Convergent and distributed effects of the schizophrenia-associated 3q29 deletion on the human neural transcriptome

Esra Sefik BA[1,4]*, Ryan H. Purcell PhD[2,3]*, The Emory 3q29 Project[1], Elaine F. Walker PhD[4], Gary J. Bassell PhD[2,3], Jennifer G. Mulle MHS PhD[1,5#]

[1]Department of Human Genetics, Emory University School of Medicine, Atlanta, GA, USA

[2]Department of Cell Biology, Emory University School of Medicine, Atlanta, GA, USA

[3]Laboratory of Translational Cell Biology, Emory University School of Medicine, Atlanta, GA, USA

[4]Department of Psychology, Emory University, Atlanta, GA, USA

[5]Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, USA

*These authors contributed equally

[#]Correspondence: Jennifer G. Mulle, MHS, PhD

Emory University School of Medicine

615 Michael St., Suite 301

Atlanta, Georgia 30322

Tel: (404) 727-3042 / Email: jmulle@emory.edu

Keywords: "Gene co-expression network", "Neuropsychiatric risk", "Systems biology", "Transcriptomics", "Human brain"

Abstract

The 3q29 deletion (3q29Del) confers >40-fold increased risk for schizophrenia. However, no single gene in this interval is definitively associated with disease, prompting the hypothesis that neuropsychiatric sequelae emerge upon loss of multiple functionally-connected genes. 3q29 genes are unevenly annotated and the impact of 3q29Del on the human neural transcriptome is unknown. To systematically formulate unbiased hypotheses about molecular mechanisms linking 3q29Del to neuropsychiatric illness, we conducted a systems-level network analysis of the non-pathological adult human cortical transcriptome and generated evidence-based predictions that relate 3q29 genes to novel functions and disease associations. The 21 protein-coding genes located in the interval segregated into seven clusters of highly co-expressed genes, demonstrating both convergent and distributed effects of 3q29Del across the interrogated transcriptomic landscape. Pathway analysis of these clusters indicated involvement in nervous-system functions, including synaptic signaling and organization, as well as core cellular functions, including transcriptional regulation, post-translational modifications, chromatin remodeling and mitochondrial metabolism. Top network-neighbors of 3q29 genes showed significant overlap with known schizophrenia, autism and intellectual disability-risk genes, suggesting that 3q29Del biology is relevant to idiopathic disease. Leveraging “guilt by association”, we propose nine 3q29 genes, including one hub gene, as prioritized drivers of neuropsychiatric risk. These results provide testable hypotheses for experimental analysis on causal drivers and mechanisms of the largest known genetic risk factor for schizophrenia and highlight the study of normal function in non-pathological post-mortem tissue to further our understanding of psychiatric genetics, especially for rare syndromes like 3q29Del, where access to neural tissue from carriers is unavailable or limited.

Introduction

Copy number variants (CNVs) offer tractable entry points to investigate the genetic contributions to complex neuropsychiatric diseases. The recurrent 1.6Mb deletion of the 3q29 interval (3q29Del) is robustly associated with schizophrenia spectrum disorders (SZ) [1-4] and is the strongest known risk allele for the disease with an estimated odds ratio >40 [5]. Autism spectrum disorders (ASD) and intellectual disability / developmental delay (IDD) are also enriched in 3q29Del carriers [6-8]. However, it is not currently known which genes within the interval are responsible for the increased neuropsychiatric risk. No single 3q29 interval gene has been definitively associated with SZ, ASD, or ID, prompting the hypothesis that haploinsufficiency of more than one gene is required [9]. The paucity of information regarding the functional roles of most 3q29 interval genes hampers the development of evidence-based hypotheses for deciphering this link. No transcriptomic investigation of 3q29Del in humans has been reported, and it is unclear what impact hemizygous loss of these genes might have in the nervous system.

Among the 21 protein-coding genes of the 3q29Del locus, several have been proposed as drivers of the behavioral phenotypes [10], yet the evidence for their individual association with neuropsychiatric disease remains suggestive. *DLG1* produces a synaptic scaffold protein that interacts with AMPA and NMDA-type glutamate receptors [11-14], the latter of which is hypothesized to be involved in SZ pathogenesis [15]. A *DLG1* polymorphism has been genetically linked to SZ [16, 17]. However, the mouse-specific phenotypes of 3q29Del are not recapitulated by haploinsufficiency of *Dlg1* alone [18]. Another prominent candidate, *PAK2* encodes a brain-expressed protein kinase involved in cytoskeletal dynamics [19] and dendritic spine morphology [20]. Both *DLG1* and *PAK2* are homologous to genes linked to ID [21, 22] and evidence from *Drosophila* indicates that joint haploinsufficiency of both genes simultaneously may be required for synaptic defects rather than either gene individually [23]. A recent study generated select combinatorial knockdowns of 3q29 gene homologs in *Drosophila* and *Xenopus laevis* and proposed that a component of the nuclear cap-binding complex, *NCBP2* [24], mediates neurodevelopmental defects in 3q29Del syndrome [9]. However, in the 3q29Del mouse model, NCBP2 is not decreased at the protein level in brain tissue [18], dampening enthusiasm for this gene as a causal element. It remains unclear which genes and their potential interactions are responsible for 3q29-associated phenotypes.

To avoid annotation-bias [25] and address the knowledge gap for 3q29 genes, we employed a gene co-expression network analysis approach that is rooted in systems biology [26-29], and generated evidence-based predictions that relate individual 3q29 genes to novel functions and disease phenotypes. Accumulating evidence indicates that genes work in conjunction, rather than in isolation, to realize most cellular functions [9, 23, 30]. Genes participating in the same molecular and biological pathways tend to show positively correlated expression with each other (co-expression), as they are often expressed under the control of a coordinated transcriptional regulatory system [31, 32]. In this holistic context, well-characterized genes can be leveraged to infer the functions of understudied genes by studying network patterns that emerge by means of co-expression [33-35]. This *in silico* approach to investigating unknown biology extends the "guilt by association" paradigm [36] that is extensively used for inductive reasoning in other domains to gene-gene interactions in complex biological systems [37-39]. Weighted gene co-expression network analysis (WGCNA) [27, 40] has been successfully deployed to study how genes embedded in network structures jointly affect complex human diseases [41-51]. We employed this paradigm to glean new biological insights into the 3q29Del syndrome.

Methods and Materials

The reference dataset

To uncover the network-level operations of genes located in the recurrent 3q29Del locus, we employed WGCNA [27, 40] and organized the non-pathological adult human cortical transcriptome into modules of highly co-expressed genes (Fig. 1a). Given the strong genetic link between 3q29Del and risk for SZ [5], we focused the present network analysis on revealing the clustering patterns of 3q29 interval genes as a function of their expression similarity during adulthood: a period when SZ typically manifests diagnostically, with peak onset in late adolescence and early adulthood [52], and a substantial proportion of patients becoming ill during middle adulthood [53]. A prior study has shown that only 0.7% of genes detected in the neocortex show a temporally regulated profile of differential expression during adulthood (between ages 20-60 years) [54]; hence, gene expression data were pooled across adulthood to derive the present dataset. We further focused our analysis spatially on gene-gene relationships in the PFC: a brain region that subserves a diverse range of cognitive and emotional operations, is most consistently implicated in the etiology of SZ [55] and may be particularly vulnerable to the effects of genetic disruption due to its protracted development [56]. For these reasons, the network was

constructed on publicly available transcriptomic data from the Genotype Tissue Expression Project (GTEx) [57], using prefrontal cortex (PFC; Brodmann Area 9) samples from male and female adults (age range = 20-79, 68.2% male) with no known history of psychiatric or neurological disorder (Fig. S1, Table S1.1). Transcriptome profiling was performed by RNA-Seq as described in [57].

Protein-coding transcripts were extracted from the dataset, normalized expression values were $\log_2$ transformed and summarized at the gene-level, and outlier samples were removed (Fig. S2). The data were corrected for covariance mediated by age, sex, death classification, post-mortem interval (PMI) and batch effect [58, 59] (Fig. 1b, Fig. S1). Genes with zero variance and genes and samples with greater than 50% missing entries (default) were removed [27, 40]. The normalized, outlier-removed, residualized expression values of 18,410 protein-coding genes from 107 samples constituted the final dataset.

Weighted and signed gene co-expression network construction

The single-block pipeline implemented in the WGCNA R package was employed for network construction [27, 40]. Co-expression similarity was defined by biweight midcorrelation [60, 61]. To capture the continuous nature of interactions and accentuate strong positive correlations, co-expression similarity was transformed into a signed and weighted adjacency matrix by a soft-thresholding procedure that yielded approximate scale-free topology [62-65] (Fig. 1c, Fig. S3). Topological overlap measures (TOM) were calculated from the resulting adjacency matrix to capture not only the univariate correlational relationship between gene-pairs but also the large-scale connections among “neighborhoods” of genes [66, 67] (Fig. 1a). Hence, we measure the interconnectedness of gene-pairs in relation to the commonality of the nodes they connect to.

