclei were captured and single-nucleus RNA-sequencing libraries were prepared following the 10X Genomics protocol CG000315 RevC: Chromium Next GEM Single Cell 3' Reagent Kits v3.1 (Dual Index). After washing and nuclei counting, approximately 8,000 nuclei from each sample were loaded on the 10X Genomics Chromium Controller. The Reverse Transcription and cDNA amplification were done following the manufacturer's protocol. 3' mRNA-seq gene expression libraries were prepared from 100ng cDNA on Sciclone G3 Liquid Handling Workstation (Revivity).

Next Generation Sequencing

Libraries were pooled and sequenced on NovaSeq S4 100 cycle flow cell with 2% PhiX spiked in. The paired-end dual index sequencing run was set up as Read1 (28 cycles, i7 Index: 10 cycles, i5 Index: 10 cycles) and Read 2 (90 cycles).

Single-Nucleus RNA-Sequencing Data Processing

Using the 10X Genomics CellRanger's processing software (v 7.0.0) [78], we demultiplexed the raw base call (BCL) files. The generated FASTQ files from CellRanger were aligned to the reference human genome GRCh38p14 to create raw feature-barcode count matrices.

Quality Control, Nuclei Filtering, and Clustering

All analyses were performed with R version 4.2.2. Data from adult (>10 years old) and infant (<1 year old) brain samples were processed separately. Starting with the adult data, we converted the single-nucleus RNA-sequencing counts into a Seurat object, which we then iteratively filtered and visualized with uniform manifold approximation and projections (UMAPs). To ensure data quality and reliability, we implemented a series of rigorous quality control and filtering steps (Figure S2). Utilizing SoupX (v 1.6.2) [79], we eliminated ambient RNA using the default values provided by the software.

Sample outliers were determined by Pearson correlation analysis of the normalized count matrix after pseudobulking across the 36,601 mapped genes. We removed two non-carrier samples (26 and 67) as they show correlation values ≤0.7, the lowest overall correlations compared to the other 67 samples (Figure S3A).

We employed scDblFinder (v 1.12.0) [68] to identify and filter out doublets in our data. scDblFinder was run on each sample using the recommended default values. We did not observe a noticeable concentration of doublets in any cluster (all <50% doublets) (Figure S2A). We thus removed any cell classified as a doublet according to scDblFinder.

We established the following criteria for initial nuclei inclusion: nuclei with >200 unique genes detected , > 500 number of transcripts, and a fraction of mitochondria-encoded genes (percent.mt) <5.0% (Figure S2C). Using the Seurat package (v

4.3.0) [19], the data underwent normalization, dimensionality reductions, and unsupervised clustering with Louvain clustering resolution of 1.0.

Harmony Integration and Additional Nuclei Filtering

To address potential batch effects biasing clustering across the dataset, we applied the Harmony algorithm (v 0.1.1) [80]. We used Harmony to control for match group, as it discerned cell type clusters with the best resolution. This was expected, since match groups were designed to control for the commonly observed confounding effects of sex, ethnicity, psychiatric diagnosis, and age. Utilizing match group as our integration variable with 50 dimensions, we improved cluster resolution between distinct cell types.

However, we still encountered noticeable bridging between major cell classes in the UMAP visualization, suggestive of persistent debris or poor-quality nuclei. The bridged clusters were also the clusters with the highest percent.mt suggesting that this bridging was an artifact of debris with high mitochondria content. We conducted quality filtering and Harmony integration by removing clusters with highest mean percent.mt. We started with clusters exceeding 2.5%, then removed clusters with percent.mt exceeding 2.0% and reran Harmony after both steps. Rigorous iterations of cluster removal by percent.mt filtering allowed for fewer bridged clusters and better cell type resolution (Figure S2B).

Cell Type Annotations and Sample Filtering

We utilized multiple approaches to annotate cell types and checked for consistency across approaches. First, we explored the expression of top marker genes associated with expected cell types. Then we employed Azimuth software (v 0.4.6) to annotate cell types, using the human motor cortex data set from the Allen Institute of Brain Research as a reference [19,20].

In picking cell type resolution, we aimed to achieve a high level of biologically meaningful resolution, while ensuring robust representation across all samples. Cells were annotated by identifying the most predominant cell type with each cluster from Seurat, based on subclass provided by Azimuth. Although Azimuth allowed us to resolve cell types in fine detail, we choose to limit our analysis to nine broad cell types based on cell type counts across samples, grouping some of the smaller populations of neurons and glia. (Figure S4A). We settled on nine well-represented cell types: astrocytes, microglia, oligodendrocytes, OPCs, vascular leptomeningeal cell (VLMC) and endothelial cells, upper- and lower-layer excitatory neurons, and medial ganglionic eminences (MGE)- and caudal ganglionic eminences (CGE)-derived inhibitory neurons (Figure 1E). VLMCs and endothelial cells were grouped together as they both had very low cell counts and serve similar biological functions. Low cell counts of VLMC and endothelial cell types are not unexpected in dlPFC and ACC[81]. We compared these annotations to marker gene expression to ensure confidence (Figure S4A-C). We observed similar proportions of the nine broad cell

types between dlPFC and ACC, after normalizing by the number of samples from both brain region (Figure S5A-B), consistent with prior work [17].

