Science Advances
AAAS
Manuscript Template

Title: Cross-species analysis identifies mitochondrial dysregulation as a functional consequence of the schizophrenia-associated 3q29 deletion

Short title: Transcriptomic effects of 3q29 deletion

Authors

Ryan H. Purcell[1,2]†*, Esra Sefik[3]†, Erica Werner[2], Alexia T. King[3], Trenell J. Mosley[3], Megan E. Merritt-Garza[2], Pankaj Chopra[3], Zachary T. McEachin[1,2], Sridhar Karne[2], Nisha Raj[1,2], Brandon J. Vaglio[4], Dylan Sullivan[4], Bonnie L. Firestein[4], Kedamawit Tilahun[2], Maxine I. Robinette[2], Stephen T. Warren[3], Zhexing Wen[2,5], Victor Faundez[2], Steven A. Sloan[3], Gary J. Bassell*[1,2], and Jennifer G. Mulle*[3,6]

Affiliations

[1]Laboratory of Translational Cell Biology, Emory University School of Medicine, Atlanta, GA, USA.
[2]Department of Cell Biology, Emory University School of Medicine, Atlanta, GA, USA.
[3]Department of Human Genetics, Emory University School of Medicine, Atlanta, GA, USA.
[4]Department of Cell Biology and Neuroscience, Rutgers University, Piscataway, NJ, USA.
[5]Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, GA, USA.
[6]Current address: Department of Psychiatry, Rutgers University, Piscataway, NJ, USA.

†These authors contributed equally to this work.
*Correspondence: RHP ryan.purcell@emory.edu; GJB gary.bassell@emory.edu; JGM jm2618@cabm.rutgers.edu

Abstract

The 1.6Mb deletion at chromosome 3q29 (3q29Del) is the strongest identified genetic risk factor for schizophrenia, but the effects of this variant on neurodevelopment are not well understood. We interrogated the developing neural transcriptome in two experimental model systems with complementary advantages: isogenic human cortical organoids and isocortex from the 3q29Del mouse model. We profiled transcriptomes from isogenic cortical organoids that were aged for 2 months and 12 months, as well as perinatal mouse isocortex, all at single-cell resolution. Systematic pathway analysis implicated dysregulation of mitochondrial function and energy metabolism. These molecular signatures were supported by analysis of oxidative phosphorylation protein complex expression in mouse brain and assays of mitochondrial function in engineered cell lines, which revealed a lack of metabolic flexibility and a contribution of the 3q29 gene *PAK2*. Together these data indicate that metabolic disruption is associated with 3q29Del and is conserved across species.

Teaser

Single-cell transcriptomic profiling reveals effects of the schizophrenia-associated 3q29 deletion on mitochondrial function.

MAIN TEXT

Introduction

Rare variants have now been identified that confer extraordinarily high risk for schizophrenia (SCZ). Functional study of these variants may yield insights into the molecular and cellular impairments that ultimately give rise to psychosis. By restricting investigation to a single variant, etiologic heterogeneity is vastly reduced, which may lead to better discrimination of causal mechanisms. To date, the strongest identified single genetic risk factor for SCZ is the 3q29 deletion (3q29Del), a copy number variant (CNV) that encompasses 22 protein-coding genes and is located near the telomeric end of human chromosome 3 (*3, 4*). Hemizygous loss of this set of genes is associated with at least a 40-fold increase in risk for SCZ (*5, 6*); this deletion also increases risk for additional neurodevelopmental and psychiatric conditions, including intellectual disability, autism spectrum disorder (ASD), and attention-deficit/hyperactivity disorder (ADHD) (OMIM #609425) (*7*).

Exciting developments in molecular neuroscience have led to powerful new tools for the investigation of neurobiology of mental health disorders. In this study, we leverage two state-of-the-art experimental model systems, which together amplify the rigor of our approach. The starting substrate for these experiments are two CRISPR-engineered experimental systems: newly-generated isogenic human induced pluripotent stem (iPS) cells, where we have precisely introduced the 3q29Del using CRIPSR/Cas9 and the 3q29 mouse model (B6.Del16$^{+/Bdh1\text{-}Tfrc}$), which bears complete synteny to the human 3q29 interval and displays neurodevelopmental and somatic correlates of human syndromic phenotypes (*8-10*). These experimental systems offer complementary advantages; cortical organoids are the current gold standard model of early human cortical development *in vitro,* whereas the syntenic 3q29Del mice provide a source of brain tissue from a physiological context. We hypothesized that a transcriptomic analysis of differentially expressed genes in developing cortical tissue would provide relatively unbiased insights into underlying mechanisms of cellular dysfunction. We reasoned that the set of transcriptomic changes observed in *both* model systems are likely attributable to the 3q29Del and may underlie core phenotypes.

To investigate the biological effects of the 3q29Del, we performed single-cell mRNA sequencing (scRNA-seq) in isogenic human cortical organoids at both early (2-month) and late (12-month) developmental time points and in perinatal (p7) mouse isocortex. We devised a strategy to systematically identify the most salient transcriptomic effects, both globally and in specific neural cell-types, to identify molecular phenotypes for functional analysis. This strategy led us to a dysregulated transcriptome linked to mitochondria, which displayed both early and prolonged changes that were subsequently supported by orthogonal analyses of protein expression and functional assays in engineered HEK cells and multiple cohorts of human iPSC-derived neural progenitor cells. In the context of emerging reports of mitochondrial phenotypes associated with other risk alleles such as 22q11.2 deletion (*11, 12*), these results point to mitochondria as a possible site of convergent biology downstream of discrete neurodevelopmental variants.

Results

Generating isogenic 3q29Del iPSC lines

To isolate the effects of the 3q29Del from variable human genetic backgrounds, we introduced the full 1.6 Mb deletion into an iPSC line derived from a neurotypical female individual by adapting a method previously used to generate isogenic iPSC lines carrying other neurodevelopmental CNVs (*13*). Like most recurrent CNVs, the 3q29Del is flanked by low-copy

repeats (LCRs) or segmental duplications (SDs), which are multi-kilobase stretches of highly homologous sequence (*14*) that are likely involved in the formation of structural variants such as CNVs (*15*). We targeted this homologous sequence with a single guide RNA that is predicted to cut at one site within each 3q29 SD (*16*) and isolated three clonal lines carrying the 3q29Del (Fig. S1). A neurotypical male iPS cell line (one clone) was also engineered to carry the 3q29Del using

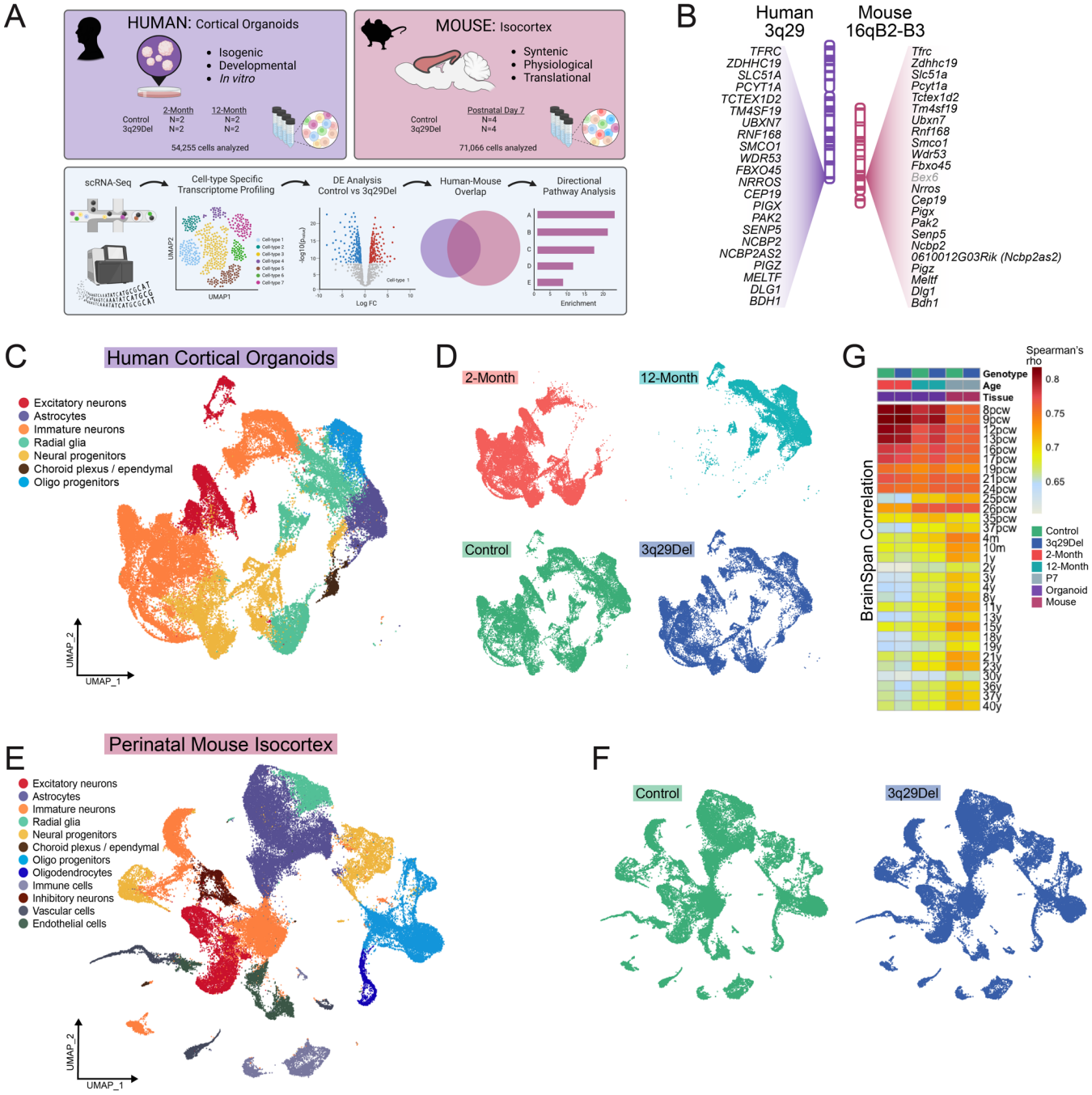


Fig 1. Cross-species single-cell sequencing. (**A**) A single-cell RNA-sequencing experiment was performed in isogenic human induced-pluripotent stem cell (iPSC)-derived cortical organoids at two time points and in postnatal day 7 mouse isocortex. An overview of the strategy to collect and filter differential gene expression data from both model systems is illustrated. (**B**) The human 3q29 deletion locus is nearly perfectly syntenic with a region of mouse chromosome 16, with the same gene order inverted. Corresponding loci are illustrated in the same orientation to facilitate clearer cross-species comparison. *Bex6* (in gray) is the only gene present in the mouse, not in the human locus. (**C** and **E**) UMAP dimensionality reduction plots colored by the main cell-types identified in human (**C**) and mouse (**E**) experiments. Human and mouse cells showed no obvious difference in gross distribution by genotype (**D**, **F**) but human cells were clearly divided in their transcriptomic clustering patterns by time point (D, top). The average expression profile of each sample was correlated (Spearman) to BrainSpan gene expression data profiling the human brain transcriptome in postmortem specimens across the lifespan (*2*) (**G**). Abbreviations: pcw, post-conception weeks (prenatal); m, months (postnatal); y, years (postnatal).

the methods described above. All clones retained normal iPSC morphology and karyotype, and genome-wide optical mapping analyses revealed no off-target structural variants.