The modular structure of the data was revealed by average linkage hierarchical clustering on TOM following its transformation into a dissimilarity metric (Fig. 1d). Module definitions used in this study do not use *a priori* knowledge of functionally defined gene-sets. Instead, modules were detected in a data-driven fashion through adaptively pruning the branches of the resulting dendrogram by the dynamic-hybrid-tree-cut method [68]. The expression profile of each identified module was subsequently summarized by a module eigengene (ME) [69] , defined as its first principal component. Calculation of MEs amounted to a data reduction method used for effectively correlating entire modules to one another and for establishing the eigengene-based module connectivity measure (kME) of individual genes [40]. To eliminate spurious assignment of genes into separate

modules, modules with strongly correlated MEs were amalgamated (Pearson's r > 0.8, $P < 0.05$, cut height = 0.2) (Fig. 1d).

Module preservation and quality analyses

To validate the reproducibility of the network modules derived from the GTEx dataset (considered the reference dataset/network), we evaluated module preservation in an independently-ascertained, demographically-comparable transcriptomic dataset, referred to as the test dataset/network (Fig. S4a). For this purpose, publicly available transcriptomic data was obtained from the BrainSpan Developmental Transcriptome Project [54]. 30 non-pathological post-mortem samples from the PFC of male and female adults (age range = 18-37, 50% male) with no known neurological or psychiatric disorder comprised the test dataset (Fig. S4b, Table S1.1). Transcriptome profiling was performed by RNA-Seq as described in [70].

The pre-processed test dataset consisted of normalized and residualized expression values for 18,339 protein-coding genes from 30 samples that were pooled from four sub-regions of the PFC to test whether the co-expression patterns derived from Brodmann Area 9 of the PFC in the reference dataset could be commonly found and robustly defined in broader sampling of tissue across the PFC. Prior to preservation analysis, a sample-level hierarchical clustering of the test dataset was conducted, which revealed no distribution bias associated with PFC sub-region, ruling out tissue of origin as a potential confound in test network construction. (Fig. S4c).

To determine whether properties of within-module topology were preserved in the test network, we calculated a composite, network-based preservation statistic for each module ($Z_{summary.pres}$) by using the *modulePreservation* function of the WGCNA package in R [71]. $Z_{summary.pres}$ is a summary statistic that encompasses multiple density-based and connectivity-based preservation statistics, which are equally important for establishing the overall preservation of a module. To determine whether the observed preservation statistics were higher than expected by chance, we randomly permuted the module assignments in the test dataset (200 times) and derived a standardized $Z_{summary.pres}$ score for each module. To account for this metric's dependence on module-size, we reduced large modules by randomly sampling 1000 intra-modular nodes. The resulting scores were evaluated according to established thresholds: $Z_{summary.pres} < 2$, no evidence for preservation; $2 < Z_{summary.pres} < 10$, moderate evidence for preservation; $Z_{summary.pres} > 10$, strong evidence for preservation.

In addition to testing preservation, we measured the quality of the modules that were defined in the reference network by employing a resampling technique that applied the preservation statistics outlined above to repeated random splits of the reference dataset. We assessed the robustness of the identified modules (i.e., how distinct a module is from the background) by calculating a standardized, composite quality statistic ($Z_{summary.qual}$), as described in [71]. The same $Z_{summary.pres}$ thresholds outlined above were used to evaluate $Z_{summary.qual}$.

Functional characterization of network modules harboring 3q29 genes

Pathway analyses of individual modules found to harbor 3q29 genes were performed by g:Profiler (http://biit.cs.ut.ee/gprofiler), using annotations from the Gene Ontology (GO), Reactome and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases. Enriched terms surpassing the adjusted g:SCS significance threshold of $P < 0.05$ were filtered for size and semantic similarity [72].

Identification of prioritized driver genes and biological mechanisms

Disease-associated genes are often more closely connected to each other than random gene-pairs in a biological network; this non-random network characteristic has enabled the identification of novel genetic risk loci for many diseases [73-77]. To generate data-driven hypotheses about which 3q29 genes are causally linked to the major neuropsychiatric phenotypes of 3q29Del, we tested the overlap between “top neighbors” of individual 3q29 genes and known risk genes for SZ and related disorders (Fig. 3a). A top neighbor was defined as any node whose gene expression profile has a moderate-to-high correlation (Spearman’s rho (ρ) ≥ 0.5), $P < 0.05$) with a given 3q29 gene (considered a “seed”) within the same module. Hence, top neighbors were identified by a hard-thresholding method applied only to intramodular edges of a seed that were initially defined by the topological overlap principle.

Lastly, to formulate testable hypotheses about key biological mechanisms that link the 3q29 locus to neuropsychiatric disease, we interrogated the functional enrichment of prioritized driver genes, using the same pathway analysis approach applied to modules. See Supplemental Methods for details.

Results

Unbiased gene co-expression network analysis reveals convergent and distributed effects of 3q29 interval genes across the adult human cortical transcriptome.

Applying an unsupervised WGCNA approach [27, 40] to publicly-available data from the GTEx Project [57] revealed that the protein-coding transcriptome of the healthy adult human PFC can be organized into a gene co-expression network of 19 modules (labeled by color) (Fig. 1d, Fig. 2a). The identified modules group genes with highly similar expression profiles and likely represent shared function and/or co-regulation. The resulting module sizes ranged from 43 (steel-blue module) to 4,746 (green module) genes, with an average module size of 1,014 genes. To obtain high-quality module definitions, one module (grey module) was reserved for genes that could not be unequivocally assigned to any module. Thus, the grey module was excluded from downstream analysis. The resulting set of modules was screened for membership of 3q29 genes; modules that were found to harbor at least one 3q29 gene are rereferred to as “3q29 modules”.

To ensure the reproducibility and robustness of our network analysis results, we tested the preservation and quality of the identified modules in an independent dataset obtained from the BrainSpan Developmental Transcriptome Project [54] (Fig. S4) and in repeated random splits of the reference dataset. These analyses revealed the replicable, well-defined and non-random nature of the identified network modules (Fig. 1d, Fig. S5, Table S1.4).

The 21 protein-coding genes located in the 3q29 interval clustered into seven modules (Fig. 2a): black (one 3q29 gene), brown (four 3q29 genes), dark-turquoise (one 3q29 gene), green (six 3q29 genes), magenta (one 3q29 gene), midnight-blue (three 3q29 genes), and turquoise (five 3q29 genes). Within this network, 18 (86%) of the 3q29 interval genes concentrate into just four modules (Fig. 2a), suggesting that the haploinsufficiency of sets of genes within the locus may perturb the same biological processes via multiple hits, cumulatively disrupting redundancy and compensatory resiliency in the normative regulation of cellular functions.

To evaluate whether modules further clustered within larger meta-modules that represent the higher-order organization of the transcriptome, we identified meta-modules as tight clusters of positively correlated MEs, detectable as major branches of the eigengene dendrogram [78] (Fig. 2a, 2b). Meta-modules were screened to identify the grouping patterns among 3q29 modules, allowing exploration of extra-modular interactions. This analysis revealed that the 3q29 modules further cluster into three higher level meta-modules (Fig. 2b), which

likely reflect dependencies and interactions between pathways involving 3q29 genes. Simultaneously, leading presumptive candidates *DLG1* and *PAK2* were found in opposite branches of the network, demonstrating the distributed effects of this CNV across the transcriptomic landscape.

Pathway analysis points to functional involvement of the 3q29 locus in nervous-system functions and core aspects of cell biology.

Since highly co-expressed genes often share similar functions [31, 32], biological processes and pathways that are enriched in a co-expression module can be used to infer functional information for poorly annotated genes of that module. Functional enrichment analysis of 3q29 modules showed that their constituent genes converge onto canonical biological processes at proportions greater than expected by chance, indicating that these modules are biologically relevant units (Fig. 2d, Table S1.5). The turquoise and green modules showed overrepresentation of roles specific to the neuronal system and implicate involvement in multiple synaptic properties. Other 3q29 modules point to biological pathways that may also underlie neuropsychiatric pathology in 3q29Del. The magenta module was predominantly enriched for protein modification, turnover, and localization. Additionally, a link was identified between the magenta module and the initiation of major histocompatibility complex class-I (MHC-I)-dependent immune responses, driven by a genomic locus implicated in the etiology of SZ [79, 80]. On the other hand, overrepresented pathways in the black module encompass regulation of gene expression and maintenance of the integrity of the cellular genome, including DNA repair, and the metabolism and processing of RNA. The midnight-blue module shared enriched roles with the black module, validating their shared meta-module structure, yet it was set apart by its involvement in cell cycle regulation. The brown module revealed primary enrichment for cellular metabolism and mitochondrial function, whereas the dark-turquoise module coalesced genes involved in epigenetic mechanisms and in signal transduction pathways mediated by Rho GTPases. This latter function may be attributable to *PAK2*, which encodes a known Rho GTPase effector. Intriguingly, this module was also enriched for estrogen receptor-mediated signaling, suggesting a potential mechanism for sex-specific effects. Taken together, functional characterization of the 3q29 modules point to novel mechanisms of shared or synchronized action for co-clustering 3q29 genes (Fig. 2d).