After classifying cell types, we filtered out samples with low neuron numbers. We removed sample 30, which contained fewer than 20 neurons across all four neuron types (Figure S3C,5C). Following the same criteria for outlier detection as used before, we analyzed Pearson correlations between samples on the normalized pseudobulked expression matrix across the 36601 genes (Figure S3B). Sample 30 also had low correlations across the 67 samples and, along with the underrepresentation of neuronal cell types, was thus excluded on this basis. After removing the outlier sample, each cell type had greater than 10 cells in each sample, except VLMCs.

Testing of Cell type Proportion

To rule out major cell type shifts and detect potential biases in the relative cell type abundance, we calculated the proportion of each cell type using the total cell count for each sample as the denominator and respective cell counts for each of the nine cell types as the numerator. Within each cell type, we ran a linear mixed model using lme4 (linear mixed-effects models) [82] procedure following the formulas:

(1) celltype_proportion~CNV +(1|BrainRegion) (1 | BrainNumber) +(1|Match_group)
(2) celltype_count~CNV +(1|BrainRegion) +(1 |BrainNumber) +(1 |Match_group)

We used square root and log transformations on the cell type proportion and counts respectively for the model to meet normality. Transformation for each cell type and model were chosen based on the Shapiro-Wilk test, requiring a p-value >0.05 of the residuals. We tested for the effect of CNV with match group, brain number, and region as random effects, since they are nested in the study design. We did not observe significant cell type shifts across most CNVs. One exception was the 15q11.2 deletion, which showed significant differences in cell type proportion for most cell types. The cell type counts for the 15q11.2 deletion, however, only showed significantly higher numbers of microglia and oligodendrocytes. These two cell types likely drive the significant results across proportions. Additionally, OPCs in the 22q11.2 deletion were significant for both cell type proportion and counts, with fewer OPCs in carriers than non-carriers (Table S5,S6). Given the limited sample size, we cannot conclude if differences in cell type proportion were due to biological or technical factors. However, the involvement of oligendrocytes, OPCs, and microglia suggests that they may be attributable to the ratio of grey and white matter in the dissection.

Differential Gene Expression

Differential gene expression (DGE) analysis was achieved using dreamlet (v 0.99.6) [18]. Dreamlet has the capability to use mixed models, which effectively addresses the nested nature of our data. Dreamlet adopts a pseudobulking strategy for single-

nucleus data, which has shown improved performance compared to single-cell approaches[83]. Dreamlet incorporates two-levels of precision weights, accommodating variations in both cell count and sequencing depth. This was beneficial to our study design to account for the observed variability in cell type counts described above.

Utilizing matched non-carriers enabled us to account for known confounders despite limited sample size and better target transcriptomic changes driven by each CNV. However, it was important to delineate how well our non-carrier controls were matched and determine correction for additional covariates as needed. To explore if or which other covariates should be included in the DGE model, we used variance partitioning [69] and linear regression analyses where we iteratively added in covariates. Using variance partitioning, we assessed correlations among the covariates and found that match group was significantly correlated with age, sex, ethnicity, etc. as expected (Figure S7A). Exploring the variance explained by each covariate indicated that CNV status and match group account for the largest amount of variance (Figure S7B).

We also performed principal component analysis and used linear regression to test for association of known covariates with the top principal components (PC), finding that match group was the most significant covariate (Figure S7C).

After residualizing for match group, brain region, and subject ID as random effects, we found no covariates were significantly associated with PC1- at p-value < 0.05 in all cell types (Figure S7D).

We therefore chose the following DGE model in dreamlet:

$$\sim CNV + (1|BrainRegion) + (1|BrainNumber) + (1|MatchGroup)$$

We followed default settings in dreamlet for gene and sample filtering, which requires a minimum number of five reads for a gene to be considered expressed in a sample, minimum of four samples passing cutoff for a cell type to be retained, and minimum of 40% of retained samples to have non-zero counts for a gene to be retained. Additionally, we also removed highly abundant mitochondrial-encoded and ribosomal genes that are often associated with sample quality: gene names starting with “MT-“, “RPL”, and “RPS”.

No covariate was significantly associated (p-value <0.05) with PC1 for all nine cell types, however some were significant in several cell types, such as cell type count. While dreamlet corrects for differences in cell type counts with precision weights, we created a more rigorous quality-covariate adjusted model for comparison to ensure differences in cell type count and sample quality between carriers and non-carriers did not influence DGE results. We selected quality covariates with a p-value < 0.01 in PC1 for at least one cell type to be added to a quality-covariate-adjusted model.

Three covariates met this criterion: RNA integrity number (RIN), mean number of genes per cell by sample (nFeature), and cell count of each cell type by sample.

$$\sim CNV + (1|BrainRegion) + (1|BrainNumber) + (1|MatchGroup) + RIN + nFeature + Cell.Type.Counts$$

We conducted a parallel analysis with these three additional fixed effects. We observed high correlations of t-statistics between the original model and the quality-covariate adjusted model, with the exception of oligodendrocytes in the 15q11.2 deletion. This was consistent with the significant differences in cell-type proportions in this sample, which had suggested more white matter in that dissection (Figure S7E). Given that this is likely driven by dissection, we believe the 15q11.2 oligodendrocyte results should be interpreted with caution.