Single-cell transcriptomics in developing mouse and human cortical tissue

Four deletion clones and four clones of the parent cell lines (3 female, 1 male per group) were differentiated to dorsal forebrain cortical organoids by established methods (Fig. S1F) (*17*). Single-cell transcriptomes were produced from multiple organoids from two clonal female lines per genotype at 2-months and 12-months into *in vitro* differentiation to capture a broad diversity of developing and mature cell-types (Fig. S2-S3). 54,255 cells were included in the human cortical organoid analysis (54% Control, Fig. S4). A mean of 2,805 genes were detected in each cell.

The 3q29Del mouse has been previously reported by two independent groups to express neuropsychiatric phenotypes including alterations in startle responses and social interactions (*8, 10*). Postnatal day 7 (P7) was chosen for tissue dissociation and single-cell sequencing to capture an array of mature and developing cell-types (Fig. S5-S6). 71,066 cortical cells were isolated from four male mice per genotype and included in the mouse scRNA-seq analysis (52.9% Control, Fig. S7). The mean number of genes detected in each cell was 2,920.

Fig. 1C and 1E show the major cell-types with distinct expression profiles that were isolated in each sequencing experiment. As expected, 2-month and 12-month human cortical organoids contained many of the cell-types that were also found in the perinatal mouse isocortex including excitatory neurons, astrocytes, immature neurons, radial glia/neural stem cells, neural progenitors, choroid plexus/ependymal cells, and oligodendrocyte progenitors (Fig. S8). In addition to these cells, we also identified immune cells, inhibitory neurons, vascular cells, and endothelial cells in the mouse experiment (Fig. S9). Notably, in both experiments, we did not observe large-scale changes in cell clustering by genotype (Fig. 1D, 1F), but did observe a stark division in human cell clustering by time point (Fig. 1D). Indeed, most human clusters were comprised almost entirely of cells from a single time point, and only one cluster (cl. 2, annotated as migrating neuroblasts) was nearly evenly split by time point, consistent with developmentally regulated shifts in cell-type composition (Fig. S4). Predictably, astrocytes, oligodendrocyte progenitors, and upper layer excitatory neurons were not yet present in 2-month organoids (*17*) but were indeed found in 12-month organoids.

To better understand the window of cortical development that our experimental models best reproduce, we compared the average expression profile of each sample to postmortem human brain transcriptomes from the BrainSpan database (*2*) (Fig. 1G). Spearman correlations revealed that both 2-month and 12-month organoids best matched very early phases of human brain development (8-9 post-conception weeks (pcw), Spearman r's > 0.80 for 2-month; > 0.78 for 12-month), and that across 2-month organoids, 12-month organoids and P7 mouse cortical cells strong correlations were maintained through the second trimester of human gestation (up to 24 pcw, r's > 0.74 for 2-month; > 0.73 for 12-month; > 0.73 for mouse cortex). From 25 pcw onwards, the average concordance between human brain tissue and 2-month organoids began a greater decline (r's > 0.60) compared to that of 12-month organoids (r's > 0.63), which overall remained lower than that of mouse cortex (r's > 0.66). As a control experiment, we compared the gene expression profiles of the human homologs of Control and 3q29Del mouse liver (*9*) to the same BrainSpan data and, as expected, found that all correlations were markedly poorer than with any human or mouse cortical sample (r's < 0.56, Fig. S7). Together, these data suggest that the gene expression profiles of 2-12 month human cortical organoids and P7 mouse cortex best model the first two trimesters of human gestation.

We performed differential expression analysis in each cluster of both experiments by genotype. In both mouse and human cells, all 3q29Del transcripts were observed to be decreased

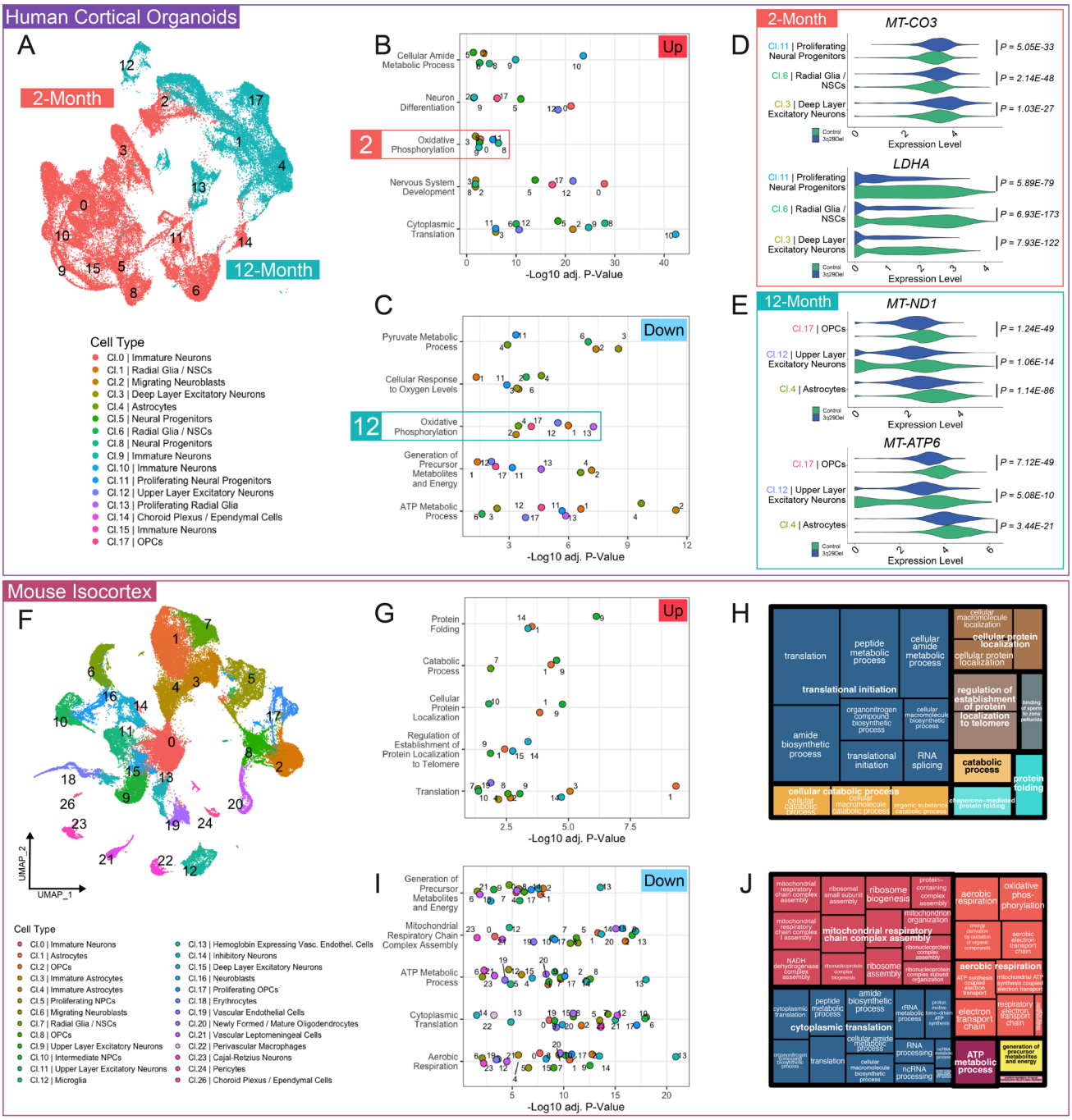


Fig 2. Transcriptomic evidence of metabolic changes in 3q29Del. The umbrella pathways most frequently found to be differentially-expressed based on up- (**B**) and down- (**C**) regulated genes in cortical organoids (**A**). Oxidative phosphorylation (OXPHOS) was enriched among both increased and decreased genes, but all clusters contributing to up-regulated OXPHOS were from 2-month organoids and all clusters contributing to down-regulated OXPHOS were from 12-month organoids. (**D**) Example violin plots visualizing log-normalized expression data of genes dysregulated in 2-month organoid clusters: *MT-CO3* (increased in 3q29Del) encodes the respiratory chain complex IV subunit COX3, *LDHA* (decreased in 3q29Del) is a key enzyme in glycolysis. (**E**) Example violin plots visualizing log-normalized expression data of genes dysregulated in 12-month organoid clusters: *MT-ND1* (decreased in 3q29Del) encodes a component of respiratory chain complex I and *MT-ATP6* (decreased in 3q29Del) encodes a component of the ATP Synthase complex. The most frequently up- (**G**) and down- (**I**) regulated umbrella pathways in mouse isocortex (**F**) are shown. Treemaps derived by Revigo analysis (**H** and **J**) display the hierarchical organization of specific Gene Ontology Biological Processes (GO:BP) identified in pathway analysis. Similar colors denote semantic similarity. The size of each rectangle is proportional to the number of clusters exhibiting over-representation of a given GO:BP term. (All p-values are adjusted for multiple comparisons). Abbreviations: OPC, oligodendrocyte progenitor cells; NSC, neural stem cells; cl, cluster.

to approximately match copy number in nearly every cell-type (Fig. S10-S12). The only mouse-specific gene located in the syntenic locus (*Bex6*) was either not detected or not differentially expressed in any cluster. Across all clusters and two time points, there were 5,244 unique DEGs in human organoids, and 3,482 DEGs across all mouse clusters. To test for a similar global impact of the 3q29Del, all unique DEGs were compiled for human and mouse experiments. Strikingly, we found that more than half of the 3,253 strictly-matched human homologs of mouse DEGs were also found to be DEGs in the human dataset (Fig. S13A, fold enrichment = 1.53, hypergeometric P = 1.64E-162).

We explored whether the total number of DEGs in a given cluster was determined by cluster size (i.e. number of cells assigned to a cluster) or the number of 3q29 locus DEGs found in that cluster (Fig. S13B). In both mouse and human datasets, the number of 3q29 DEGs within the cluster (but not cluster size) was found to be a significant predictor of the total number of DEGs (negative binomial regression, $P<0.0001$), suggesting that haploinsufficiency of genes in the 3q29 locus is a significant driver of total differential gene expression.