UBXN7 is a highly connected cortical hub-gene predicted to play a crucial role in the neuropsychiatric sequela of 3q29Del.

Targeted disruption of a highly-connected "hub" gene produces a more deleterious effect on network function and yields a larger number of phenotypic outcomes than randomly selected or less connected genes [81, 82]. Hence, we sought to measure how strongly connected individual 3q29 genes are to their modules by evaluating their intramodular kME (Fig. 2c, Table S1.3), defined as the Pearson's correlation between the expression profile of a gene and the eigengene of its assigned module [27, 40, 69].

Genes with high intramodular kME are considered hub genes that are predicted to be critical components of the overall function of their module [27]. Nodes with high intramodular kME often have high intramodular connectivity (kIM), which reflects sum of adjacencies to other nodes [40]. However, an advantage of using kME over other connectivity metrics, is its defined *P*-value and values that lie between -1 and 1, allowing comparison across modules that differ in size. To generate rigorous predictions about which 3q29 genes, if any, are intramodular hub genes, we adopted a conservative criterion that defines hub genes as nodes with kME > 0.8 ($P < 0.05$). Only one 3q29 interval gene was identified as a hub gene: *UBXN7* (kME = 0.84, $P = 8.33E-30$), which encodes a ubiquitin ligase-substrate adaptor [83, 84].

SMCO1, SLC51A, and *MFI2* had non-significant kMEs ($P < 0.05$) for their module, suggesting low intramodular connectivity. These 3q29 genes are detected but display very low abundance in the human cerebral cortex [85] (Table S1.6), which may relate to their peripheral network assignments in our analysis. Consequently, *SMCO1, SLC51A* and *MFI2* were excluded from downstream analysis to derive the most parsimonious prioritization of driver genes based on tight network connections. A complete list of gene-sets for each module and kME values are provided in Tables S1.2-3.

Nine 3q29 interval genes form transcriptomic subnetworks enriched for known SZ, ASD and IDD-risk genes.

We next identified a refined subset of target genes (top neighbors) that not only co-cluster based on TOM but also have a strong pairwise correlation with 3q29 genes (Fig. 3a). Several 3q29 genes were found to be top neighbors of one another. *FBXO45* ($\rho = 0.5$, $P = 5.43E-09$) and *PIGX* ($\rho = 0.6$, $P = 1.24E-10$) are top-neighbors of *CEP19*, while *SENP5* and *WDR53* are top-neighbors of each other ($\rho = 0.5$, $P = 1.05E-07$) (Table S2.1). On

the other hand, intramodular connections of *TM4SF19* and *ZDHHC19* did not meet top neighborhood criteria; hence they were not included in downstream analysis. Similar to *SMCO1, SLC51A* and *MFI2,* their mRNA expression profiles indicate low abundance in the cerebral cortex (Table S1.6), which likely reflects their lack of strong network connections.

The human transcriptome is theorized to demonstrate non-random topological characteristics, where disease genes interact with other disease genes that underlie a common pathophenotype [86]. Concordant with this prediction, within the top neighbors of 3q29 genes, we found several genes that have been extensively implicated in neuropsychiatric disease (Table S2.1). These include *MECP2*, *NRXN1*, *GRIN2A*, *GRIN2B*, *CHD8*, *SATB2*, *CNTNAP2*, *FOXP1*, *PTEN*, and *SCN2A*. Motivated by this observation, we asked whether top neighbors of individual 3q29 genes significantly overlap with known SZ, ASD or intellectual disability / developmental delay (IDD)-risk genes (Fig. 3a). We curated six evidence-based lists of SZ [87-91], ASD [92, 93] and IDD-risk genes [94], which span loci across a wide range of the allele frequency spectrum and include post-mortem findings from case-control gene expression studies (Table S2.2). Hypergeometric tests were conducted to gauge the probability of the overlap between curated gene-sets and top neighbors, as implemented in the *GeneOverlap* package in R [95], followed by Benjamini-Hochberg multiple testing correction. 3q29 genes whose top neighbors showed an overrepresentation of SZ, ASD and/or IDD risk genes (adjusted $P < 0.05$) were subsequently prioritized as likely genetic drivers of neuropsychiatric risk in 3q29Del syndrome, along with their SZ, ASD, and/or IDD-related top neighbors from the enrichment findings (Fig. 3a). We found overrepresentation of one or more established risk gene-sets among the top neighbors of nine 3q29 genes (Benjamini-Hochberg adjusted-$P < 0.05$) (Fig. 3b, Table S2.3).

To evaluate the specificity of the identified patterns of polygenic disease burden, we also tested these top neighbors for overlap with known Parkinson's disease (PD) [96], late-onset Alzheimer's disease (AD) [97], and inflammatory bowel disease (IBD) risk genes [98] (Table S2.2). These phenotypes have no known link to 3q29Del, thus, their risk loci were considered negative controls. Common variants associated with height [99] (Table S2.2) were included as a fourth negative control to rule out a potential bias associated with large differences in the sizes of curated gene-sets.

Concurrently, we found no statistically significant evidence for overrepresentation of AD or IBD-risk genes among the interrogated top neighbors (Fig. 3b). Only the top neighbors of *SENP5* showed a significant overlap

with height-associated genes (adjusted-*P* = 2.36E-02), and the top neighbors of *NRROS*, which did not show an enrichment for known IDD, ASD, or SZ risk genes, exhibited a small but significant overlap with known PD-risk genes (adjusted-*P* = 2.00E-02) (Fig. 3b). Overall, 19 out of 96 hypergeometric tests (20%) revealed a significant overrepresentation of SZ, ASD, and/or IDD-risk gene-sets among the top neighbors of 3q29 genes. By contrast, only 2 out of 64 (3%) hypergeometric tests indicated a significant overlap with the negative control gene-sets. The substantial margin between these two enrichment ratios supports the high specificity of our network-derived inferences for uncovering biology relevant to 3q29Del. By leveraging guilt by association, we prioritize *BDH1, CEP19, DLG1, FBXO45, PIGZ, RNF168, SENP5, UBXN7, and WDR53*, along with their 284 unique SZ, ASD, and/or IDD-related top neighbors from significant overlap tests as likely drivers of the neuropsychiatric consequences of 3q29Del (Fig. 4a, Table S2.3).

Disease-relevant driver genes prioritized by network analysis load onto key biological pathways linked to neuropsychiatric disorders.

To formulate testable hypotheses about the biological mechanisms linking the 3q29 locus to neuropsychiatric phenotypes, we interrogated whether the prioritized driver genes identified in our network analysis assemble into known biological pathways. Functional enrichment analysis on the union of 293 prioritized driver genes (including nine 3q29 genes) revealed significant overrepresentation of eight biological pathways annotated by the Reactome and KEGG databases (Fig. 4b, Table S2.4). These include axon guidance (adjusted-*P* = 3.64E-03), long-term potentiation (adjusted-*P* = 7.29E-03), and regulation of actin cytoskeleton (adjusted-*P* = 1.17E-02). Additionally, several GO biological processes (GO:BP), including chromosome organization (adjusted-*P* = 3.81E-09), histone modification (adjusted-*P* = 3.31E-08), neuron differentiation (adjusted-*P* = 1.88E-04), neurogenesis (adjusted-*P* = 1.89E-03), and excitatory postsynaptic potential (adjusted-*P* = 8.97E-03) were overrepresented among the predicted drivers (Fig. 4b-c, Table S2.4). We hypothesize that the disruption of one or more of these biological pathways and processes, some of which have been demonstrated to be altered in idiopathic SZ and ASD [2, 100], lie on the casual pathway to neuropsychopathology in 3q29Del syndrome. For extended results, see Supplemental Results.

Discussion

The 3q29Del has been reliably associated with extraordinary risk for serious neuropsychiatric illness and therefore may offer key insights to advance our understanding of the biological basis of these complex disorders. Currently, the driver genes and affected biological pathways that link 3q29Del to neuropsychiatric pathology remain unknown. To avoid bias introduced by annotation-based criteria in the formulation of mechanistic hypotheses, we engaged a system-level vantage point and interrogated the collective behavior of 3q29 interval genes with the global protein-coding transcriptome of the healthy human brain. We leveraged publicly available transcriptomic data from the GTEx Project [57] to perform WGCNA [27, 40] on postmortem cortical samples from donors with no known history of psychiatric or neurological disease. Given that precise actions of genes are often dependent on tissue and age [54], we focused our analysis on the adult PFC, a brain region most consistently implicated in SZ [55], which typically strikes during adulthood. We analyzed the resulting network to identify the modular properties and undirected connectivity patterns of the 3q29 interval, which yielded key predictions into inter-related functions and disease associations. Our findings provide foundational information to formulate rigorous, targeted and testable hypotheses on the causal drivers and mechanisms underlying the largest known single genetic risk factor for SZ.