To test for CNV gene dosage effects, we split carriers into groups by CNV, comprised of deletions, reciprocal duplications, and kept all non-carriers. We then classified samples by CNV gene dosage, with deletions classified with a 1, non-carriers a 2, and duplications a 3. This continuous gene dosage classification was then used as the fixed effect in the model, keeping the same grouping factors:

$$\sim GeneDosage + (1|BrainRegion) + (1|BrainNumber) + (1|MatchGroup)$$

This yielded genes that were positively associated with CNV gene dosage (positive logFC) and negatively associated with CNV gene dosage (negative logFC).

Functional Enrichments

Rank-based functional enrichment testing approaches were used to mitigate noise and unreliable results due to arbitrary p-value thresholds, known to influence threshold-based functional enrichment testing results [70,84]. GSEA was our primary functional enrichment testing method as it is widely used and acknowledged to be a strong candidate for discovering biological insights for RNA-seq data analysis [70,85]. The GSEA algorithm was implemented using fgsea (v 1.24) [86].

We tested for functional enrichments using Gene Ontology (GO) Biological Processes (BP) database[71,87–89]. Mitocarta [74] was used to obtain more detailed curated terms related to mitochondrial functions, including OXPHOS complexes. We also used SynGO to select terms related to the synapse [75] (Figure S16).

Results were adjusted for multiple testing using the Benjamini-Hochberg procedure. We applied a filtering criterion for functional enrichments, retaining only those with false discover rate (FDR) < 0.1 as a threshold for significance. To assess convergence across the dataset, we explored terms with the highest overlap in significant terms across both CNVs and cell types, in all 81 models (DGE across 9 CNVs and 9 cell types). Any term with the same intersection size was then ranked by FDR for selection of top terms. To explore direction of enrichment, we utilized the

normalized enrichment score (NES). This value accounts for the differences in gene set size.

Additionally, we explored convergence at the level of the CNV and at the level of the cell type, by extracting terms with the highest intersection size (1) across cell types for each CNV and conversely (2) across CNVs within each cell type. When terms had the same number of intersections, they were ranked by decreasing summed FDR value.

LOEUF Scores

To characterize genes within the CNV regions, the loss-of-function observed/expected upper fraction (LOEUF score) was used as a metric for risk. LOEUF scores are a continuous metric to quantify the intolerance of a gene to a loss of function variant and were obtained from the GnomAD database [26] (v4.1.0). To aggregate constraint scores for each CNV, inverse LOEUF scores (1/LOEUF) were summed for all CNV genes with available constraint metrics for each DGE model, as previously described[27].

Infant Samples Clustering, Cell Type Annotations

We also analyzed a smaller infant dataset, which included samples from the dlPFC and ACC of infant carriers of the 7q11.23 (WBS) or 15q11.2 (BP1-BP2) deletion, along with matched infant non-carriers (Figure 1A). The infant data underwent the same pipeline as the adult data. Moreover, the same QC steps were applied to the infant samples as with the adult samples, with Harmony integrations and removal of clusters with the highest percent mt. The nuclei were then annotated for cell type using Azimuth software (v 0.4.6) [19,20], with the human motor cortex data set from the Allen Institute of Brain Research as a reference [19] (Figure S6A-D). The only change in cell type classification was that oligodendrocytes and OPCs were merged into a single category labeled OPC_Oligo, due to the low number of oligodendrocytes in infant brain samples and the biological similarities between the two cell types.

We followed the same analysis pipeline as for the adult dataset for differential gene expression with exploration of metadata by variance partitioning and linear regression of covariates (Figure S14A-B). The final model stayed consistent with the adult dataset as:

$$\sim CNV + (1|BrainRegion) + (1|BrainNumber) + (1|MatchGroup)$$

Similarly, the default filtering values were retained from dreamlet for the infant dataset. Functional enrichments were achieved using fGSEA [86] on the GO BP and MitoCarta database [74].

References

1. Collins, R.L., Brand, H., Karczewski, K.J., Zhao, X., Alföldi, J., Francioli, L.C., Khera, A.V., Lowther, C., Gauthier, L.D., Wang, H., et al. (2020). A structural variation reference for medical and population genetics. Nature *581*, 444–451. https://doi.org/10.1038/s41586-020-2287-8.

2. Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Månér, S., Massa, H., Walker, M., Chi, M., et al. (2004). Large-Scale Copy Number Polymorphism in the Human Genome. Science *305*, 525–528. https://doi.org/10.1126/science.1098918.

3. Chen, S., Francioli, L.C., Goodrich, J.K., Collins, R.L., Kanai, M., Wang, Q., Alföldi, J., Watts, N.A., Vittal, C., Gauthier, L.D., et al. (2024). A genomic mutational constraint map using variation in 76,156 human genomes. Nature *625*, 92–100. https://doi.org/10.1038/s41586-023-06045-0.

4. Gudmundsson, S., Singer-Berk, M., Watts, N.A., Phu, W., Goodrich, J.K., Solomonson, M., Genome Aggregation Database Consortium, Rehm, H.L., MacArthur, D.G., and O'Donnell-Luria, A. (2022). Variant interpretation using population databases: Lessons from gnomAD. Hum. Mutat. *43*, 1012–1030. https://doi.org/10.1002/humu.24309.

5. Zhang, F., Gu, W., Hurles, M.E., and Lupski, J.R. (2009). Copy Number Variation in Human Health, Disease, and Evolution. Annu. Rev. Genomics Hum. Genet. *10*, 451–481. https://doi.org/10.1146/annurev.genom.9.081307.164217.