To further test the degree of similarity in differential gene expression across mouse and human experiments, we determined the average log fold change of DEGs in 10 comparable clusters and calculated Pearson correlations (Fig. S13C-L). From proliferating neural progenitors to deep layer excitatory neurons, we found significant positive correlations (moderate to strong) in mouse and human 3q29Del gene expression changes in all comparisons except in astrocytes (Fig. S13H), which may reflect differences in maturation state between 12-month cortical organoid cells and postnatal mouse brain.

Effects of 3q29Del on expression of mitochondrial and metabolic genes

We developed two systematic approaches to understand the most salient effects of the 3q29Del on the developing cortical transcriptome in mouse and human models. First, we sought to determine the pathways that were most frequently enriched across mouse and human clusters regardless of cell-type. To identify these frequently implicated pathways, DEGs from each cluster were split by direction of change (up-regulated vs down-regulated) and pathway analysis was performed as described in methods. All significantly enriched Gene Ontology: Biological Process (GO:BP) pathways were compiled and filtered by Revigo (*18*) to identify umbrella terms that were frequently dysregulated across multiple clusters.

We found that *Oxidative Phosphorylation* (OXPHOS) was both down-regulated and up-regulated across multiple clusters (Fig. 2B). A closer examination revealed that all down-regulated OXPHOS clusters were found in 12-month organoids and all up-regulated OXPHOS clusters were found in 2-month organoids. Moreover, the glycolysis-related *Pyruvate Metabolic Process* was found to be down-regulated in several of the 2-month clusters (cl.3, cl.11) that also up-regulated OXPHOS. Indeed, *Glycolytic Process* specifically was found to be among the top-10 down-regulated pathways in 2-month clusters of 3q29Del deep layer excitatory neurons (cl. 3), radial glia / NSCs (cl.6), and proliferating neural progenitors (cl.11, Fig. S14). This observation inspired the hypothesis that 3q29Del cells may exhibit altered metabolic maturation. A critical stage of neuronal differentiation is the switch from the heavily glycolytic progenitor state to mitochondrial aerobic respiration in mature neurons, which involves down-regulation of several key genes including *LDHA* (*19*), which encodes the enzyme lactate dehydrogenase A. Cluster-level analysis revealed a striking decrease in the expression of *LDHA* in 3q29Del early-born deep layer excitatory neurons (cl. 3) and radial glia/neural progenitors (cl. 6, cl. 11), which also showed increased mitochondrial gene expression (Fig. 2D). This result indicates a possible alteration in neuronal metabolic transition. Notably, all 12-month clusters displayed down-regulated OXPHOS

(Fig. 2C, E), which may indicate a long-term effect of 3q29Del on cellular aerobic respiration across multiple neural cell-types.

Differential pathway analysis of the mouse cortical data further supported the notion of a long-term decrease in mitochondria-related gene expression. Of the top 5 most frequently down-regulated umbrella terms, 4 were related to mitochondrial function and energy metabolism (Fig. 2I-J). Most mouse clusters were found to have down-regulated genes enriched for at least one of *Aerobic Respiration, ATP Metabolic Process, Mitochondrial Respiratory Chain Complex Assembly,* and *Generation of Precursor Metabolites and Energy* (Fig. S15).

Cross-species analysis in astrocytes and neurons

A second, parallel analysis strategy that we employed was to stringently filter exact-match DEGs in homologous human and mouse clusters by direction of change. First, we identified the human homologs of mouse DEGs in astrocytes (mm cl.1) and compared them to human DEGs in organoid cl. 4 (Fig. 3A). We found a 2.02-fold enrichment of commonly down-regulated genes (hypergeometric P = 2.8E-4) but no significant overlap among up-regulated DEGs (Fig. 3B).

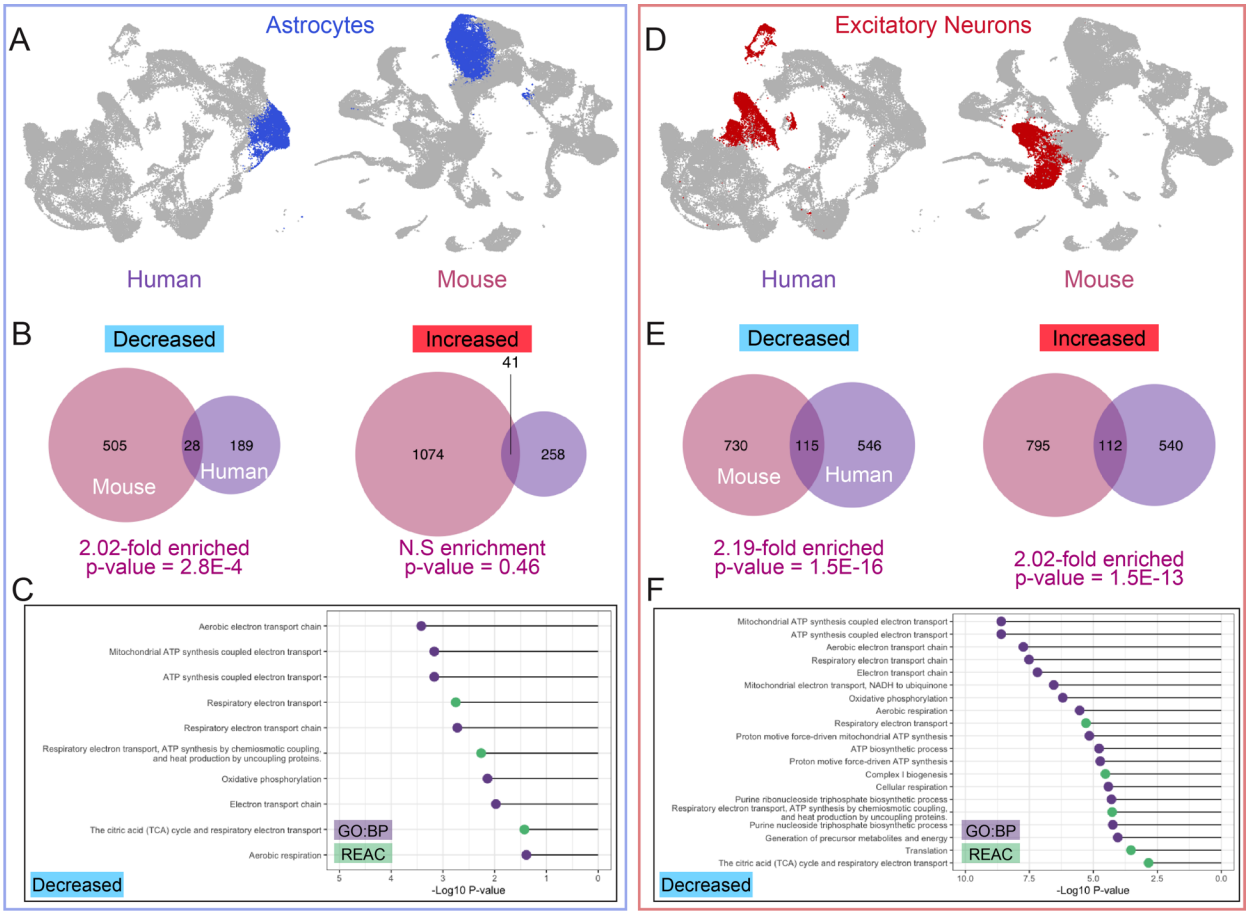


Fig 3. Common patterns of differential gene expression in two major mouse and human cell-types. Astrocytes were identified in human cortical organoids (12-month) and mouse isocortex. Corresponding clusters are color coded in blue in UMAP projections (**A**). The human homologs of mouse DEGs identified by MAST analysis were compared to organoid DEGs based on direction of change and a significant overlap was observed between the down-regulated DEGs of mouse and organoid astrocyte clusters (**B**). Pathway analysis of overlapping DEGs showed that all significantly enriched Gene Ontology: Biological Process (GO:BP) and Reactome (REAC) terms were related to mitochondrial function and metabolism (**C**). Upper and deep layer excitatory neuron DEGs were pooled and unique organoid DEGs were compared to the human homologs of mouse DEGs based on direction of change. Corresponding clusters are color coded in red in UMAP projections (**D**). There was a significant overlap between the DEGs of mouse and organoid excitatory neuron clusters for both up-regulated and down-regulated genes (**E**). Decreased genes were heavily enriched for GO:BP and REAC terms related to mitochondrial function and cellular respiration (**F**).

Pathway analysis revealed that the 28 commonly down-regulated genes were heavily enriched for terms related to the electron transport chain and OXPHOS (Fig. 3C).

We performed a parallel analysis in excitatory neurons pooling the unique DEGs of organoid cluster 3 (deep layer) and 12 (upper layer) and compared this list to the human