Genomic studies have identified several recurrent CNVs that confer high risk for neuropsychiatric disorders [2, 8]. The current challenge is to understand which genes within these loci are the major drivers of risk. In the 3q29 locus, *DLG1* and *PAK2* have been most often proposed as candidate drivers of neuropsychiatric phenotypes [10, 23, 101]. Indeed, a recent literature search revealed more publications related to these genes than of all other 3q29Del genes combined (Fig. S6). Consistent with previous reports of an association between *DLG1* and SZ [16, 17], the current study presents network-level evidence for prioritizing *DLG1* as a neuropsychiatric disease-linked gene. Surprisingly, however, our analysis does not support inclusion of *PAK2* as a predicted driver of neuropsychiatric risk. Instead, our results lend support to *DLG1* and eight other 3q29 genes, most of which are largely understudied, as key players in 3q29Del syndrome. Our unbiased approach prioritizes *BDH1, CEP19*, *DLG1*, *FBXO45, PIGZ, RNF168, SENP5, UBXN7* and *WDR53* as primary drivers.

It is currently unknown whether the biological basis of neuropsychiatric risk associated with recurrent CNVs overlaps with that of individuals who share the same clinical diagnosis but do not share the same rare genetic variant. Our findings suggest that molecular perturbations caused by the hemizygous deletion of select 3q29

genes may overlap with the genetic etiologies contributing to idiopathic forms of SZ, ASD and ID. Disease-relevant driver genes prioritized by our network analysis are enriched for canonical biological pathways, such as neurogenesis, neuron differentiation, synapse organization, excitatory postsynaptic potential, long-term potentiation, axon guidance, regulation of actin cytoskeleton, signal transduction, post-translational protein modifications, chromatin organization and histone modification. We hypothesize that the disruption of one or more of these biological processes, some of which are altered in idiopathic SZ and ASD [2, 100], lie on the casual pathway to neuropsychopathology in 3q29Del syndrome.

No single gene within the interval has been definitively associated with neuropsychiatric disease, prompting the hypothesis that neuropathology in 3q29Del emerges upon loss of multiple genes that are functionally connected. While a single nucleotide polymorphism in *DLG1* has been associated with SZ in a case-control study [16, 17], the risk associated with this variant does not approach that of 3q29Del, suggesting that the neuropsychiatric risk associated with this CNV is distributed across more than one gene in the locus. To investigate functional connections across multiple 3q29 genes, we conducted an unsupervised analysis of the modular organization of the adult human PFC. We found the 21 3q29 genes distributed into three meta-modules and seven modules, with 18 genes converging into just four modules. Hence, 3q29 genes display both distributed and convergent effects in the adult human cortical transcriptome. Rather than functioning as independent agents, sets of 3q29 genes may have shared and/or synchronized function and constitute interacting sources of pathology. It is conceivable that the consequences of the haploinsufficiency arise through the weakening of multiple distinct pathways that normally provide protective redundancies (distributed model), and/or through multiple insults to a functionally connected module that cumulatively disrupt resiliency (convergent biology model). These hypotheses warrant further testing.

A major goal of this study was to infer unknown functions for understudied 3q29 genes by leveraging well-studied co-clustering genes. Pathway analysis of modules harboring 3q29 genes revealed likely functional involvement of the 3q29 locus in not only nervous-system specific functions, but also in core aspects of cell biology that are non-specific to an organ system. The closely-related black and midnight-blue modules were significantly associated with regulation of gene expression, chromatin organization and DNA repair. The green and turquoise modules were both associated with nervous system development and function, and in particular, regulation and organization of synaptic signaling and components. This finding is surprising because most of the

3q29 genes located in these latter two modules have not been identified as synaptic genes. Similarly, the 3q29 genes in the black and midnight-blue modules have not been implicated in gene regulatory pathways or DNA repair. We maintain that biological functions of poorly annotated genes can be inferred through the graph-based modeling of inter-gene relationships. Thereby, we predict novel roles for individual 3q29 genes in functions related to synaptic transmission, modulation of neurotransmission, synapse structure and function, mitochondrial metabolism, transcriptional and translational regulation, chromatin remodeling, cell cycle regulation, and protein modification, localization and turnover. We propose that the subset of predicted functions that are non-specific to an organ system likely contribute to global developmental outcomes in 3q29Del.

Analysis of eigengene-based connectivity revealed that *UBXN7* is a hub gene, with top neighbors enriched for known association with all three major neuropsychiatric phenotypes of 3q29Del. Hub nodes of biological networks are often associated with human disease [102, 103]. Disease-genes, identified from OMIM's Morbid Map of the Human Genome, disproportionately exhibit hub-gene characteristics, with protein products participating in more known protein-protein interactions than that of non-disease genes [104]. Supported by this literature, we predict that (1) *UBXN7* exerts critical influence on a large network of co-expressed genes, and (2) loss-of-function (LoF) mutations in *UBXN7* can cause major dysfunction in affiliated biological pathways (indeed, its pLI score = 0.99, i.e. extremely intolerant to LoF [105]). We prioritize *UBXN7* as a major driver in 3q29Del syndrome. *UBXN7* has not been previously linked to neuropsychiatric disorders or proposed as a candidate driver of 3q29Del syndrome. However, *UBXN7* has been reported to regulate the ASD-associated E3 ubiquitin ligase Cullin-3 (*CUL3*) [106], an interaction that deserves more attention in light of our findings [84].

One limitation of the current study is its singular focus on protein-coding elements. How the non-coding elements of the interval, along with splice variants, integrate into the predictions formulated in this study is ripe for future investigation. Overall, the transcriptomic network identified in this study is predicted to connect 3q29 interval genes with gene-sets outside the interval that participate in the same or overlapping biological process and associate with similar disease phenotypes. Perturbation of 3q29 interval gene dosage is expected to also perturb the functioning of network-partners outside the recurrent 3q29Del locus. However, note that the underlying structure of weighted gene co-expression networks is agnostic to the mechanistic order of cellular and molecular events. The information necessary to derive the order of biological interactions is not an explicit outcome of gene co-expression itself, since such inferences require time-dependent analysis of combinatorial

interactions between nodes. As a result, some of the network partners identified in this study are expected to function upstream of their co-expressed 3q29 interval partner, and likely not be affected by 3q29Del. Moreover, the complex interactions between molecules can be dynamic across time and space [54]. Hence, a future direction will be to ask whether the network connections formed by 3q29 interval genes in the adult PFC show differential expression in the neural tissue of 3q29Del carriers and whether they show temporal and/or spatial variation.

Finally, our analysis does not preclude the possibility that other 3q29 interval genes moderate phenotypic expressivity. For example, while the dark turquoise module (including *PAK2*) did not harbor prioritized driver genes, it was significantly associated with estrogen receptor-mediated signaling. An intriguing emerging feature of 3q29Del syndrome is the markedly reduced sex bias in risk for ASD [7]. Additional studies will be required to assess the drivers of sex-specific phenotypes of 3q29Del syndrome.

Now that recurrent, highly penetrant CNV loci have been identified as important risk factors for neuropsychiatric disorders, determination of the component genes driving this risk is the next step toward deciphering mechanisms. We used an unbiased systems biology approach that leveraged the power of open data to infer unknown functions for understudied 3q29 interval genes, and to refine the 3q29 locus to nine prioritized driver genes, including one hub gene. Importantly, this approach can partially overcome barriers to formulation of relevant hypotheses that are introduced by poor annotation of interval genes, without requiring laborious, expensive, and time-consuming experiments to functionally characterize all genes within the interval. Our results reveal the power of this approach for prioritization of putative drivers. Ongoing and future studies will be directed at understanding how these genes work in concert and how multiple haploinsufficiencies confer risk for neuropsychiatric disease.

Acknowledgments

This study was supported by NIH R01 MH110701 (J.G.M. and G.J.B.), F32 MH124273 (R.H.P.), and the Emory University School of Medicine.

The GTEx data (release version 6) used for the analyses described in this manuscript were downloaded from the GTEx portal on 01/08/2019 (http://www.gtexportal.org/home/datasets/, file name: “GTEx_Analysis_v6_RNA-seq_RNA-SeQCv1.1.8_gene_rpkm.gct.gz”); dbGaP accession number: phs000424.v6.p1. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS.

The BrainSpan Developmental Transcriptome dataset used to construct the test network was downloaded from the Allen Brain Atlas portal on 01/12/2019 (https://www.brainspan.org/static/download/, file name: “RNA-Seq Gencode v10 summarized to genes”); dbGaP accession number: phs000755.v2.p1.

The authors gratefully acknowledge the contributions of the members of the Emory 3q29 Project: Katrina Aberizk, Hallie Averbach, T. Lindsey Burrell, Shanthi Cambala, Grace Carlock, Tamara Caspary, Joseph F. Cubells, David Cutler, Paul A. Dawson, Michael P. Epstein, Roberto Espana, Michael J. Gambello, Katrina Goines, Ryan Guest, Henry R. Johnston, Cheryl Klaiman, Sookyong Koh, Elizabeth J. Leslie, Longchuan Li, Bryan Mak, Tamika Malone, Michael Mortillo, Trenell Mosley, Melissa M. Murphy, Derek Novacek, Rebecca M. Pollak, Rossana Sanchez, Celine A. Saulnier, Jason Schroeder, Sarah Shultz, Nikisha Sisodiya, Steven Sloan, Stephen T. Warren, David Weinshenker, Zhexing Wen, Stormi White, and Michael E. Zwick.