6. Bassett, A.S., Chow, E.W.C., Husted, J., Hodgkinson, K.A., Oechslin, E., Harris, L., and Silversides, C. (2009). Premature death in adults with 22q11.2 deletion syndrome. J. Med. Genet. *46*, 324–330. https://doi.org/10.1136/jmg.2008.063800.

7. Kendall, K.M., Rees, E., Bracher-Smith, M., Legge, S., Riglin, L., Zammit, S., O'Donovan, M.C., Owen, M.J., Jones, I., Kirov, G., et al. (2019). Association of Rare Copy Number Variants With Risk of Depression. JAMA Psychiatry *76*, 818. https://doi.org/10.1001/jamapsychiatry.2019.0566.

8. Mollon, J., Almasy, L., Jacquemont, S., and Glahn, D.C. (2023). The contribution of copy number variants to psychiatric symptoms and cognitive ability. Mol. Psychiatry *28*, 1480–1493. https://doi.org/10.1038/s41380-023-01978-4.

9. Rees, E., and Kirov, G. (2021). Copy number variation and neuropsychiatric illness. Curr. Opin. Genet. Dev. *68*, 57–63. https://doi.org/10.1016/j.gde.2021.02.014.

10. Sanders, S.J., Sahin, M., Hostyk, J., Thurm, A., Jacquemont, S., Avillach, P., Douard, E., Martin, C.L., Modi, M.E., Moreno-De-Luca, A., et al. (2019). A framework for the investigation of rare genetic disorders in neuropsychiatry. Nat. Med. *25*, 1477–1487. https://doi.org/10.1038/s41591-019-0581-5.

11. Malhotra, D., and Sebat, J. (2012). CNVs: Harbingers of a Rare Variant Revolution in Psychiatric Genetics. Cell *148*, 1223–1241. https://doi.org/10.1016/j.cell.2012.02.039.

12. Flaherty, E.K., and Brennand, K.J. (2017). Using hiPSCs to model neuropsychiatric copy number variations (CNVs) has potential to reveal underlying disease mechanisms. Brain Res. *1655*, 283–293. https://doi.org/10.1016/j.brainres.2015.11.009.

13. Modenato, C., Kumar, K., Moreau, C., Martin-Brevet, S., Huguet, G., Schramm, C., Jean-Louis, M., Martin, C.-O., Younis, N., Tamer, P., et al. (2021). Effects of eight neuropsychiatric copy number variants on human brain structure. Transl. Psychiatry *11*, 399. https://doi.org/10.1038/s41398-021-01490-9.

14. Mulle, J.G., Sullivan, P.F., and Hjerling-Leffler, J. (2021). Editorial overview: Rare CNV disorders and neuropsychiatric phenotypes: opportunities, challenges, solutions. Curr. Opin. Genet. Dev. *68*, iii–ix. https://doi.org/10.1016/j.gde.2021.05.002.

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

16. Parikshak, N.N., Swarup, V., Belgard, T.G., Irimia, M., Ramaswami, G., Gandal, M.J., Hartl, C., Leppa, V., Ubieta, L. de la T., Huang, J., et al. (2016). Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. Nature *540*, 423–427. https://doi.org/10.1038/nature20612.

17. Kim, B., Kim, D., Schulmann, A., Patel, Y., Caban-Rivera, C., Kim, P., Jambhale, A., Johnson, K.R., Feng, N., Xu, Q., et al. (2023). Cellular Diversity in Human Subgenual Anterior Cingulate and Dorsolateral Prefrontal Cortex by Single-Nucleus RNA-Sequencing. J. Neurosci. *43*, 3582–3597. https://doi.org/10.1523/JNEUROSCI.0830-22.2023.

18. Gabriel E. Hoffman, Donghoon Lee, Jaroslav Bendl, Prashant Fnu, Aram Hong, Clara Casey, Marcela Alvia, Zhiping Shao, Stathis Argyriou, Karen Therrien, et al. (2023). Efficient differential expression analysis of large-scale single cell transcriptomics data using dreamlet. bioRxiv, 2023.03.17.533005. https://doi.org/10.1101/2023.03.17.533005.

19. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573-3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.

20. Bakken, T.E., Jorstad, N.L., Hu, Q., Lake, B.B., Tian, W., Kalmbach, B.E., Crow, M., Hodge, R.D., Krienen, F.M., Sorensen, S.A., et al. (2021). Comparative cellular analysis of motor cortex in human, marmoset and mouse. Nature *598*, 111–119. https://doi.org/10.1038/s41586-021-03465-8.

21. Rehm, H.L., Berg, J.S., Brooks, L.D., Bustamante, C.D., Evans, J.P., Landrum, M.J., Ledbetter, D.H., Maglott, D.R., Martin, C.L., Nussbaum, R.L., et al. (2015). ClinGen — The Clinical Genome Resource. N. Engl. J. Med. *372*, 2235–2242. https://doi.org/10.1056/NEJMsr1406261.

22. Weiner, D.J., Ling, E., Erdin, S., Tai, D.J.C., Yadav, R., Grove, J., Fu, J.M., Nadig, A., Carey, C.E., Baya, N., et al. (2022). Statistical and functional convergence of common and rare genetic influences on autism at chromosome 16p. Nat. Genet. *54*, 1630–1639. https://doi.org/10.1038/s41588-022-01203-y.