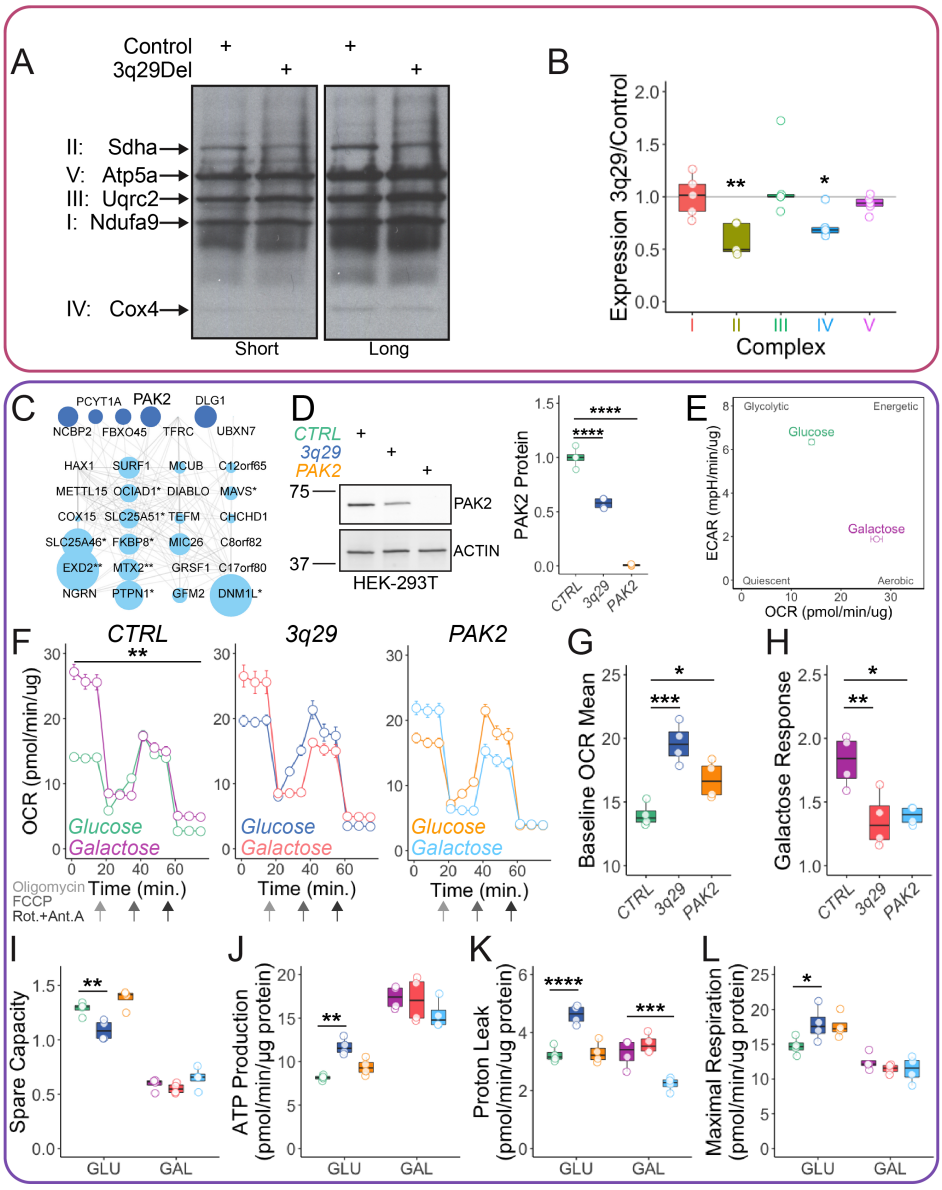


Fig 4. Mitochondrial phenotypes in 3q29 mice and engineered cell lines. Mitochondrial fractions from adult mouse brain lysates were found to have selective decreases in components of OXPHOS Complexes II and IV (**A**, quantified in **B**, N=5). At least 7 3q29-encoded proteins interact with mitochondria-localized proteins (**C**, from (*1*)). Symbol size reflects topological coefficients. HEK cell lines were engineered to carry either the heterozygous 3q29Del or completely lack PAK2 as shown by Western blot (**D**, one-way ANOVA, F(2, 9)=237.7, *CTRL vs 3q29* or *PAK2* ****P<0.0001). Control HEK-293T cells (*CTRL*) transition from a glycolytic to more aerobic cellular respiration state in galactose medium (**E**). Oxygen consumption rate (OCR) is significantly increased by 48-hour galactose medium challenge in *CTRL* cells (**F**, two-way ANOVA, main effect of medium F(1, 6)=23.99, **P=0.0027) but not in *3q29* (two-way ANOVA, F(1, 6)=0.08808, P=0.7766) or *PAK2* cells (two-way ANOVA, F(1, 6)=0.6221, P=0.4603). Both *3q29* and *PAK2* cells displayed increased baseline OCR (**G**, one-way ANOVA, effect of genotype F(2, 9)=17.24, P=0.0008; *CTRL vs. 3q29* ***P=0.0005, **CTRL vs PAK2* P=0.0332) and decreased response to galactose (**H**, one-way ANOVA, effect of genotype F(2, 9)=8.838, P=0.0075; *CTRL vs. 3q29* **P=0.0075, *CTRL vs PAK2* *P=0.0138). In glucose medium, *3q29* cells showed reduced spare capacity (**I**, two-way ANOVA effect of genotype, F(2, 18)=448.0, P=0.0003; *CTRL vs 3q29* **P=0.0047) and increased ATP production (**J**, two-way ANOVA effect of genotype, F(2, 18)=4.309, P=0.0296; *CTRL vs 3q29* **P=0.0079). Proton leak (**K**) was found to be increased in *3q29* cells in glucose (two-way ANOVA, main effect of genotype F(2, 18)=31.16, P<0.0001; *CTRL vs 3q29* ****P<0.0001) and decreased in *PAK2* cells in galactose (*CTRL vs PAK2* ***P=0.0007). Maximal respiration was significantly elevated in *3q29* cells in glucose (**L**, two-way ANOVA interaction of genotype and medium, F(2, 18)=4.219, P=0.0314; *CTRL vs 3q29* *P=0.0364) but was unchanged from *CTRL* in galactose conditions.

homologs of mouse DEGs from excitatory neuron clusters 9, 11, and 15 (Fig. 3D). In this case, we found approximately twice the overlap that would be expected by chance among both increased and decreased genes (P = 1.5E-16). Nearly all significantly enriched GO:BP and Reactome (REAC) terms among down-regulated genes are related to cellular energy metabolism and mitochondrial function (Fig. 3F, top 20 enriched pathways by p-value shown). Statistically significant intersections were also observed across human and mouse oligodendrocyte progenitor cells (OPCs), radial glia, and migrating neuroblasts (Fig. S16).

Components of respiratory complexes II and IV are altered in mouse brain

The single-cell transcriptomic data from human 12-month organoids and mouse isocortex strongly indicated a long-term effect of the 3q29Del on mitochondrial function. More specifically, a top pathway dysregulated across 17 mouse clusters was *Mitochondrial Respiratory Chain Complex Assembly*. To test the hypothesis that the 3q29Del compromises the integrity of the mitochondrial respiratory chain at the protein level, we probed by Western blot for OXPHOS complex components in Percoll-isolated mitochondrial fractions from adult male and female mouse brains in control and 3q29Del backgrounds. Across 5 independent experiments with 2 mice per genotype pooled in each replicate, we observed selective decreases in components of Complexes II and IV (Fig. 4A-B, one sample Wilcoxon signed-rank test, $P<0.05$) indicating a shift in the stoichiometry of respiratory chain complexes.

Mitochondrial function is altered in 3q29Del engineered cell lines

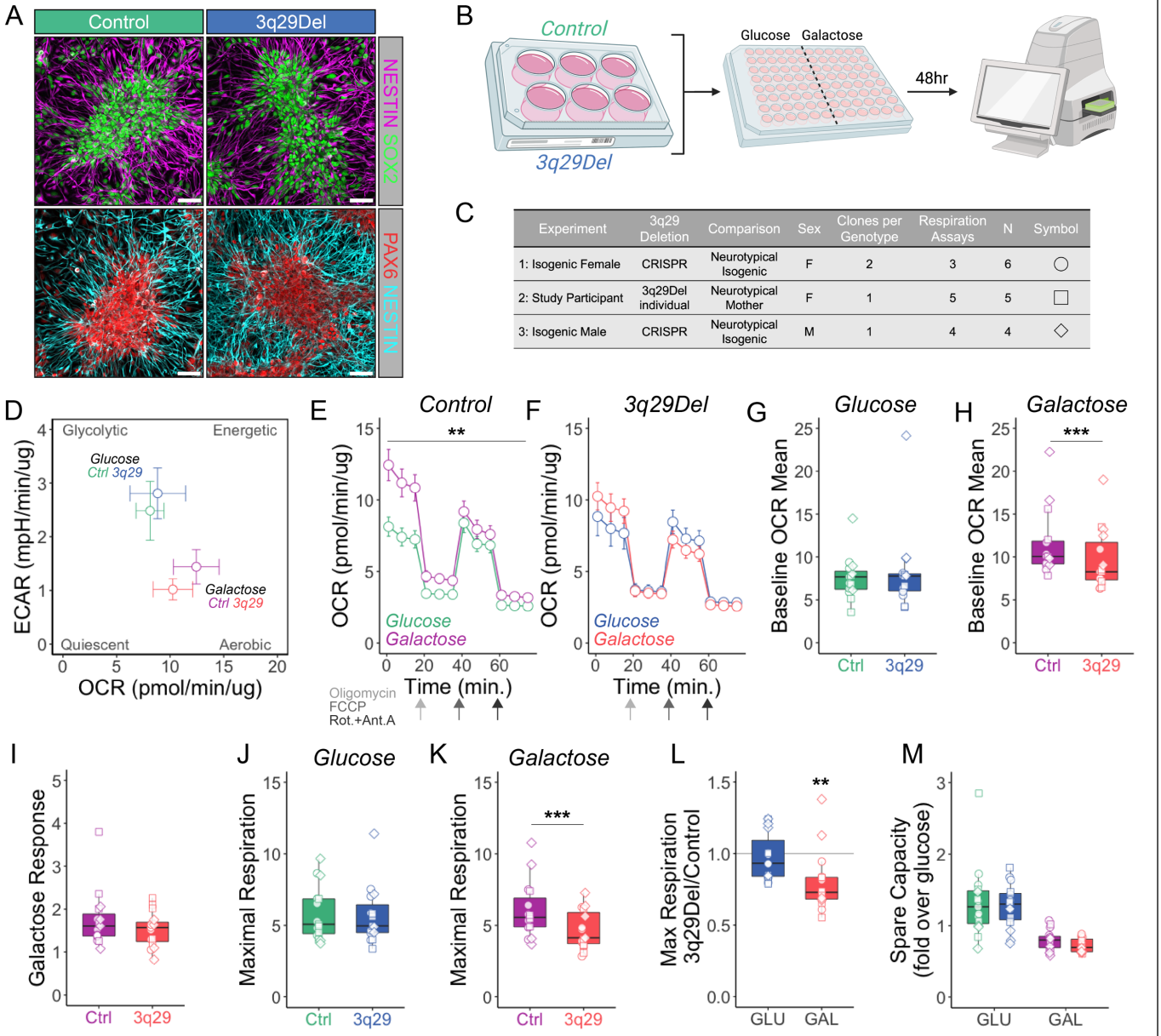
Transcriptomic data from human 2-month 3q29Del organoids suggested an alteration in the timing and/or efficiency of the canonical glycolysis to OXPHOS shift in developing neural cells. We sought to mimic this shift by challenging engineered HEK-293T cell lines with galactose medium (GAL), which forces cells to utilize OXPHOS for energy production (*20-22*). Indeed, 48-hr galactose treatment dramatically shifted control HEK cells (*CTRL)* from glycolytic to aerobic metabolism (Fig. 4E). For these experiments, we used similar CRISPR/Cas9 methods to engineer a complete version of the hemizygous 3q29Del into HEK-293T cells (*TFRC-BDH1*). In a separate HEK line, we ablated expression of the 3q29 gene *PAK2* (Fig. 4D, Fig. S17), which was found to be among the most highly-expressed genes in the 3q29Del interval in nearly all mouse and human cell-types and was recently shown to be involved in cellular energy metabolism (*23*). Indeed, we found that PAK2 is one of seven 3q29-encoded proteins identified in a proximity labeling protein-protein interaction mitochondria map (*1*).

After a 48-hr galactose challenge, we measured mitochondrial function in the Seahorse mitochondrial stress test, which isolates contributions of the respiratory chain complexes to oxygen consumption in cultured cells through sequential addition of complex-specific inhibitor molecules. Galactose medium significantly increased oxygen consumption rate in control (*CTRL*) cells (Fig. 4F) but not in *3q29* or *PAK2* cells, consistent with the disrupted stoichiometry of OXPHOS protein complexes identified in 3q29 mouse brain (Fig. 4). Both *3q29* and *PAK2* cells displayed increased baseline OCR (Fig. 4G) and decreased response to galactose (Fig. 4H). Not all 3q29 phenotypes were recapitulated by *PAK2* knockout. For example, *3q29* cells showed reduced spare capacity (Fig. 4I) and increased ATP production-dependent on respiration (Fig. 4J) in glucose medium, which were unaffected by loss of PAK2. Proton leak (Fig. 4K) was found to be increased in *3q29* cells in glucose and decreased in *PAK2* cells in galactose. Maximal respiration (Fig. 4L) was significantly elevated in *3q29* cells in glucose but was unchanged from *CTRL* in galactose conditions. Thus, *PAK2* is likely contributing to, but not solely sufficient for, the metabolic phenotypes of 3q29Del.