Conflict of Interest

The authors have no competing financial interests related to this study to declare.

References

1. Levinson DF, Duan J, Oh S, Wang K, Sanders AR, Shi J *et al.* Copy number variants in schizophrenia: confirmation of five previous findings and new evidence for 3q29 microdeletions and VIPR2 duplications. *Am J Psychiatry* 2011; **168**(3)**:** 302-316.

2. Marshall CR, Howrigan DP, Merico D, Thiruvahindrapuram B, Wu W, Greer DS *et al.* Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nat Genet* 2017; **49**(1)**:** 27-35.

3. Mulle JG, Dodd AF, McGrath JA, Wolyniec PS, Mitchell AA, Shetty AC *et al.* Microdeletions of 3q29 confer high risk for schizophrenia. *Am J Hum Genet* 2010; **87**(2)**:** 229-236.

4. Szatkiewicz JP, O'Dushlaine C, Chen G, Chambert K, Moran JL, Neale BM *et al.* Copy number variation in schizophrenia in Sweden. *Mol Psychiatry* 2014; **19**(7)**:** 762-773.

5. Mulle JG. The 3q29 deletion confers >40-fold increase in risk for schizophrenia. *Mol Psychiatry* 2015; **20**(9)**:** 1028-1029.

6. Glassford MR, Rosenfeld JA, Freedman AA, Zwick ME, Mulle JG, Unique Rare Chromosome Disorder Support G. Novel features of 3q29 deletion syndrome: Results from the 3q29 registry. *Am J Med Genet A* 2016; **170A**(4)**:** 999-1006.

7. Pollak RM, Murphy MM, Epstein MP, Zwick ME, Klaiman C, Saulnier CA *et al.* Neuropsychiatric phenotypes and a distinct constellation of ASD features in 3q29 deletion syndrome: results from the 3q29 registry. *Mol Autism* 2019; **10:** 30.

8. Sanders SJ, He X, Willsey AJ, Ercan-Sencicek AG, Samocha KE, Cicek AE *et al.* Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron* 2015; **87**(6)**:** 1215-1233.

9. Singh MD, Jensen M, Lasser M, Huber E, Yusuff T, Pizzo L *et al.* NCBP2 modulates neurodevelopmental defects of the 3q29 deletion in Drosophila and Xenopus laevis models. *PLoS Genet* 2020; **16**(2)**:** e1008590.

10. Carroll LS, Williams HJ, Walters J, Kirov G, O'Donovan MC, Owen MJ. Mutation screening of the 3q29 microdeletion syndrome candidate genes DLG1 and PAK2 in schizophrenia. *Am J Med Genet B Neuropsychiatr Genet* 2011; **156B**(7)**:** 844-849.

11. Gardoni F, Mauceri D, Fiorentini C, Bellone C, Missale C, Cattabeni F *et al.* CaMKII-dependent phosphorylation regulates SAP97/NR2A interaction. *J Biol Chem* 2003; **278**(45)**:** 44745-44752.

12. Leonard AS, Davare MA, Horne MC, Garner CC, Hell JW. SAP97 is associated with the alpha-amino-3-hydroxy-5-methylisoxazole-4-propionic acid receptor GluR1 subunit. *J Biol Chem* 1998; **273**(31)**:** 19518-19524.

13. Lin EI, Jeyifous O, Green WN. CASK regulates SAP97 conformation and its interactions with AMPA and NMDA receptors. *J Neurosci* 2013; **33**(29)**:** 12067-12076.

14. Zhou W, Zhang L, Guoxiang X, Mojsilovic-Petrovic J, Takamaya K, Sattler R *et al.* GluR1 controls dendrite growth through its binding partner, SAP97. *J Neurosci* 2008; **28**(41)**:** 10220-10233.

15. Moghaddam B, Javitt D. From revolution to evolution: the glutamate hypothesis of schizophrenia and its implication for treatment. *Neuropsychopharmacology* 2012; **37**(1)**:** 4-15.

16. Uezato A, Kimura-Sato J, Yamamoto N, Iijima Y, Kunugi H, Nishikawa T. Further evidence for a male-selective genetic association of synapse-associated protein 97 (SAP97) gene with schizophrenia. *Behav Brain Funct* 2012; **8:** 2.

17. Uezato A, Yamamoto N, Jitoku D, Haramo E, Hiraaki E, Iwayama Y *et al.* Genetic and molecular risk factors within the newly identified primate-specific exon of the SAP97/DLG1 gene in the 3q29 schizophrenia-associated locus. *Am J Med Genet B Neuropsychiatr Genet* 2017; **174**(8)**:** 798-807.

18. Rutkowski TP, Purcell RH, Pollak RM, Grewenow SM, Gafford GM, Malone T *et al.* Behavioral changes and growth deficits in a CRISPR engineered mouse model of the schizophrenia-associated 3q29 deletion. *Mol Psychiatry* 2019.

19. Bokoch GM. Biology of the p21-activated kinases. *Annu Rev Biochem* 2003; **72:** 743-781.

20. Wang Y, Zeng C, Li J, Zhou Z, Ju X, Xia S *et al.* PAK2 Haploinsufficiency Results in Synaptic Cytoskeleton Impairment and Autism-Related Behavior. *Cell Rep* 2018; **24**(8)**:** 2029-2041.

21. Allen KM, Gleeson JG, Bagrodia S, Partington MW, MacMillan JC, Cerione RA *et al.* PAK3 mutation in nonsyndromic X-linked mental retardation. *Nat Genet* 1998; **20**(1)**:** 25-30.

22. Tarpey P, Parnau J, Blow M, Woffendin H, Bignell G, Cox C *et al.* Mutations in the DLG3 gene cause nonsyndromic X-linked mental retardation. *Am J Hum Genet* 2004; **75**(2)**:** 318-324.

23. Grice SJ, Liu JL, Webber C. Synergistic interactions between Drosophila orthologues of genes spanned by de novo human CNVs support multiple-hit models of autism. *PLoS Genet* 2015; **11**(3)**:** e1004998.

24. Pabis M, Neufeld N, Shav-Tal Y, Neugebauer KM. Binding properties and dynamic localization of an alternative isoform of the cap-binding complex subunit CBP20. *Nucleus* 2010; **1**(5)**:** 412-421.

25. Johnson EC, Border R, Melroy-Greif WE, de Leeuw CA, Ehringer MA, Keller MC. No Evidence That Schizophrenia Candidate Genes Are More Associated With Schizophrenia Than Noncandidate Genes. *Biol Psychiatry* 2017; **82**(10)**:** 702-708.

26. Eisenberg D, Marcotte EM, Xenarios I, Yeates TO. Protein function in the post-genomic era. *Nature* 2000; **405**(6788)**:** 823-826.

27. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* 2005; **4:** Article17.

28. Ma'ayan A, Blitzer RD, Iyengar R. Toward predictive models of mammalian cells. *Annu Rev Biophys Biomol Struct* 2005; **34:** 319-349.

29. Barabasi AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nat Rev Genet* 2011; **12**(1)**:** 56-68.

30. Qiu Y, Arbogast T, Lorenzo SM, Li H, Tang SC, Richardson E *et al.* Oligogenic Effects of 16p11.2 Copy-Number Variation on Craniofacial Development. *Cell Rep* 2019; **28**(13)**:** 3320-3328 e3324.

31. Singer GA, Lloyd AT, Huminiecki LB, Wolfe KH. Clusters of co-expressed genes in mammalian genomes are conserved by natural selection. *Mol Biol Evol* 2005; **22**(3)**:** 767-775.

32. de la Fuente A. From 'differential expression' to 'differential networking' - identification of dysfunctional regulatory networks in diseases. *Trends Genet* 2010; **26**(7)**:** 326-333.

33. Wolfe CJ, Kohane IS, Butte AJ. Systematic survey reveals general applicability of "guilt-by-association" within gene coexpression networks. *BMC Bioinformatics* 2005; **6:** 227.

34. Horan K, Jang C, Bailey-Serres J, Mittler R, Shelton C, Harper JF *et al.* Annotating genes of known and unknown function by large-scale coexpression analysis. *Plant Physiol* 2008; **147**(1)**:** 41-57.

35. Ahn AC, Tewari M, Poon CS, Phillips RS. The limits of reductionism in medicine: could systems biology offer an alternative? *PLoS Med* 2006; **3**(6)**:** e208.

36. Oliver S. Guilt-by-association goes global. *Nature* 2000; **403**(6770)**:** 601-603.

37. Lee HK, Hsu AK, Sajdak J, Qin J, Pavlidis P. Coexpression analysis of human genes across many microarray data sets. *Genome Res* 2004; **14**(6)**:** 1085-1094.