23. Stefansson, H., Meyer-Lindenberg, A., Steinberg, S., Magnusdottir, B., Morgen, K., Arnarsdottir, S., Bjornsdottir, G., Walters, G.B., Jonsdottir, G.A., Doyle, O.M., et al. (2014). CNVs conferring risk of autism or schizophrenia affect cognition in controls. Nature *505*, 361–366. https://doi.org/10.1038/nature12818.

24. Tansey, K.E., Rees, E., Linden, D.E., Ripke, S., Chambert, K.D., Moran, J.L., McCarroll, S.A., Holmans, P., Kirov, G., Walters, J., et al. (2016). Common alleles contribute to schizophrenia in CNV carriers. Mol. Psychiatry *21*, 1085–1089. https://doi.org/10.1038/mp.2015.143.

25. Moreau, C.A., Ching, C.R., Kumar, K., Jacquemont, S., and Bearden, C.E. (2021). Structural and functional brain alterations revealed by neuroimaging in CNV carriers. Curr. Opin. Genet. Dev. *68*, 88–98. https://doi.org/10.1016/j.gde.2021.03.002.

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

27. Alexander-Bloch, A., Huguet, G., Schultz, L.M., Huffnagle, N., Jacquemont, S., Seidlitz, J., Saci, Z., Moore, T.M., Bethlehem, R.A.I., Mollon, J., et al. (2022). Copy Number Variant Risk Scores Associated With Cognition, Psychopathology, and Brain Structure in Youths in the Philadelphia Neurodevelopmental Cohort. JAMA Psychiatry *79*, 699–709. https://doi.org/10.1001/jamapsychiatry.2022.1017.

28. Huguet, G., Schramm, C., Douard, E., Tamer, P., Main, A., Monin, P., England, J., Jizi, K., Renne, T., Poirier, M., et al. (2021). Genome-wide analysis of gene dosage in 24,092 individuals estimates that 10,000 genes modulate cognitive ability. Mol. Psychiatry *26*, 2663–2676. https://doi.org/10.1038/s41380-020-00985-z.

29. Collins, R.L., Glessner, J.T., Porcu, E., Lepamets, M., Brandon, R., Lauricella, C., Han, L., Morley, T., Niestroj, L.-M., Ulirsch, J., et al. (2022). A cross-disorder dosage sensitivity map of the human genome. Cell *185*, 3041-3055.e25. https://doi.org/10.1016/j.cell.2022.06.036.

30. Kopal, J., Kumar, K., Saltoun, K., Modenato, C., Moreau, C.A., Martin-Brevet, S., Huguet, G., Jean-Louis, M., Martin, C.-O., Saci, Z., et al. (2023). Rare CNVs and phenome-wide profiling highlight brain structural divergence and phenotypical convergence. Nat. Hum. Behav. *7*, 1001–1017. https://doi.org/10.1038/s41562-023-01541-9.

31. Raznahan, A., Won, H., Glahn, D.C., and Jacquemont, S. (2022). Convergence and Divergence of Rare Genetic Disorders on Brain Phenotypes: A Review. JAMA Psychiatry *79*, 818. https://doi.org/10.1001/jamapsychiatry.2022.1450.

32. Rosenfeld, J.A., Traylor, R.N., Schaefer, G.B., McPherson, E.W., Ballif, B.C., Klopocki, E., Mundlos, S., Shaffer, L.G., Aylsworth, A.S., and 1q21.1 Study Group (2012). Proximal microdeletions and microduplications of 1q21.1 contribute to variable abnormal phenotypes. Eur. J. Hum. Genet. *20*, 754–761. https://doi.org/10.1038/ejhg.2012.6.

33. Béland-Millar, A., Kirby, A., Truong, Y., Ouellette, J., Yandiev, S., Bouyakdan, K., Pileggi, C., Naz, S., Yin, M., Carrier, M., et al. (2023). 16p11.2 haploinsufficiency reduces mitochondrial biogenesis in brain endothelial cells and alters brain metabolism in adult mice. Cell Rep. *42*, 112485. https://doi.org/10.1016/j.celrep.2023.112485.

34. Tai, D.J.C., Razaz, P., Erdin, S., Gao, D., Wang, J., Nuttle, X., De Esch, C.E., Collins, R.L., Currall, B.B., O'Keefe, K., et al. (2022). Tissue- and cell-type-specific molecular and functional signatures of 16p11.2 reciprocal genomic disorder across mouse brain and human neuronal models. Am. J. Hum. Genet. *109*, 1789–1813. https://doi.org/10.1016/j.ajhg.2022.08.012.

35. Tebbenkamp, A.T.N., Varela, L., Choi, J., Paredes, M.I., Giani, A.M., Song, J.E., Sestan-Pesa, M., Franjic, D., Sousa, A.M.M., Liu, Z.-W., et al. (2018). The 7q11.23 Protein DNAJC30 Interacts with ATP Synthase and Links Mitochondria to Brain Development. Cell *175*, 1088-1104.e23. https://doi.org/10.1016/j.cell.2018.09.014.

36. Ferrari, M., and Stagi, S. (2021). Oxidative Stress in Down and Williams-Beuren Syndromes: An Overview. Molecules *26*, 3139. https://doi.org/10.3390/molecules26113139.