3q29Del neural progenitors lack metabolic flexibility

To directly test the hypothesis that the 3q29Del disrupts the metabolic transition from the glycolytic state of neural progenitor cells (NPCs) to aerobic respiration in differentiated neurons, we challenged Control and 3q29Del NPCs with galactose medium for 48hrs prior to Seahorse mitochondrial stress assays. These assays were performed in 8 clones of 6 independent cell lines differentiated from both isogenic cell lines and lines derived from 3q29Del study participants, over 12 respiration assays (Fig. 5C, Supp. File 1). As we observed in HEK cell lines, galactose medium shifted NPCs toward a more aerobic metabolic profile (Fig. 5D), and while Control NPCs significantly increased oxygen consumption rate (OCR) in response to galactose challenge, 3q29Del cells did not (Fig. 5E). Unlike HEK cells, we found no differences in OCR under glucose conditions, but found that 3q29Del NPCs had a lower baseline OCR mean (Fig. 5H) and decreased maximal respiration (Fig. 5K) in galactose medium. These data demonstrate that like HEK-293T cells, 3q29Del NPCs are impaired in the transition from glycolysis to oxidative metabolism induced by a galactose challenge. Importantly, these changes were observed in patient-derived lines as well as in our lab-generated isogenic lines.

Taken together, these experiments reveal that the 3q29Del produces a convergent mitochondrial phenotype at the level of transcriptome, protein expression and function.



Experiment	3q29 Deletion	Comparison	Sex	Clones per Genotype	Respiration Assays	N	Symbol
1: Isogenic Female	CRISPR	Neurotypical Isogenic	F	2	3	6	○
2: Study Participant	3q29Del individual	Neurotypical Mother	F	1	5	5	□
3: Isogenic Male	CRISPR	Neurotypical Isogenic	M	1	4	4	◇

Figure 5. Lack of metabolic flexibility in 3q29Del neural progenitor cells. A) Control and 3q29Del neural progenitor cells (NPCs) exhibited normal morphology and stained positive for the neurofilament protein Nestin, multipotency marker SOX2, and NPC marker PAX6 (scale = 50um, quantified in Fig. S18). **B**) Illustration of experimental design. NPCs were challenged for 48-hr in neural medium containing glucose (GLU) or galactose (GAL). **C**) Table of cell lines used in this experiment. Data from three separate cohorts was combined in plots D-M. N=15 from 6 independent NPC lines for all experiments. **D**) Energy map indicates that galactose treatment pushes cells from more glycolytic to more aerobic metabolic profile. **E**) Control NPCs significantly increase oxygen consumption rate (OCR) in galactose medium (two-way ANOVA main effect of medium, $F(1, 28)=9.295$, **P=0.0050). **F**) 3q29Del NPCs exhibited no significant change in OCR in galactose medium (two-way ANOVA, medium effect $F(1, 28)=0.01219$, P=0.9129). **G**) No significant difference in baseline OCR mean in glucose medium was observed (two-tailed ratio paired t-test, P=0.7015), but 3q29 NPCs displayed significantly lower baseline OCR mean in galactose medium (**H**, two-tailed ratio paired t-test, ***P=0.0009). **I**) Galactose response (i.e. basal OCR fold change over glucose) was unchanged in 3q29Del NPCs (two-tailed ratio paired t-test, P=0.0935). **J**) Maximal respiration was unchanged in glucose medium (two-tailed ratio paired t-test, P=0.5028), but was significantly decreased in 3q29Del NPCs in galactose medium (**K**, two-tailed ratio paired t-test, ***P=0.0007). Similarly, (**L**) the maximal respiration ratio of 3q29Del:Control NPCs was unchanged in glucose medium (GLU) but was significantly reduced in galactose conditions (GAL, one sample two-tailed t-test, **P=0.0026). There was no significant change in spare capacity in 3q29Del NPCs (**M**, two-way ANOVA genotype effect $F(1, 56)=0.5930$, P=0.4445).

Discussion

As the strongest known genetic risk factor for SCZ, the 3q29Del is a high priority target for mechanistic investigation. This study employed a rigorous cross-species strategy to identify transcriptomic phenotypes in human and mouse 3q29Del neural tissue. By leveraging the complementary advantages of two highly relevant experimental model systems, we have isolated the effects of the 3q29Del in the developing cortex. We further designed an analysis strategy to systematically filter widespread DEG findings down to the most likely processes and pathways for in-depth functional analysis. This approach led to neural mitochondria as a site of consistent gene dysregulation. Subsequent testing revealed changes in mitochondrial protein expression and function at the cellular level, consistently observed in multiple cell types including study participant-derived cell lines. These data strongly implicate the mitochondrion as an organelle impacted by the 3q29Del. These results also highlight the strength of our cross-species approach, which rapidly and efficiently led to productive avenues for functional study.

In human cortical organoids, we found that *Oxidative Phosphorylation* was among the most frequently implicated pathways across cell-type clusters (Fig. 2). However, the direction of change was dependent on developmental timepoint: it was up-regulated in 2-month cells and decreased in 12-month cells. Moreover, the key glycolysis gene *LDHA* was found to be strongly decreased in multiple 2-month clusters that were overexpressing OXPHOS genes. These findings suggest that the glycolysis to OXPHOS transition, which is critical for neuronal differentiation and maturation (*19*), may be disrupted in 3q29Del cells. We tested this prediction in an independent cellular model system, 3q29-engineered HEK cells, and found further evidence for a lack of metabolic flexibility; 3q29Del cells showed almost no spare capacity under baseline conditions and, when challenged with a galactose based medium to force aerobic respiration via OXPHOS, 3q29Del cells had a notably blunted response (Fig. 4). In addition, ablation of the 3q29-encoded gene *PAK2* recapitulated two of these effects – decreased response to galactose medium and increased baseline aerobic respiration – while leaving spare capacity intact. These results suggest that *PAK2* is likely one of multiple 3q29 locus-encoded genes that contribute~~s~~ to metabolic phenotypes. We further tested the hypothesis of an altered metabolic transition in 3q29Del cells by challenging neural progenitor cultures from two independent isogenic pairs (male and female) and a pair of study participant derived cell lines in a similar galactose medium paradigm. We found that 3q29Del neural progenitors did not significantly increase oxygen consumption in galactose medium and had reduced maximal respiration compared to Control cells. These data further indicate that the 3q29Del imparts a vulnerability to metabolic challenges. Further studies will be required to dissect the contributions of 3q29Del locus genes to this phenotype.

A major strength of the design of the current study is the utilization of human and mouse models as well as two time points *in vitro*. In both 12-month human cortical organoids and perinatal mouse cortical tissue, we found a widespread decrease in expression of genes related to mitochondrial energy production. In particular, down-regulated gene lists from 17 mouse clusters were enriched for *Mitochondrial Respiratory Chain Complex Assembly* (Fig. 2). In support of this transcriptomic prediction, we found evidence for a shift in the stoichiometry of respiratory chain complex proteins in mitochondrial fractions from mouse brain with specific decreases in protein components of complex II and IV (Fig. 4). This result indicates that the transcriptomic changes that we observed in perinatal mouse cortical tissue and 12-month *in vitro* human cortical organoids may translate into long-term, persistent effects at the protein level. Future work may be aimed determining if there is a mechanistic connection between the cellular energy metabolism phenotypes that we have described and the consistent finding that mouse and human individuals with 3q29Del are significantly smaller than expected (*7, 8, 10*).

Mitochondria have been previously implicated in the pathophysiology of neurodevelopmental CNV disorders and idiopathic schizophrenia (*24-27*). Interestingly, a CNV disorder with perhaps the most similar phenotypic profile to 3q29Del in human carriers, 22q11.2 deletion (22q11.2Del), harbors at least eight genes that encode mitochondria-linked proteins, several of which are also enriched at synapses (*28*): *MRPL40, SLC25A1, PRODH, TXNRD2, AIFM3, COMT, RTL10, SNAP29* (*29*). Several mitochondrial phenotypes have now been reported in 22q11.2Del models. Similar to our findings, the activity of OXPHOS complexes I and IV was found to be decreased in human 22q11.2Del iPSC-derived neurons, which resulted in reduced ATP production (*11, 12*). This suggests a convergent biology of mitochondrial dysfunction between 3q29Del and 22q11 deletion. This phenotype was attributed to haploinsufficiency of the 22q11.2Del locus gene *MRPL40*, which is a component of mitochondrial ribosome. Interestingly, loss of one copy of *Mrpl40* in mice is sufficient to produce short-term neuroplasticity phenotypes (*30*), potentially linking mitochondrial phenotypes to more well-established synaptic defects in SCZ models. A separate study that utilized a cross-species strategy to prioritize 22q11.2Del-associated effects in mouse brain and human patient fibroblasts identified the 22q11.2 gene encoding the mitochondrial citrate transporter *SLC25A1* as a key component of a dysregulated mitochondrial protein hub (*31*). Further studies indicated an interaction between SLC25A1 and MRPL40 at the protein level (*32*). Additionally, a large transcriptomic study of 22q11.2Del cortical organoids also found enrichment of DEGs related to mitochondrial function (*33*).

Other neurodevelopmental CNVs have been associated with mitochondrial phenotypes as well. The Williams Syndrome (deletion) and SCZ-associated (duplication) locus 7q11.23 contains the gene *DNAJC30*, which encodes a protein that interacts with the ATP synthase complex (*34*). Complete loss of *DNAJC30* was found to disrupt sociability in mice and severely impair mitochondrial function in mouse neurons (*34*). Human fibroblasts from individuals with Williams Syndrome (i.e. hemizygous for *DNAJC30*) were also found to have impaired mitochondrial function and reduced ATP production (*34*). In addition, a recent study of reciprocal CNVs at the neurodevelopmental disorder associated locus 16p11.2 also found a strong signal for enrichment of DEGs related to energy metabolism and mitochondrial function in both mouse brain and cultured human neural cell lines (*35*) and loss of the 16p11.2 locus-encoded gene *TAOK2* was found to disrupt mitochondrial morphology and function in mouse neurons (*36*). Another study of convergent biology in mouse models of the neurodevelopmental CNVs 1q21.1, 15q13.3, and 22q11.2 found dysregulation of a transcriptomic module related to neuronal energetics (*37*). Finally, recent data indicates that complete loss of one of the top single gene risk factors for SCZ, *SETD1A* (*38*), impairs basal glycolysis and respiratory capacity in human neurons (*39*). Together, these data indicate that our findings in 3q29Del mouse and human cortical tissue fit with reports of other mitochondrial phenotypes associated with neurodevelopmental variants and suggest that neural mitochondria may be a key site of biological convergence downstream of these high-risk alleles.