38. Moreau Y, Tranchevent LC. Computational tools for prioritizing candidate genes: boosting disease gene discovery. *Nat Rev Genet* 2012; **13**(8)**:** 523-536.

39. Oti M, Brunner HG. The modular nature of genetic diseases. *Clin Genet* 2007; **71**(1)**:** 1-11.

40. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008; **9:** 559.

41. Radulescu E, Jaffe AE, Straub RE, Chen Q, Shin JH, Hyde TM *et al.* Identification and prioritization of gene sets associated with schizophrenia risk by co-expression network analysis in human brain. *Mol Psychiatry* 2020; **25**(4)**:** 791-804.

42. Etemadikhah M, Niazi A, Wetterberg L, Feuk L. Transcriptome analysis of fibroblasts from schizophrenia patients reveals differential expression of schizophrenia-related genes. *Sci Rep* 2020; **10**(1)**:** 630.

43. Gordon A, Forsingdal A, Klewe IV, Nielsen J, Didriksen M, Werge T *et al.* Transcriptomic networks implicate neuronal energetic abnormalities in three mouse models harboring autism and schizophrenia-associated mutations. *Mol Psychiatry* 2019.

44. Chen C, Cheng L, Grennan K, Pibiri F, Zhang C, Badner JA *et al.* Two gene co-expression modules differentiate psychotics and controls. *Mol Psychiatry* 2013; **18**(12)**:** 1308-1314.

45. Oron O, Getselter D, Shohat S, Reuveni E, Lukic I, Shifman S *et al.* Gene network analysis reveals a role for striatal glutamatergic receptors in dysregulated risk-assessment behavior of autism mouse models. *Transl Psychiatry* 2019; **9**(1)**:** 257.

46. Lin M, Pedrosa E, Hrabovsky A, Chen J, Puliafito BR, Gilbert SR *et al.* Integrative transcriptome network analysis of iPSC-derived neurons from schizophrenia and schizoaffective disorder patients with 22q11.2 deletion. *BMC Syst Biol* 2016; **10**(1)**:** 105.

47. Brennand K, Savas JN, Kim Y, Tran N, Simone A, Hashimoto-Torii K *et al.* Phenotypic differences in hiPSC NPCs derived from patients with schizophrenia. *Mol Psychiatry* 2015; **20**(3)**:** 361-368.

48. Steinberg J, Webber C. The roles of FMRP-regulated genes in autism spectrum disorder: single- and multiple-hit genetic etiologies. *Am J Hum Genet* 2013; **93**(5)**:** 825-839.

49. Jalbrzikowski M, Lazaro MT, Gao F, Huang A, Chow C, Geschwind DH *et al.* Transcriptome Profiling of Peripheral Blood in 22q11.2 Deletion Syndrome Reveals Functional Pathways Related to Psychosis and Autism Spectrum Disorder. *PLoS One* 2015; **10**(7)**:** e0132542.

50. Wang P, Zhao D, Lachman HM, Zheng D. Enriched expression of genes associated with autism spectrum disorders in human inhibitory neurons. *Transl Psychiatry* 2018; **8**(1)**:** 13.

51. Maschietto M, Tahira AC, Puga R, Lima L, Mariani D, Paulsen Bda S *et al.* Co-expression network of neural-differentiation genes shows specific pattern in schizophrenia. *BMC Med Genomics* 2015; **8:** 23.

52. Hafner H, Maurer K, Loffler W, Fatkenheuer B, an der Heiden W, Riecher-Rossler A *et al.* The epidemiology of early schizophrenia. Influence of age and gender on onset and early course. *Br J Psychiatry Suppl* 1994; (23)**:** 29-38.

53. Howard R, Rabins PV, Seeman MV, Jeste DV. Late-onset schizophrenia and very-late-onset schizophrenia-like psychosis: an international consensus. The International Late-Onset Schizophrenia Group. *Am J Psychiatry* 2000; **157**(2)**:** 172-178.

54. Kang HJ, Kawasawa YI, Cheng F, Zhu Y, Xu X, Li M *et al.* Spatio-temporal transcriptome of the human brain. *Nature* 2011; **478**(7370)**:** 483-489.

55. Selemon LD. Regionally diverse cortical pathology in schizophrenia: clues to the etiology of the disease. *Schizophr Bull* 2001; **27**(3)**:** 349-377.

56. Fuster JM. Frontal lobe and cognitive development. *J Neurocytol* 2002; **31**(3-5)**:** 373-385.

57. Consortium GT. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 2015; **348**(6235)**:** 648-660.

58. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 2012; **28**(6)**:** 882-883.

59. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 2007; **8**(1)**:** 118-127.

60. Wilcox R. *Introduction to Robust Estimation and Hypothesis Testing*. Academic Press: San Diego, 1997.

61. Song L, Langfelder P, Horvath S. Comparison of co-expression measures: mutual information, correlation, and model based indices. *BMC Bioinformatics* 2012; **13:** 328.

62. Barabasi AL, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nat Rev Genet* 2004; **5**(2)**:** 101-113.

63. Fu J, Keurentjes JJ, Bouwmeester H, America T, Verstappen FW, Ward JL *et al.* System-wide molecular evidence for phenotypic buffering in Arabidopsis. *Nat Genet* 2009; **41**(2)**:** 166-167.

64. Ouma WZ, Pogacar K, Grotewold E. Topological and statistical analyses of gene regulatory networks reveal unifying yet quantitatively different emergent properties. *PLoS Comput Biol* 2018; **14**(4)**:** e1006098.

65. Lachowiec J, Queitsch C, Kliebenstein DJ. Molecular mechanisms governing differential robustness of development and environmental responses in plants. *Ann Bot* 2016; **117**(5)**:** 795-809.

66. Yip AM, Horvath S. Gene network interconnectedness and the generalized topological overlap measure. *BMC Bioinformatics* 2007; **8:** 22.

67. Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabasi AL. Hierarchical organization of modularity in metabolic networks. *Science* 2002; **297**(5586)**:** 1551-1555.

68. Langfelder P, Zhang B, Horvath S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics* 2008; **24**(5)**:** 719-720.

69. Horvath S, Dong J. Geometric interpretation of gene coexpression network analysis. *PLoS Comput Biol* 2008; **4**(8)**:** e1000117.

70. BrainSpan Atlas of the Developing Human Brain. Technical white paper: transcriptome profiling by RNA sequencing and exon microarray (v.5). 2013.

71. Langfelder P, Luo R, Oldham MC, Horvath S. Is My Network Module Preserved and Reproducible? *Plos Computational Biology* 2011; **7**(1).

72. Reimand J, Isserlin R, Voisin V, Kucera M, Tannus-Lopes C, Rostamianfar A *et al.* Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nat Protoc* 2019; **14**(2)**:** 482-517.

73. Ala U, Piro RM, Grassi E, Damasco C, Silengo L, Oti M *et al.* Prediction of human disease genes by human-mouse conserved coexpression analysis. *PLoS Comput Biol* 2008; **4**(3)**:** e1000043.

74. Segal E, Friedman N, Koller D, Regev A. A module map showing conditional activity of expression modules in cancer. *Nat Genet* 2004; **36**(10)**:** 1090-1098.

75. Torkamani A, Dean B, Schork NJ, Thomas EA. Coexpression network analysis of neural tissue reveals perturbations in developmental processes in schizophrenia. *Genome Res* 2010; **20**(4)**:** 403-412.

76. van Dam S, Cordeiro R, Craig T, van Dam J, Wood SH, de Magalhaes JP. GeneFriends: an online co-expression analysis tool to identify novel gene targets for aging and complex diseases. *BMC Genomics* 2012; **13:** 535.

77. McCarroll SA, Murphy CT, Zou S, Pletcher SD, Chin CS, Jan YN *et al.* Comparing genomic expression patterns across species identifies shared transcriptional profile in aging. *Nat Genet* 2004; **36**(2)**:** 197-204.

78. Langfelder P, Horvath S. Eigengene networks for studying the relationships between co-expression modules. *BMC Syst Biol* 2007; **1:** 54.

79. Mokhtari R, Lachman HM. The Major Histocompatibility Complex (MHC) in Schizophrenia: A Review. *J Clin Cell Immunol* 2016; **7**(6).

80. Sekar A, Bialas AR, de Rivera H, Davis A, Hammond TR, Kamitaki N *et al.* Schizophrenia risk from complex variation of complement component 4. *Nature* 2016; **530**(7589)**:** 177-183.

81. Cooper TF, Morby AP, Gunn A, Schneider D. Effect of random and hub gene disruptions on environmental and mutational robustness in Escherichia coli. *BMC Genomics* 2006; **7:** 237.

82. Yu H, Braun P, Yildirim MA, Lemmens I, Venkatesan K, Sahalie J *et al.* High-quality binary protein interaction map of the yeast interactome network. *Science* 2008; **322**(5898)**:** 104-110.

83. Alexandru G, Graumann J, Smith GT, Kolawa NJ, Fang R, Deshaies RJ. UBXD7 binds multiple ubiquitin ligases and implicates p97 in HIF1alpha turnover. *Cell* 2008; **134**(5)**:** 804-816.