37. Devaraju, P., and Zakharenko, S.S. (2017). Mitochondria in complex psychiatric disorders: Lessons from mouse models of 22q11.2 deletion syndrome: Hemizygous deletion of several mitochondrial genes in the 22q11.2 genomic region can lead to symptoms associated with neuropsychiatric disease. BioEssays *39*, 1600177. https://doi.org/10.1002/bies.201600177.

38. Li, J., Ryan, S.K., Deboer, E., Cook, K., Fitzgerald, S., Lachman, H.M., Wallace, D.C., Goldberg, E.M., and Anderson, S.A. (2019). Mitochondrial deficits in human iPSC-derived neurons from patients with 22q11.2 deletion syndrome and schizophrenia. Transl. Psychiatry *9*, 302. https://doi.org/10.1038/s41398-019-0643-y.

39. Purcell, R.H., Sefik, E., Werner, E., King, A.T., Mosley, T.J., Merritt-Garza, M.E., Chopra, P., McEachin, Z.T., Karne, S., Raj, N., et al. (2023). Cross-species analysis identifies mitochondrial dysregulation as a functional consequence of the schizophrenia-associated 3q29 deletion. Sci. Adv. *9*, eadh0558. https://doi.org/10.1126/sciadv.adh0558.

40. Forsyth, J.K., Nachun, D., Gandal, M.J., Geschwind, D.H., Anderson, A.E., Coppola, G., and Bearden, C.E. (2020). Synaptic and Gene Regulatory Mechanisms in Schizophrenia, Autism, and 22q11.2 Copy Number Variant–Mediated Risk for Neuropsychiatric Disorders. Biol. Psychiatry *87*, 150–163. https://doi.org/10.1016/j.biopsych.2019.06.029.

41. Nehme, R., Pietiläinen, O., Artomov, M., Tegtmeyer, M., Valakh, V., Lehtonen, L., Bell, C., Singh, T., Trehan, A., Sherwood, J., et al. (2022). The 22q11.2 region regulates presynaptic gene-products linked to schizophrenia. Nat. Commun. *13*, 3690. https://doi.org/10.1038/s41467-022-31436-8.

42. Sun, Z., Williams, D.J., Xu, B., and Gogos, J.A. (2018). Altered function and maturation of primary cortical neurons from a 22q11.2 deletion mouse model of schizophrenia. Transl. Psychiatry *8*, 85. https://doi.org/10.1038/s41398-018-0132-8.

43. Urresti, J., Zhang, P., Moran-Losada, P., Yu, N.-K., Negraes, P.D., Trujillo, C.A., Antaki, D., Amar, M., Chau, K., Pramod, A.B., et al. (2021). Cortical organoids model early brain development disrupted by 16p11.2 copy number variants in autism. Mol. Psychiatry *26*, 7560–7580. https://doi.org/10.1038/s41380-021-01243-6.

44. Portmann, T., Yang, M., Mao, R., Panagiotakos, G., Ellegood, J., Dolen, G., Bader, P.L., Grueter, B.A., Goold, C., Fisher, E., et al. (2014). Behavioral Abnormalities and Circuit Defects in the Basal Ganglia of a Mouse Model of 16p11.2 Deletion Syndrome. Cell Rep. *7*, 1077–1092. https://doi.org/10.1016/j.celrep.2014.03.036.

45. Gao, M.C., Bellugi, U., Dai, L., Mills, D.L., Sobel, E.M., Lange, K., and Korenberg, J.R. (2010). Intelligence in Williams Syndrome Is Related to STX1A, Which Encodes a Component of the Presynaptic SNARE Complex. PLoS ONE *5*, e10292. https://doi.org/10.1371/journal.pone.0010292.

46. Sun, T., Qiao, H., Pan, P.-Y., Chen, Y., and Sheng, Z.-H. (2013). Motile Axonal Mitochondria Contribute to the Variability of Presynaptic Strength. Cell Rep. *4*, 413–419. https://doi.org/10.1016/j.celrep.2013.06.040.

47. Devine, M.J., and Kittler, J.T. (2018). Mitochondria at the neuronal presynapse in health and disease. Nat. Rev. Neurosci. *19*, 63–80. https://doi.org/10.1038/nrn.2017.170.

48. Sullivan, C.R., O'Donovan, S.M., McCullumsmith, R.E., and Ramsey, A. (2018). Defects in Bioenergetic Coupling in Schizophrenia. Biol. Psychiatry *83*, 739–750. https://doi.org/10.1016/j.biopsych.2017.10.014.

49. Das, S.C., Hjelm, B.E., Rollins, B.L., Sequeira, A., Morgan, L., Omidsalar, A.A., Schatzberg, A.F., Barchas, J.D., Lee, F.S., Myers, R.M., et al. (2022). Mitochondria DNA copy number, mitochondria DNA total somatic deletions, Complex I activity, synapse number, and synaptic mitochondria number are altered in schizophrenia and bipolar disorder. Transl. Psychiatry *12*, 353. https://doi.org/10.1038/s41398-022-02127-1.

50. Dienel, S.J., Schoonover, K.E., and Lewis, D.A. (2022). Cognitive Dysfunction and Prefrontal Cortical Circuit Alterations in Schizophrenia: Developmental Trajectories. Biol. Psychiatry *92*, 450–459. https://doi.org/10.1016/j.biopsych.2022.03.002.