Unlike 22q11.2, which encodes several proteins that function within mitochondria, the mechanistic link to 3q29 genes is not known. Our data implicates the highly expressed kinase *PAK2* in the metabolic phenotypes associated with 3q29Del, though likely in conjunction with additional driver genes, as predicted by our earlier network-based inferences on 3q29 neuropathology emerging upon loss of multiple functionally connected genes in the interval (*40*). Mitochondria are involved in many cellular pathways and processes in addition to energy production including apoptosis signaling. Thus, our findings of transcriptomic and functional phenotypes at 3q29Del mitochondria fit with a previous report of increased susceptibility to apoptosis in *Drosophila* models (*41*). Further studies will be required to determine if mitochondrial phenotypes are a primary consequence of 3q29Del, and the specific driver genes for these effects.

Given the hierarchical structure of scRNA-seq data, treating individual cells as independent sampling units can yield false positives in differential expression results due to the underestimation of true standard errors. To improve the reproducibility and validity of our findings, we focused our investigation on disrupted gene expression signals and corresponding signaling pathways with independent statistical support from two separate model systems. We note that while this approach increases our confidence in capturing true associations, additional signals relevant to disease mechanism may be hidden among unshared findings between model systems. We also note that non-coding genes make up a small fraction of the detected transcripts in our present library construction approach, hence, the extent to which non-coding genes of the 3q29Del interval contribute to pathogenesis is unknown; more work is needed to decode the involvement of the non-coding elements of this region in future studies.

Bridging the gap between genetic risk and biological mechanisms is a major challenge for psychiatry. In this study, we sought to use systematic methods to identify the most salient, conserved transcriptomic effects of the SCZ-associated 3q29Del in disease-relevant tissues as an important step toward determining cellular and molecular phenotypes of this important variant. These findings should motivate further work to determine the mechanisms of these 3q29Del sequelae and their relevance to various clinical phenotypes.

Materials and Methods

Cell Culture and Genome Engineering

Whole blood samples of 5-10 mL were collected in EDTA Vacutainer test tubes and processed for the isolation of erythroid progenitor cells (EPCs) using the Erythroid Progenitor Kit (StemCell Technologies). EPCs were reprogrammed using Sendai particles (CytoTune-iPS 2.0 Reprogramming kit, Invitrogen) and plated onto Matrigel coated six-well plates (Corning). Cultures were transitioned from erythroid expansion media to ReproTesR (StemCell Technologies) and then fed daily with ReproTesR until clones were isolated. iPSCs were maintained on Matrigel coated tissue culture plates with mTeSR Plus (StemCell Technologies).

Cell lines were characterized for stem cell markers by RT-PCR and immunocytochemistry after at least 10 passages in culture. Total RNA was isolated from each cell line with the RNeasy Plus Kit (Qiagen) according to manufacturer's protocol. mRNA was reverse transcribed into cDNA using the High Capacity cDNA Synthesis Kit (Applied Biosystems). Expression of pluripotency genes *OCT4*, *SOX2*, *REX1* and *NANOG* was determined by RT-PCR. Sendai virus inactivity was confirmed using Sendai genome specific primers.

Isogenic 3q29Del iPSC and HEK cell lines were generated using the SCORE method (*13*). To identify low-copy repeat (LCR) target sequences, the reference sequence (hg38) between *TNK2 – TFRC* (centromeric) and *BDH1 – RUBCN* (telomeric) was downloaded and aligned in NCBI BLAST. A ~20 Kb segment was found to be 97% identical and was searched for gRNA sequences using CHOPCHOP (https://chopchop.cbu.uib.no) (*16*). Three single gRNA sequences (IDT) that were predicted to each cut at a single site in both LCRs were identified and cloned into pSpCas9(BB)-2A-Puro (PX459) V2.0, which was a gift from Feng Zhang (Addgene plasmid #62988; http://n2t.net/addgene:62988; RRID: Addgene_62988) (*42*).

Single gRNA plasmids were transfected into a neurotypical control iPSC line (IRB#CR002-IRB00088012, maintained in mTeSR or mTeSR+ (STEMCELL, Vancouver) on Matrigel (Corning)-coated plates using a reverse transfection method and Mirus TransIT-LT1 reagent (Mirus Bio, Madison, WI) and transfected cells were transiently selected for puromycin resistance. Genome cleavage efficiency for each gRNA was calculated using the GeneArt Genomic Cleavage Detection Kit (Thermo) and gRNA_2 (5'-CAGTCTTGGCTACATGACAA-

3', directed to -strand, hg38 chr3:195,996,820 - chr3:197,634,397) was found to be the most efficient with cleaved bands at the predicted sizes. Cells transfected with gRNA_2 were dissociated and cloned out by limiting dilution in mTeSR supplemented with 10% CloneR (STEMCELL). Putative clonal colonies were manually transferred to Matrigel-coated 24-well plates for expansion and screened for change in copy number of the 3q29Del locus gene *PAK2* (Hs03456434_cn) using TaqMan Copy Number Assays (Thermo). Three (of 100) clones showed an apparent loss of one copy of *PAK2* and were subsequently screened for loss of the 3q29 genes *TFRC* (Hs03499383_cn), *DLG1* (Hs04250494_cn), and *BDH1* (Hs03458594_cn) and for no change in copy number to external (non-deleted) 3q29 genes *TNK2* (Hs03499383_cn) and *RUBCN* (Hs03499806_cn) all referenced to *RNASEP* (Thermo #4401631). All cell lines retained normal karyotypes (WiCell, Madison, WI) and were free of mycoplasma contamination (LookOut, Sigma).

To generate 3q29Del HEK-293T cell lines, HEK cells (RRID:CVCL_0063) were transfected with either empty px459 or px459+gRNA_1 (5'-ttagatgtatgccccagacg-3', directed to the +strand) and screened and verified with TaqMan copy number assays as described above. *PAK2* was deleted from a control HEK-293T line as detailed above (*PAK2* gRNA 5'-TTTCGTATGATCCGGTCGCG-3', directed to -strand). Clones were screened by Western blot (Rabbit monoclonal PAK2 from Abcam; ab76293; RRID AB_1524149; 1:5000 dilution) and confirmed by Sanger sequencing PCR-amplified gDNA. HEK cell lines were also negative for mycoplasma contamination.

Genome-wide Optical Mapping

1.5E6 iPSCs were pelleted, washed with DPBS, and frozen at -80°C following aspiration of all visible supernatant. 750ng of DNA was labeled, stained, and homogenized using the DNA Labeling Kit-DLS (Bionano; 80005). Stained DNA was loaded onto the Saphyr chip G1.2 and the chip was scanned in order to image the labeled DNA using the Saphyr System. Structural variants were called relative to the reference genome (hg38) using Bionano Solve. Structural variants were compared to the parent (unedited) cell line using the Bionano Solve Variant Annotation Pipeline.

Cortical Organoid Differentiation

Engineered isogenic 3q29Del iPSC lines and the unedited parent line, along with two additional clonal lines from the same donor, were expanded in mTeSR or mTeSR+ on Matrigel-coated plates. On DIV 0, colonies were gently released from plates in 0.35mg/ml Dispase according to an established protocol (*17*). Floating colonies were re-suspended in mTeSR supplemented with 10uM Y-27632 (Reprocell, Beltsville, MD) in ultra-low attachment 10cm dishes (Corning). After 48hr, spheroids were transitioned to Neural Induction Medium (20% Knockout Serum Replacement, 1% Non-essential amino acids, 100U/mL Pen/Strep, 0.5% GlutaMAX, 0.1mM 2-mercaptoethanol in DMEM/F12 w/ HEPES), supplemented with 5uM Dorsomorphin and 10uM SB-431542 (added fresh) with daily media changes through DIV 6. On DIV 7, Neural Induction Medium was replaced with Neural Medium (Neurobasal-A with 2% B-27 w/o vitamin A, 1% GlutaMAX, 100U/mL Pen/Strep) supplemented with fresh EGF (20ng/ml, R&D Systems) and FGF (20ng/ml, R&D Systems) for daily media changes through day 16. From day 17-25, organoids were fed Neural Medium with EGF and FGF every two days. From day 26-42, Neural Medium was supplemented with BDNF (20ng/ml, R&D Systems) and NT-3 (20ng/ml, R&D Systems) every two days. From day 43 onwards, organoids were fed Neural Medium without supplements twice weekly.

Mouse Genotyping and Maintenance

All animal experiments were performed under guidelines approved by the Emory University Institutional Animal Care and Use Committee. Mice were genotyped as described previously (*8*) and noted as either Control (wild-type, C57BL/6 N Charles River Laboratories) or 3q29Del (B6.Del16$^{+/Bdh1\text{-}Tfrc}$, MGI:6241487). Male 3q29Del mice and Control littermates were included in the scRNA-seq study. Both male and female mice were included in mitochondrial fractionation experiments.

Tissue Dissociation and Sorting

Single-cell suspensions from cortical organoids (DIV 50 = "2-month" N=2 Control, N=2 3q29Del, and DIV 360 = "12-month" N=2 Control, N=2 3q29Del) and postnatal day 7 (P7) mouse cortices (N=4 Control, N=4 3q29Del) were produced by a papain dissociation method based on a published protocol (*43*). Organoids were dissociated in three batches that were each balanced for genotype and "age". Mouse samples were also dissociated in three batches each balanced by genotype. In both sets the experimenter was blinded to genotype. Tissue was coarsely chopped with a sterile scalpel and digested for 1hr at 34°C in a pH-equilibrated papain solution (Worthington, Lakewood, NJ) with constant CO_2 flow over the enzyme solution. Digested tissue was gently spun out of papain, through ovomucoid solutions, and sequentially triturated with P1000 and P200 pipet tips. Live cells were counted by manual and automated methods (Countess II, Thermo) and in organoid samples were isolated from cellular debris by fluorescence-activated cell sorting on a FACSAria-II instrument (calcein AM-high, Ethidium Homodimer-1 low).

Single-cell Library Prep and RNA-Sequencing

Single-cell suspensions were loaded into the 10X Genomics Controller chip for the Chromium Next GEM Single Cell 3' kit workflow as instructed by the manufacturer with a goal capture of 10,000 cells per sample. The resulting 10X libraries were sequenced using Illumina chemistry. Mouse samples and libraries were prepared and sequenced at a separate time from human samples.

scRNA-seq Data Processing and Analysis

To quantify gene expression at single-cell resolution, the standard Cell Ranger (10x Genomics) and Seurat (*44*) data processing pipelines were followed for demultiplexing base call files into FASTQ files, alignment of scRNA-seq reads to species-specific reference transcriptomes with STAR (mouse: mm10, human: GRCh38), cellular barcode and unique molecular identifier (UMI) counting, and gene- and cell-level quality control (QC). To filter out low-quality cells, empty droplets and multiplets, genes expressed in <10 cells, cells with >30% reads mapping to the mitochondrial genome, and cells with unique feature (gene) counts >7,000 were removed based on manual inspection of the distributions of each QC metric individually and jointly. Outlier cells with low unique feature counts were further removed via sample-specific thresholding of corresponding distributions (<250 for mice; <700 for organoids). Thresholds were set as permissive as possible to avoid filtering out viable cell populations, consistent with current best-practice recommendations (*45*).