84. Tao S, Liu P, Luo G, Rojo de la Vega M, Chen H, Wu T *et al.* p97 Negatively Regulates NRF2 by Extracting Ubiquitylated NRF2 from the KEAP1-CUL3 E3 Complex. *Mol Cell Biol* 2017; **37**(8).

85. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A *et al.* Proteomics. Tissue-based map of the human proteome. *Science* 2015; **347**(6220)**:** 1260419.

86. Chan SY, Loscalzo J. The emerging paradigm of network medicine in the study of human disease. *Circ Res* 2012; **111**(3)**:** 359-374.

87. Fromer M, Roussos P, Sieberts SK, Johnson JS, Kavanagh DH, Perumal TM *et al.* Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat Neurosci* 2016; **19**(11)**:** 1442-1453.

88. Li J, Cai T, Jiang Y, Chen H, He X, Chen C *et al.* Genes with de novo mutations are shared by four neuropsychiatric disorders discovered from NPdenovo database. *Mol Psychiatry* 2016; **21**(2)**:** 290-297.

89. Meng Q, Wang K, Brunetti T, Xia Y, Jiao C, Dai R *et al.* The DGCR5 long noncoding RNA may regulate expression of several schizophrenia-related genes. *Sci Transl Med* 2018; **10**(472).

90. Psych EC, Akbarian S, Liu C, Knowles JA, Vaccarino FM, Farnham PJ *et al.* The PsychENCODE project. *Nat Neurosci* 2015; **18**(12)**:** 1707-1712.

91. Schizophrenia Working Group of the Psychiatric Genomics C. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014; **511**(7510)**:** 421-427.

92. Banerjee-Basu S, Packer A. SFARI Gene: an evolving database for the autism research community. *Dis Model Mech* 2010; **3**(3-4)**:** 133-135.

93. Abrahams BS, Arking DE, Campbell DB, Mefford HC, Morrow EM, Weiss LA *et al.* SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol Autism* 2013; **4**(1)**:** 36.

94. Deciphering Developmental Disorders S. Prevalence and architecture of de novo mutations in developmental disorders. *Nature* 2017; **542**(7642)**:** 433-438.

95. Shen L. GeneOverlap: Test and visualize gene overlaps. R package version 1.20.0 edn2019.

96. Chang D, Nalls MA, Hallgrimsdottir IB, Hunkapiller J, van der Brug M, Cai F *et al.* A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nat Genet* 2017; **49**(10)**:** 1511-1516.

97. Kunkle BW, Grenier-Boley B, Sims R, Bis JC, Damotte V, Naj AC *et al.* Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. *Nat Genet* 2019; **51**(3)**:** 414-430.

98. Momozawa Y, Dmitrieva J, Theatre E, Deffontaine V, Rahmouni S, Charloteaux B *et al.* IBD risk loci are enriched in multigenic regulatory modules encompassing putative causative genes. *Nat Commun* 2018; **9**(1)**:** 2427.

99. Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN *et al.* Meta-analysis of genome-wide association studies for height and body mass index in approximately 700000 individuals of European ancestry. *Hum Mol Genet* 2018; **27**(20)**:** 3641-3649.

100. Guan J, Cai JJ, Ji G, Sham PC. Commonality in dysregulated expression of gene sets in cortical brains of individuals with autism, schizophrenia, and bipolar disorder. *Transl Psychiatry* 2019; **9**(1)**:** 152.

101. Willatt L, Cox J, Barber J, Cabanas ED, Collins A, Donnai D *et al.* 3q29 microdeletion syndrome: clinical and molecular characterization of a new syndrome. *Am J Hum Genet* 2005; **77**(1)**:** 154-160.

102. Wachi S, Yoneda K, Wu R. Interactome-transcriptome analysis reveals the high centrality of genes differentially expressed in lung cancer tissues. *Bioinformatics* 2005; **21**(23)**:** 4205-4208.

103. Jonsson PF, Bates PA. Global topological features of cancer proteins in the human interactome. *Bioinformatics* 2006; **22**(18)**:** 2291-2297.

104. Xu J, Li Y. Discovering disease-genes by topological features in human protein-protein interaction network. *Bioinformatics* 2006; **22**(22)**:** 2800-2805.

105. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016; **536**(7616)**:** 285-291.

106. O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP *et al.* Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* 2012; **485**(7397)**:** 246-250.

Figure legends

Figure 1. Unbiased weighted gene co-expression network analysis (WGCNA) of the human transcriptome in the healthy adult prefrontal cortex (PFC). (a) A schematic of the data analysis workflow underlying WGCNA-derived predictions for functional interrogation of the 3q29Del interval. The reference dataset was obtained from the GTEx Project to construct a systems-level network representation of coordinated gene expression patterns across 107 non-pathological post-mortem samples collected from the Brodmann Area 9 (BA9) of male and female adults with no known history of psychiatric or neurological disorder. Modules of highly co-expressed genes were identified based on their topological overlap measure (TOM). The TOM between two genes is high if the genes have many overlapping network connections, yielding an interconnectedness measure that is proportional to the number of shared neighbors between pairs of genes. The resulting network was screened for modules harboring 3q29 interval genes (3q29 modules), which were then interrogated for biological function and hub genes. A test dataset obtained from the BrainSpan Project was used to validate the reproducibility of this network in an independent sample of 30 non-pathological post-mortem specimens collected from four sub-regions of the PFC from adult males and females with no known history of psychiatric or neurological disorder. These sub-regions are the OFC: orbital frontal cortex, DLPFC: dorsolateral PFC, VLPFC: ventrolateral PFC, and MPFC: medial PFC. **(b)** Sample-level dendrogram and trait heatmaps of the reference dataset. The dendrogram was yielded by hierarchical clustering of 107 GTEx samples using normalized, outlier-removed and residualized gene expression values for 18,410 protein-coding genes. Color bars represent trait heatmaps for sex, age-group (range = 20-79 years), death-classification based on the Hardy scale (range = 0-4), post-mortem interval (PMI) and batch id. The color intensity (from light yellow to red) is proportional to continuous or categorical values (in increasing order) of each variable. For sex, yellow and orange indicate female and male, respectively. Transcriptomic data were corrected for covariance mediated by these variables prior to network construction. Adjusted data reveal no distribution bias associated with the interrogated confounds in sample-level clustering patterns. **(c)** Determination of the soft-thresholding power (β) used for WGCNA. A β of 8 (black arrow) was identified as the lowest possible power yielding a degree distribution that results in approximate scale-free network topology (SFT R^2 fit index = 0.8; red line). **(d)** Clustering dendrogram and module assignments of genes, with dissimilarity based on TOM. 18,410 protein coding genes (leaves = genes) clustered into 19 final modules (bottom color bar), detected by the dynamic hybrid tree cut method. Modules with strongly correlated eigengenes

(Pearson's r > 0.8, $P < 0.05$) were amalgamated to eliminate spurious assignment of highly co-expressed genes into separate modules. Color bars reflect module assignments before and after the merging of close modules. **(e)** Composite *Zsummary* scores for module-preservation (how well-defined modules are in an independent test dataset) and module-quality (how well-defined modules are in repeated random splits of the reference dataset). Permutation tests were performed to adjust the observed preservation and quality statistics of each module for random chance by defining Z statistics. All modules (labeled by color) identified in the reference network were preserved (reproducible) in the test network (*Zsummary* > 2; blue line). 15/18 modules, including all 3q29 modules (red arrows), exhibited strong preservation (*Zsummary* > 10; green line). 3/18 modules exhibited moderate preservation (2 < *Zsummary* < 10). All modules demonstrated strong evidence for high quality (*Zsummary* > 10), confirming that the modules identified in the reference network were well-defined and non-random.

Figure 2. Network-based inference of the functional impact of 3q29Del on the adult human prefrontal cortex (PFC). (a) Hierarchical clustering of module eigengenes (ME) that summarize the 19 modules identified by WGCNA. The 21 protein-coding genes located in the recurrent 3q29Del locus were found to be distributed into 7 co-expression modules (3q29 modules; framed), The numbers next to dendrogram branches indicate the total number of 3q29 interval genes found in each 3q29 module. **(b)** Heatmap representing the strength of Pearson's correlation (*r*) between ME-pairs. The seven 3q29 modules (arrows) further clustered into three higher-level meta-modules, corresponding to squares of blue color (high positive correlation) along the diagonal, also detected as major dendrogram branches in (a). **(c)** Eigengene-based connectivity strength (kME; y-axis) of 3q29 interval genes (x-axis; in chromosomal order) within their respective modules. kME is defined as the Pearson's correlation between a query gene and a given ME. The line graph indicates the $-\log_{10}$(*P-value*) of the plotted correlation coefficients (z-axis); the asterisks above the graph indicate $P < 0.05$. kME > 0.8 ($P < 0.05$; dotted line) indicates hub (highly connected) gene status. *UBXN7* (red frame) was found to be the only hub gene (kME > 0.8, P = 8.33E-30) within the 3q29Del locus. *SMCO1 (*kME = 0.11, P = 0.25*), SLC51A (*kME = 0.17, P = 0.09) and *MFI2* (kME = 0.09, P = 0.35) were found to have non-significant kMEs (P > 0.05) for their respective modules, suggesting peripheral membership. Color indicates module label. **(d)** Top 10 biological pathways (Reactome database) significantly enriched in 3q29 modules (adjusted-$P < 0.05$; capped at $-\log_{10}$ (adjusted-P =

10)). The g:SCS method was used for multiple testing correction. The observed enrichment profile of the queried modules for known biological processes and pathways indicates that genes co-clustering in 3q29 modules show coordinated expression and converge upon overlapping biological functions, more than expected by chance. The functional associations of gene-sets comprising individual 3q29 modules were leveraged to infer likely molecular consequences of 3q29Del in the adult human PFC.