51. Sheng, Z.-H., and Cai, Q. (2012). Mitochondrial transport in neurons: impact on synaptic homeostasis and neurodegeneration. Nat. Rev. Neurosci. *13*, 77–93. https://doi.org/10.1038/nrn3156.

52. López-Tobón, A., Shyti, R., Villa, C.E., Cheroni, C., Fuentes-Bravo, P., Trattaro, S., Caporale, N., Troglio, F., Tenderini, E., Mihailovich, M., et al. GTF2I dosage regulates neuronal differentiation and social behavior in 7q11.23 neurodevelopmental disorders. Sci. Adv. *9*, eadh2726. https://doi.org/10.1126/sciadv.adh2726.

53. Escartin, C., Galea, E., Lakatos, A., O'Callaghan, J.P., Petzold, G.C., Serrano-Pozo, A., Steinhäuser, C., Volterra, A., Carmignoto, G., Agarwal, A., et al. (2021). Reactive astrocyte nomenclature, definitions, and future directions. Nat. Neurosci. *24*, 312–325. https://doi.org/10.1038/s41593-020-00783-4.

54. Schiweck, J., Eickholt, B.J., and Murk, K. (2018). Important Shapeshifter: Mechanisms Allowing Astrocytes to Respond to the Changing Nervous System During Development, Injury and Disease. Front. Cell. Neurosci. *12*, 261. https://doi.org/10.3389/fncel.2018.00261.

55. Kulkarni, A., Dong, A., Kulkarni, V.V., Chen, J., Laxton, O., Anand, A., and Maday, S. (2020). Differential regulation of autophagy during metabolic stress in astrocytes and neurons. Autophagy *16*, 1651–1667. https://doi.org/10.1080/15548627.2019.1703354.

56. Qaisiya, M., Mardešić, P., Pastore, B., Tiribelli, C., and Bellarosa, C. (2017). The activation of autophagy protects neurons and astrocytes against bilirubin-induced cytotoxicity. Neurosci. Lett. *661*, 96–103. https://doi.org/10.1016/j.neulet.2017.09.056.

57. Allen, N.J., and Lyons, D.A. (2018). Glia as architects of central nervous system formation and function. Science *362*, 181–185. https://doi.org/10.1126/science.aat0473.

58. Allen, N.J., and Eroglu, C. (2017). Cell Biology of Astrocyte-Synapse Interactions. Neuron *96*, 697–708. https://doi.org/10.1016/j.neuron.2017.09.056.

59. Ling, E., Nemesh, J., Goldman, M., Kamitaki, N., Reed, N., Handsaker, R.E., Genovese, G., Vogelgsang, J.S., Gerges, S., Kashin, S., et al. (2024). A concerted neuron–astrocyte program declines in ageing and schizophrenia. Nature *627*, 604–611. https://doi.org/10.1038/s41586-024-07109-5.

60. Bolaños, J.P. (2016). Bioenergetics and redox adaptations of astrocytes to neuronal activity. J. Neurochem. *139*, 115–125. https://doi.org/10.1111/jnc.13486.

61. Yellen, G. (2018). Fueling thought: Management of glycolysis and oxidative phosphorylation in neuronal metabolism. J. Cell Biol. *217*, 2235–2246. https://doi.org/10.1083/jcb.201803152.

62. Napoli, E., Tassone, F., Wong, S., Angkustsiri, K., Simon, T.J., Song, G., and Giulivi, C. (2015). Mitochondrial Citrate Transporter-dependent Metabolic Signature in the 22q11.2 Deletion Syndrome. J. Biol. Chem. *290*, 23240–23253. https://doi.org/10.1074/jbc.M115.672360.

63. Khan, T.A., Revah, O., Gordon, A., Yoon, S.-J., Krawisz, A.K., Goold, C., Sun, Y., Kim, C.H., Tian, Y., Li, M.-Y., et al. (2020). Neuronal defects in a human cellular model of 22q11.2 deletion syndrome. Nat. Med. *26*, 1888–1898. https://doi.org/10.1038/s41591-020-1043-9.

64. Gokhale, A., Hartwig, C., Freeman, A.A.H., Bassell, J.L., Zlatic, S.A., Sapp Savas, C., Vadlamudi, T., Abudulai, F., Pham, T.T., Crocker, A., et al. (2019). Systems Analysis of the 22q11.2 Microdeletion Syndrome Converges on a Mitochondrial Interactome Necessary for Synapse Function and Behavior. J. Neurosci., 1983–18. https://doi.org/10.1523/JNEUROSCI.1983-18.2019.

65. Kumar, K., Modenato, C., Moreau, C., Ching, C.R.K., Harvey, A., Martin-Brevet, S., Huguet, G., Jean-Louis, M., Douard, E., Martin, C.-O., et al. (2023). Subcortical Brain Alterations in Carriers of Genomic Copy Number Variants. Am. J. Psychiatry *180*, 685–698. https://doi.org/10.1176/appi.ajp.20220304.

66. Warland, A., Kendall, K.M., Rees, E., Kirov, G., and Caseras, X. (2020). Schizophrenia-associated genomic copy number variants and subcortical brain volumes in the UK Biobank. Mol. Psychiatry *25*, 854–862. https://doi.org/10.1038/s41380-019-0355-y.

67. Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nat. Biotechnol. *33*, 495–502. https://doi.org/10.1038/nbt.3192.