The *sctransform* function in Seurat was used for normalization and variance stabilization of raw UMI counts based on regularized negative binomial regression models of the count by cellular sequencing depth relationship for each gene, while controlling for mitochondrial mapping percentage as a confounding source of variation (*46*). Resulting Pearson's residuals were used to

identify the most variable features in each dataset (n=3,000 by default), followed by dimensionality reduction by PCA and UMAP, shared nearest neighbor (SNN) graph construction on the basis of the Euclidean distance between cells in principal component space, and unbiased clustering of cells by Louvain modularity optimization. Optimal clustering solutions for each dataset was determined by building cluster trees and evaluating the SC3 stability index for every cluster iteratively at ten different clustering resolutions with the *clustree* function in R (*47*). The effect of cell-cycle variation on clustering was examined by calculating and regressing out cell-cycle phase scores in a second iteration of *sctransform*, based on the expression of canonical G2/M and S phase markers (*48*). Consistent with the developmental context of the interrogated datasets, cell-cycle differences were found to covary with cell-type and retained in final analyses as biologically relevant sources of heterogeneity. Cluster compositions were checked to confirm comparable distributions of experimental batch, replicate ID, and genotype metadata. Cluster annotations for cell-type were determined based on the expression of known cell-type and cortical layer markers curated from the literature (*49-53*). Clusters exhibiting cell-type ambiguity were further sub-clustered to refine annotations or dropped from downstream analysis in case of inconclusive results (human cl.7 and cl.16; mouse cl. 25 and cl. 27).

Seahorse Mitochondrial Stress Assay

HEK-293T cells that had been engineered to carry the 3q29Del (*3q29Del*), PAK2 knockout (*PAK2*), and mock-edited control cells (*CTRL*) were plated on poly-D-lysine coated 96-well Seahorse assay plates (XF96, Agilent) in DMEM (Gibco A144300) supplemented with 10% FBS, 2mM L-glutamine, 1mM sodium pyruvate, and either 10mM D-(+)-glucose ("Glu", 7.5E3 cells/well) or 10mM galactose ("Gal",15E3 cells/well). After 48hr, cells were washed twice in XF DMEM Assay Medium (Agilent) with either glucose or galactose (10mM) supplemented with 1mM pyruvate, 2mM glutamine.

Neural progenitor cells (NPCs) were plated at 5E4 cells/well in poly-L-ornithine (Sigma P4957; 15ug/mL) and laminin (Sigma 23017-015; 5ug/uL) coated 96-well Seahorse assay plates in STEMdiff Neural Progenitor Medium (STEMCELL 05834). After 24hr, all media was aspirated and exchanged for Neural Medium (Neurobasal minus glucose, glutamine, pyruvate Thermo A2477501) supplemented with B-27 minus insulin (Thermo A1895601), 2mM L-glutamine, 1mM sodium pyruvate with either 17.5mM glucose or galactose for 48hrs.

Mitochondrial stress test compounds were loaded into injection ports as indicated by the manufacturer to achieve the following final concentrations for HEK cells: 1uM oligomycin, 0.25uM FCCP, 0.5uM rotenone, 0.5uM antimycin A (all sourced from Sigma). NPC assays were performed with 2uM oligomycin, 0.5uM FCCP, 1uM rotenone, and 1um antimycin A. Cells equilibrated at 37°C with ambient CO_2 for approximately 1hr prior to assay initiation. At the end of the experiment, cells were washed twice in PBS+Ca^{2+}+Mg^{2+} and lysed at 4C for 30 min in 0.5% Triton X-100 protein buffer (150mM NaCl, 10mM HEPES, 0.1mM $MgCl_2$, 1mM EGTA, 1x HALT Protease + Phosphatase inhibitor). Protein concentrations in each well were determined by BCA (Pierce) to normalize oxygen consumption rate data. Data were analyzed in Wave (Agilent). Assay wells that did not show responses to drug injections were excluded from analysis.

Mouse Brain Mitochondrial Isolation

A protocol for mitochondrial isolation was adapted from prior work (*54*). Two whole brains per genotype were dissected from adult mice (2-6mos.) and pooled in 2.5mL of ice cold Medium I (0.32M sucrose, 5mM HEPES pH 7.5, 0.1mM EDTA, Complete protease inhibitor)

and homogenized with 16 strokes at approx. 800rpm in a Teflon glass homogenizer (0.125mm clearance) with a rest on ice mid-way through. Crude homogenate was cleared by centrifugation at 1000 x *g* for 10min and the supernatant was further centrifuged at 12,000 x *g* for 20min. All centrifugations were carried out at 4°C.

Isoosmotic Percoll (9 parts Percoll to 1 part 2.5M sucrose vol/vol) gradients were prepared in Medium II (0.25M sucrose, 5mM HEPES pH 7.2, 0.1mM EDTA). The second pellet was carefully re-suspended in an appropriate volume of 8.5% Percoll to produce a 7.5% Percoll solution and then was gently homogenized by twisting the Teflon pestle through the solution. The 7.5% Percoll solution containing the re-suspended tissue fraction was carefully layered on top of a gradient containing 16% and 10% Percoll. Gradients were centrifuged for 20min at 15,000 x *g* and mitochondrial fractions were extracted from the bottom of the tube and solubilized in 0.5% Triton X-100 protein buffer (150mM NaCl, 10mM HEPES, 0.1mM $MgCl_2$, 1mM EGTA, 1x Complete). Protein concentrations were determined by BCA (Pierce) and normalized. 20ug of protein was loaded to each lane of Criterion gels for SDS-PAGE. Gels were transferred onto PVDF membranes by standard protocols and blocked in 5% milk. OXPHOS complex component proteins were probed for with an OXPHOS antibody cocktail (1:250, Abcam ab110412). Protein levels were determined by band densitometry and quantified by normalizing to the most stable complex component (V).

Statistical Analysis

Differential gene expression testing for genotype was performed on log normalized expression values (scale.factor=10,000) of each cluster separately with a two-part generalized linear model that parameterizes the bimodal expression distribution and stochastic dropout characteristic of scRNA-seq data, using the MAST algorithm, while controlling for cellular detection rate (*55*). A threshold of 0.1 was implemented as the minimum cut off for average log-fold change (logfc.threshold) and detection rates (min.pct) of each gene in either genotype to increase the stringency of differential expression analysis. Multiple hypothesis testing correction was applied conservatively using the Bonferroni method to reduce the likelihood of type 1 errors, based on the total number of genes in the dataset. To facilitate comparative transcriptomics, human homologs (including multiple paralogs) were identified for all differentially-expressed genes (DEGs) in the mouse dataset via the NCBI's HomoloGene database (ncbi.nlm.nih.gov/homologene/). Data processing and analysis pipelines were harmonized across the mouse and organoid datasets, yielding parallel computational approaches for cross-species comparison of differential expression signals. The BrainSpan Developmental Transcriptome dataset used for developmental stage estimations was obtained by bulk RNA-Sequencing of postmortem human brain specimens collected from donors with no known history of neurological or psychiatric disorders, as described previously (*2, 56*). This large-scale resource is accessible via the Allen Brain Atlas data portal (https://www.brainspan.org/static/download/, file name: "RNA-Seq Gencode v10 summarized to genes"); dbGaP accession number: phs000755.v2.p1. All statistical analyses of scRNA-seq data were performed in R (v.4.0.3).

To interpret differential gene expression results, pathways likely impacted by the 3q29Del were determined based on statistically over-represented gene-sets with known functions using g:Profiler (*57*). DEGs (Bonferroni adj. $p<0.05$) for each cluster were identified as described above and input with an experiment-specific background gene set (genes with min.pct > 0.1 in any cluster). GO:Biological Process (GO:BP) and Reactome (REAC) databases were searched with $10 <$ term size < 2000. Significantly enriched pathways below a threshold of g:SCS < 0.05 (*58*) were compiled and filtered in Revigo (*18*) to reduce redundancy and determine umbrella terms.

Western blot analysis of mouse brain OXPHOS complex components: (Fig. 4B) one sample Wilcoxon signed-rank test (two-tailed), *p<0.05, **p<0.01, N = 5.

Seahorse Mitochondrial Stress Test Analysis: (Fig. 4F) two-way ANOVA, main effect of medium F(1, 6)=23.99, **P=0.0027; (Fig. 4G) one-way ANOVA, effect of genotype F(2, 9)=17.24, P=0.0008; *CTRL vs. 3q29* ***P=0.0005, **CTRL vs PAK2* P=0.0332; (Fig. 4H) one-way ANOVA, effect of genotype F(2, 9)=8.838, P=0.0075; *CTRL vs. 3q29* **P=0.0075, *CTRL vs PAK2* *P=0.0138; (Fig. 4I) two-way ANOVA effect of genotype, F(2, 18)=448.0, P=0.0003; *CTRL vs 3q29* **P=0.0047; (Fig. 4J) two-way ANOVA effect of genotype, F(2, 18)=4.309, P=0.0296; *CTRL vs 3q29* **P=0.0079; (Fig. 4K) two-way ANOVA, main effect of genotype F(2, 18)=31.16, P<0.0001; *CTRL vs 3q29* ****P<0.0001, galactose *CTRL vs PAK2* ***P=0.0007; (Fig. 4L) two-way ANOVA interaction of genotype and medium, F(2, 18)=4.219, P=0.0314; *CTRL vs 3q29* *P=0.0364. N=4 for all HEK Seahorse experiments. (Fig. 5E) two-way ANOVA main effect of medium, F(1, 28)=9.295, **P=0.0050; (Fig. 5F) two-way ANOVA, medium effect F(1, 28)=0.01219, P=0.9129; (Fig. 5G) two-tailed ratio paired t-test, P=0.7015; (Fig. 5H) two-tailed ratio paired t-test, ***P=0.0009; (Fig. 5I) two-tailed ratio paired t-test, P=0.0935; (Fig. 5J) two-tailed ratio paired t-test, P=0.5028; (Fig. 5K) two-tailed ratio paired t-test, ***P=0.0007; (Fig. 5L) one sample two-tailed t-test, **P=0.0026; (Fig. 5M) two-way ANOVA genotype effect F(1, 56)=0.5930, P=0.4445. N=15 for all NPC Seahorse experiments according to the following table:

Table 1. Respiration assay cell lines.

Parent Cell Line	Clone	Genotype	Group	Respiration Assays
1003-0031	c27	Control	Control	3
1003-0031	c33	Control	Control	3
1003-0031	3G2b	3q29Del	Isogenic	3
1003-0031	11D7	3q29Del	Isogenic	3
4258-2096	c13	Control	Control	5
4258-1031	c11	3q29Del	Study Participant	5
4258-2046	c17	Control	Control	4
4258-2046	1C7	3q29Del	Isogenic	4

Acknowledgments

Research reported here was supported in part by Imagine, Innovate and Impact (I3) from the Emory School of Medicine, a gift from Woodruff Fund Inc., and through the Georgia CTSA NIH award (UL1-TR002378). This study was also supported in part by the Emory Integrated Genomics Core (EIGC), Emory Integrated Computational Core (EICC), Emory Integrated Cellular Imaging, Emory Flow Cytometry Core, and the Emory Stem Cell Core, which are subsidized by the Emory University School of Medicine and are part of the Emory Integrated Core Facilities. Figure illustrations generated with Biorender.com.

Funding

National Institutes of Health grant F32MH124273 (RHP)

Brain & Behavior Research Foundation (RHP)
Emory University School of Medicine University Research Committee (RHP)
National Institutes of Health grant R56MH116994 (STW and JGM)
National Institutes of Health grant R01MH110701 (GJB and JGM)
National Institutes of Health grant R01MH118534 (JGM)

Author Contributions
Conceptualization: RHP, ES, STW, SAS, GJB, JGM
Methodology: RHP, ES, EW, ATK, TJM, MEMG, PC, ZTE, NR, BJV, DS, BLF, KT, ZW, VF, SAS, GJB, JGM
Investigation: RHP, EW, TJM, SK, MIR, VF
Visualization: RHP, ES, EW, VF
Supervision: RHP, GJB, JGM
Writing—original draft: RHP, ES, JGM
Writing—review & editing: RHP, ES, EW, TJM, NR, VF, SAS, GJB, JGM

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: All data, code, and materials used in the experiments and analyses are available to interested researchers here doi.org/10.5061/dryad.b2rbnzsmm. Cell lines are available for research use upon Material Transfer Agreement. Requests for cell lines should be submitted to: RHP, GJB, and JGM. All data are available in the main text or the supplementary materials.

Figures

Fig 1. Cross-species single-cell sequencing. (**A**) A single-cell RNA-sequencing experiment was performed in isogenic human induced-pluripotent stem cell (iPSC)-derived cortical organoids at two time points and in postnatal day 7 mouse isocortex. An overview of the strategy to collect and filter differential gene expression data from both model systems is illustrated. (**B**) The human 3q29 deletion locus is nearly perfectly syntenic with a region of mouse chromosome 16, with the same gene order inverted. Corresponding loci are illustrated in the same orientation to facilitate clearer cross-species comparison. *Bex6* (in gray) is the only gene present in the mouse, not in the human locus. (**C** and **E**) UMAP dimensionality reduction plots colored by the main cell-types identified in human (**C**) and mouse (**E**) experiments. Human and mouse cells showed no obvious difference in gross distribution by genotype (**D**, **F**) but human cells were clearly divided in their transcriptomic clustering patterns by time point (d, top). The average expression profile of each sample was correlated (Spearman) to BrainSpan gene expression data profiling the human brain transcriptome in postmortem specimens across the lifespan (*2*) (**G**). Abbreviations: pcw, post-conception weeks (prenatal); m, months (postnatal); y, years (postnatal).

Fig 2. Transcriptomic evidence of metabolic changes in 3q29Del. The umbrella pathways most frequently found to be differentially-expressed based on up- (**B**) and down- (**C**) regulated genes in cortical organoids (**A**). Oxidative phosphorylation (OXPHOS) was enriched among both increased and decreased genes, but all clusters contributing to up-regulated OXPHOS were from 2-month organoids and all clusters contributing to down-regulated OXPHOS were from 12-month organoids. (**D**) Example violin plots visualizing log-normalized expression data of genes dysregulated in 2-month organoid clusters: *MT-CO3* (increased in 3q29Del) encodes the respiratory chain complex IV subunit COX3, *LDHA* (decreased in 3q29Del) is a key enzyme in glycolysis. (**E**) Example violin plots visualizing log-normalized expression data of genes

dysregulated in 12-month organoid clusters: *MT-ND1* (decreased in 3q29Del) encodes a component of respiratory chain complex I and *MT-ATP6* (decreased in 3q29Del) encodes a component of the ATP Synthase complex. The most frequently up- (**G**) and down- (**I**) regulated umbrella pathways in mouse isocortex (**F**) are shown. Treemaps derived by Revigo analysis (**H** and **J**) display the hierarchical organization of specific Gene Ontology Biological Processes (GO:BP) identified in pathway analysis. Similar colors denote semantic similarity. The size of each rectangle is proportional to the number of clusters exhibiting over-representation of a given GO:BP term. (All p-values are adjusted for multiple comparisons). Abbreviations: OPC, oligodendrocyte progenitor cells; NSC, neural stem cells; cl, cluster.

Fig 3. Common patterns of differential gene expression in two major mouse and human cell-types. Astrocytes were identified in human cortical organoids (12-month) and mouse isocortex. Corresponding clusters are color coded in blue in UMAP projections (**A**). The human homologs of mouse DEGs identified by MAST analysis were compared to organoid DEGs based on direction of change and a significant overlap was observed between the down-regulated DEGs of mouse and organoid astrocyte clusters (**B**). Pathway analysis of overlapping DEGs showed that all significantly enriched Gene Ontology: Biological Process (GO:BP) and Reactome (REAC) terms were related to mitochondrial function and metabolism (**C**). Upper and deep layer excitatory neuron DEGs were pooled and unique organoid DEGs were compared to the human homologs of mouse DEGs based on direction of change. Corresponding clusters are color coded in red in UMAP projections (**D**). There was a significant overlap between the DEGs of mouse and organoid excitatory neuron clusters for both up-regulated and down-regulated genes (**E**). Decreased genes were heavily enriched for GO:BP and REAC terms related to mitochondrial function and cellular respiration (**F**).

Fig 4. Mitochondrial phenotypes in 3q29 mice and engineered cell lines. Mitochondrial fractions from adult mouse brain lysates were found to have selective decreases in components of OXPHOS Complexes II and IV (**A**, quantified in **B**, N=5). At least 7 3q29-encoded proteins interact with mitochondria-localized proteins (**C**, from Antonicka et al. 2020). Symbol size reflects topological coefficients. HEK cell lines were engineered to carry either the heterozygous 3q29Del or completely lack PAK2 as shown by Western blot (**D**, one-way ANOVA, $F(2, 9)=237.7$, *CTRL vs 3q29* or *PAK2* ****$P<0.0001$). Control HEK-293T cells (*CTRL*) transition from a glycolytic to more aerobic cellular respiration state in galactose medium (**E**). Oxygen consumption rate (OCR) is significantly increased by 48-hour galactose medium challenge in *CTRL* cells (**F**, two-way ANOVA, main effect of medium $F(1, 6)=23.99$, **$P=0.0027$) but not in *3q29* (two-way ANOVA, $F(1, 6)=0.08808$, $P=0.7766$) or *PAK2* cells (two-way ANOVA, $F(1, 6)=0.6221$, $P=0.4603$). Both *3q29* and *PAK2* cells displayed increased baseline OCR (**G**, one-way ANOVA, effect of genotype $F(2, 9)=17.24$, $P=0.0008$; *CTRL vs. 3q29* ***$P=0.0005$, **CTRL vs PAK2* $P=0.0332$) and decreased response to galactose (**H**, one-way ANOVA, effect of genotype $F(2, 9)=8.838$, $P=0.0075$; *CTRL vs. 3q29* **$P=0.0075$, *CTRL vs PAK2* *$P=0.0138$). In glucose medium, *3q29* cells showed reduced spare capacity (**I**, two-way ANOVA effect of genotype, $F(2, 18)=448.0$, $P=0.0003$; *CTRL vs 3q29* **$P=0.0047$) and increased ATP production (**J**, two-way ANOVA effect of genotype, $F(2, 18)=4.309$, $P=0.0296$; *CTRL vs 3q29* **$P=0.0079$). Proton leak (**K**) was found to be increased in *3q29* cells in glucose (two-way ANOVA, main effect of genotype $F(2, 18)=31.16$, $P<0.0001$; *CTRL vs 3q29* ****$P<0.0001$) and decreased in *PAK2* cells in galactose (*CTRL vs PAK2* ***$P=0.0007$). Maximal respiration was significantly elevated in *3q29* cells in glucose (two-way ANOVA interaction of genotype and medium, $F(2, 18)=4.219$, $P=0.0314$; *CTRL vs 3q29* *$P=0.0364$) but was unchanged from *CTRL* in galactose conditions.

Figure 5. Lack of metabolic flexibility in 3q29Del neural progenitor cells. A) Control and 3q29Del neural progenitor cells (NPCs) exhibited normal morphology and stained positive for the

neurofilament protein Nestin, multipotency marker SOX2, and NPC marker PAX6 (scale = 50um, quantified in Fig. S15). **B**) Illustration of experimental design. NPCs were challenged for 48-hr in neural medium containing glucose (GLU) or galactose (GAL). **C**) Table of cell lines used in this experiment. Data from three separate cohorts was combined in plots D-M. N=15 from 6 independent NPC lines for all experiments. **D**) Energy map indicates that galactose treatment pushes cells from more glycolytic to more aerobic metabolic profile. **E**) Control NPCs significantly increase oxygen consumption rate (OCR) in galactose medium (two-way ANOVA main effect of medium, $F(1, 28)=9.295$, **P=0.0050). **F**) 3q29Del NPCs exhibited no significant change in OCR in galactose medium (two-way ANOVA, medium effect $F(1, 28)=0.01219$, P=0.9129). **G**) No significant difference in baseline OCR mean in glucose medium was observed (two-tailed ratio paired t-test, P=0.7015), but 3q29 NPCs displayed significantly lower baseline OCR mean in galactose medium (**H**, two-tailed ratio paired t-test, ***P=0.0009). **I**) Galactose response (i.e. basal OCR fold change over glucose) was unchanged in 3q29Del NPCs (two-tailed ratio paired t-test, P=0.0935). **J**) Maximal respiration was unchanged in glucose medium (two-tailed ratio paired t-test, P=0.5028), but was significantly decreased in 3q29Del NPCs in galactose medium (**K**, two-tailed ratio paired t-test, ***P=0.0007). Similarly, (**L**) the maximal respiration ratio of 3q29Del:Control NPCs was unchanged in glucose medium (GLU) but was significantly reduced in galactose conditions (GAL, one sample two-tailed t-test, **P=0.0026). There was no significant change in spare capacity in 3q29Del NPCs (**M**, two-way ANOVA genotype effect $F(1, 56)=0.5930$, P=0.4445).