68. Germain, P.-L., Lun, A., Garcia Meixide, C., Macnair, W., and Robinson, M.D. (2022). Doublet identification in single-cell sequencing data using scDblFinder. F1000Research *10*, 979. https://doi.org/10.12688/f1000research.73600.2.

69. Hoffman, G.E., and Schadt, E.E. (2016). variancePartition: interpreting drivers of variation in complex gene expression studies. BMC Bioinformatics *17*, 483. https://doi.org/10.1186/s12859-016-1323-z.

70. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. *102*, 15545–15550. https://doi.org/10.1073/pnas.0506580102.

71. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene Ontology: tool for the unification of biology. Nat. Genet. *25*, 25–29. https://doi.org/10.1038/75556.

72. Kanehisa, M. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. *28*, 27–30. https://doi.org/10.1093/nar/28.1.27.

73. Wujak, M., Veith, C., Wu, C.-Y., Wilke, T., Kanbagli, Z.I., Novoyatleva, T., Guenther, A., Seeger, W., Grimminger, F., Sommer, N., et al. (2021). Adenylate Kinase 4—A Key Regulator of Proliferation and Metabolic Shift in Human Pulmonary Arterial Smooth Muscle Cells via Akt and HIF-1α Signaling Pathways. Int. J. Mol. Sci. *22*, 10371. https://doi.org/10.3390/ijms221910371.

74. Rath, S., Sharma, R., Gupta, R., Ast, T., Chan, C., Durham, T.J., Goodman, R.P., Grabarek, Z., Haas, M.E., Hung, W.H.W., et al. (2021). MitoCarta3.0: an updated mitochondrial proteome now with sub-organelle localization and pathway annotations. Nucleic Acids Res. *49*, D1541–D1547. https://doi.org/10.1093/nar/gkaa1011.

75. Koopmans, F., Van Nierop, P., Andres-Alonso, M., Byrnes, A., Cijsouw, T., Coba, M.P., Cornelisse, L.N., Farrell, R.J., Goldschmidt, H.L., Howrigan, D.P., et al. (2019). SynGO: An Evidence-Based, Expert-Curated Knowledge Base for the Synapse. Neuron *103*, 217-234.e4. https://doi.org/10.1016/j.neuron.2019.05.002.

76. Duncan, L., Shen, H., Schulmann, A., Li, T., Kolachana, B., Mandal, A., Feng, N., Auluck, P., and Marenco, S. (2023). Polygenic scores for psychiatric disorders in a diverse postmortem brain tissue cohort. Neuropsychopharmacology *48*, 764–772. https://doi.org/10.1038/s41386-022-01524-w.

77. Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S.F.A., Hakonarson, H., and Bucan, M. (2007). PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res. *17*, 1665–1674. https://doi.org/10.1101/gr.6861907.

78. Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. Nat. Commun. *8*, 14049. https://doi.org/10.1038/ncomms14049.

79. Young, M.D., and Behjati, S. (2020). SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. GigaScience *9*, giaa151. https://doi.org/10.1093/gigascience/giaa151.

80. Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P., and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. Nat. Methods *16*, 1289–1296. https://doi.org/10.1038/s41592-019-0619-0.

81. Huuki-Myers, L.A., Spangler, A., Eagles, N.J., Montgomery, K.D., Kwon, S.H., Guo, B., Grant-Peters, M., Divecha, H.R., Tippani, M., Sriworarat, C., et al. (2024). A data-driven single-cell and spatial transcriptomic map of the human prefrontal cortex. Science *384*, eadh1938. https://doi.org/10.1126/science.adh1938.

82. Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using **lme4**. J. Stat. Softw. *67*. https://doi.org/10.18637/jss.v067.i01.

83. Junttila, S., Smolander, J., and Elo, L.L. (2022). Benchmarking methods for detecting differential states between conditions from multi-subject single-cell RNA-seq data. Brief. Bioinform. *23*, bbac286. https://doi.org/10.1093/bib/bbac286.

84. Zyla, J., Marczyk, M., Weiner, J., and Polanska, J. (2017). Ranking metrics in gene set enrichment analysis: do they matter? BMC Bioinformatics *18*, 256. https://doi.org/10.1186/s12859-017-1674-0.

85. Candia, J., and Ferrucci, L. (2024). Assessment of Gene Set Enrichment Analysis using curated RNA-seq-based benchmarks. PLOS ONE *19*, e0302696. https://doi.org/10.1371/journal.pone.0302696.

86. Korotkevich, G., Sukhov, V., Budin, N., Shpak, B., Artyomov, M.N., and Sergushichev, A. (2016). Fast gene set enrichment analysis. Preprint, https://doi.org/10.1101/060012 https://doi.org/10.1101/060012.

87. Luecken, M.D., and Theis, F.J. (2019). Current best practices in single-cell RNA-seq analysis: a tutorial. Mol. Syst. Biol. *15*, e8746. https://doi.org/10.15252/msb.20188746.

88. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888-1902.e21. https://doi.org/10.1016/j.cell.2019.05.031.

89. The Gene Ontology Consortium, Aleksander, S.A., Balhoff, J., Carbon, S., Cherry, J.M., Drabkin, H.J., Ebert, D., Feuermann, M., Gaudet, P., Harris, N.L., et al. (2023). The Gene Ontology knowledgebase in 2023. GENETICS *224*, iyad031. https://doi.org/10.1093/genetics/iyad031.