Figure 3. 3q29 interval genes form transcriptomic subnetworks enriched for known schizophrenia, autism and intellectual / developmental disability-risk genes. (**a**) Schematic of strategy to test the neuropsychiatric disease burden associated with top network neighbors of 3q29 interval genes and to refine a list of prioritized driver genes. To minimize false positives, 3q29 modules were reduced to strongly connected top neighbors (yellow nodes) of individual 3q29 genes, which were then screened for a significant overlap with known risk genes (red nodes) for schizophrenia (SZ), autism spectrum disorders (ASD), and intellectual / developmental disability (IDD), spanning known associations over a wide spectrum of allele frequencies. A top neighbor was defined as any node whose gene expression profile had a moderate-to-high pairwise correlation (Spearman's rho (ρ) ≥ 0.5, $P < 0.05$) with a 3q29 interval gene within the same module. By leveraging the guilt by association principle, the 3q29 interval genes that showed a significant enrichment of known SZ, ASD, and/or IDD risk genes among their respective top neighbors were prioritized as likely drivers of the neurodevelopmental and psychiatric consequences of 3q29Del, along with their SZ, ASD, and/or IDD-related top neighbors from significant overlap tests. (**b**) Adjusted p-values from hypergeometric tests identifying the significance of the overlap between top neighbors of individual 3q29 genes and known risk genes for SZ, ASD, and IDD. Risk gene-sets for three traits with no known association to the 3q29Del syndrome were also tested for over-representation as negative controls. Common variants associated with height were included as another negative control to rule out a potential bias introduced by gene-set size. Nine protein-coding genes from the 3q29 interval formed transcriptomic subnetworks that are significantly enriched for known SZ, ASD and/or IDD risk genes (orange highlight, adjusted-$P < 0.05$). The proportion of hypergeometric tests with significant overrepresentation of SZ, ASD, and IDD gene-sets (19/96) was found to be an order of magnitude larger than that of the negative control tests (2/64), demonstrating the high specificity of the identified enrichment patterns for reported 3q29Del-associated phenotypes.

Figure 4. Network of prioritized drivers predicted to contribute to the neuropsychiatric sequelae of 3q29Del. (a) Nine protein-coding genes from the 3q29 interval formed top-neighbor-based transcriptomic subnetworks that were significantly enriched for known schizophrenia (SZ), autism spectrum disorder (ASD) and intellectual / developmental disability (IDD)-risk genes (adjusted-$P < 0.05$). Black and red nodes illustrated in this network diagram represent these 9 3q29 genes and their 284 top neighbors with known SZ, ASD and/or IDD-association, respectively. The union of these prioritized genes constitute 293 genetic drivers predicted to contribute to the neurodevelopmental and psychiatric phenotypes of 3q29Del. The color of network edges that connect node-pairs represents module assignment. **(b)** Top 20 biological processes and pathways with significant enrichment among prioritized drivers (adjusted-$P < 0.05$). The identified Gene Ontology biological processes (GO: BP) and Reactome and KEGG biological pathways point to key mechanisms through which select genes within the 3q29Del locus and their likely partners outside the interval are predicted to influence susceptibility to SZ, ASD, and IDD. **(c)** Organization of all statistically significant biological processes enriched in prioritized drivers into a network of related functional annotation categories. GO:BP terms are connected if they have a high overlap (share many genes); edge width represents magnitude of the overlap**.**

Figure 1

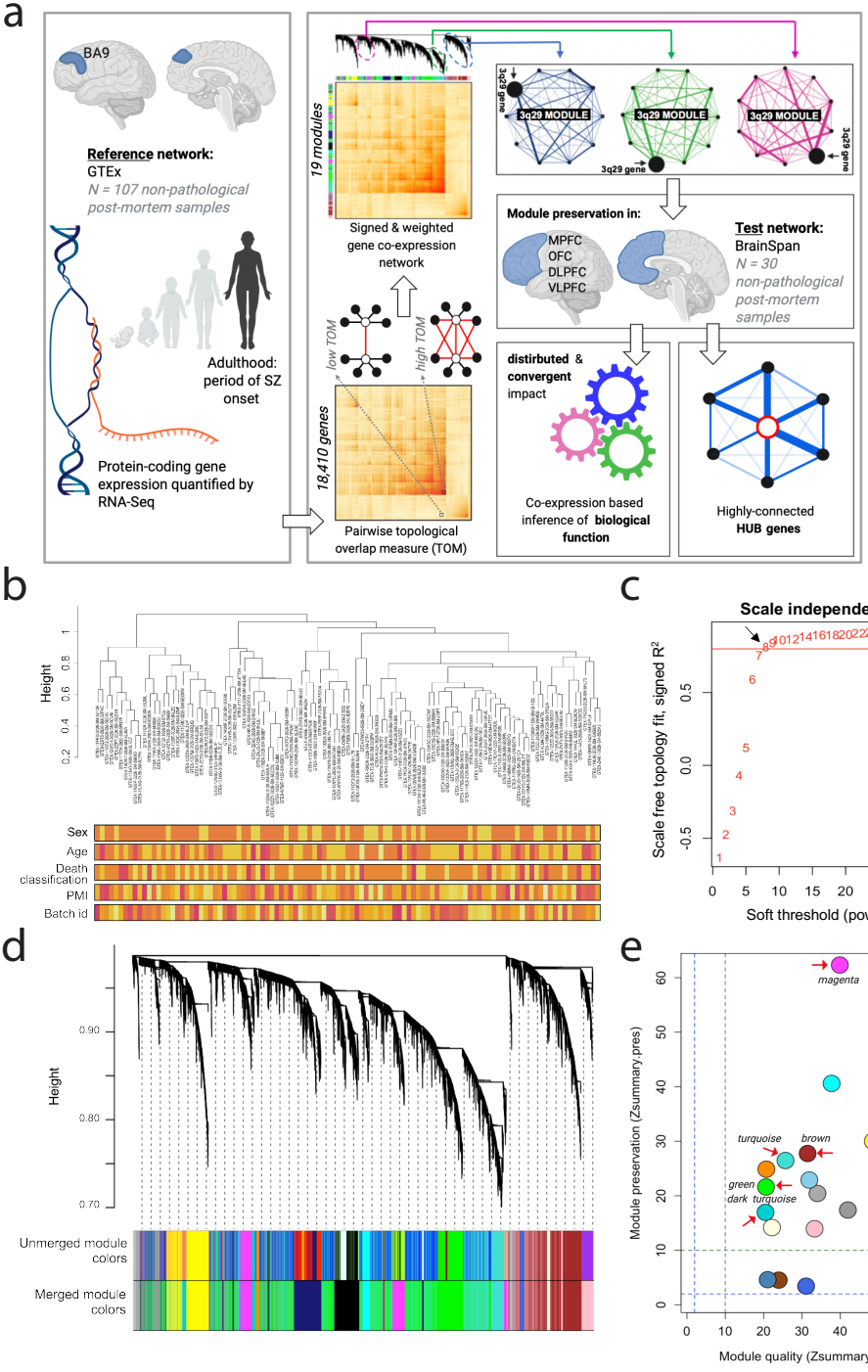

a
BA9
Reference network:
GTEx
N = 107 non-pathological post-mortem samples
Adulthood: period of SZ onset
Protein-coding gene expression quantified by RNA-Seq
19 modules
Signed & weighted gene co-expression network
low TOM
high TOM
18,410 genes
Pairwise topological overlap measure (TOM)
3q29 gene
3q29 MODULE
3q29 MODULE
3q29 MODULE
3q29 gene
3q29 gene
Module preservation in:
MPFC
OFC
DLPFC
VLPFC
Test network:
BrainSpan
N = 30 non-pathological post-mortem samples
distirbuted & convergent impact
Co-expression based inference of biological function
Highly-connected HUB genes
b
Height
Sex
Age
Death classification
PMI
Batch id
c
Scale independence
Scale free topology fit, signed R²
Soft threshold (power)
d
Height
Unmerged module colors
Merged module colors
e
magenta
midnight blue
black
turquoise
brown
green
dark turquoise
Module preservation (Zsummary.pres)
Module quality (Zsummary.qual)

Figure 2

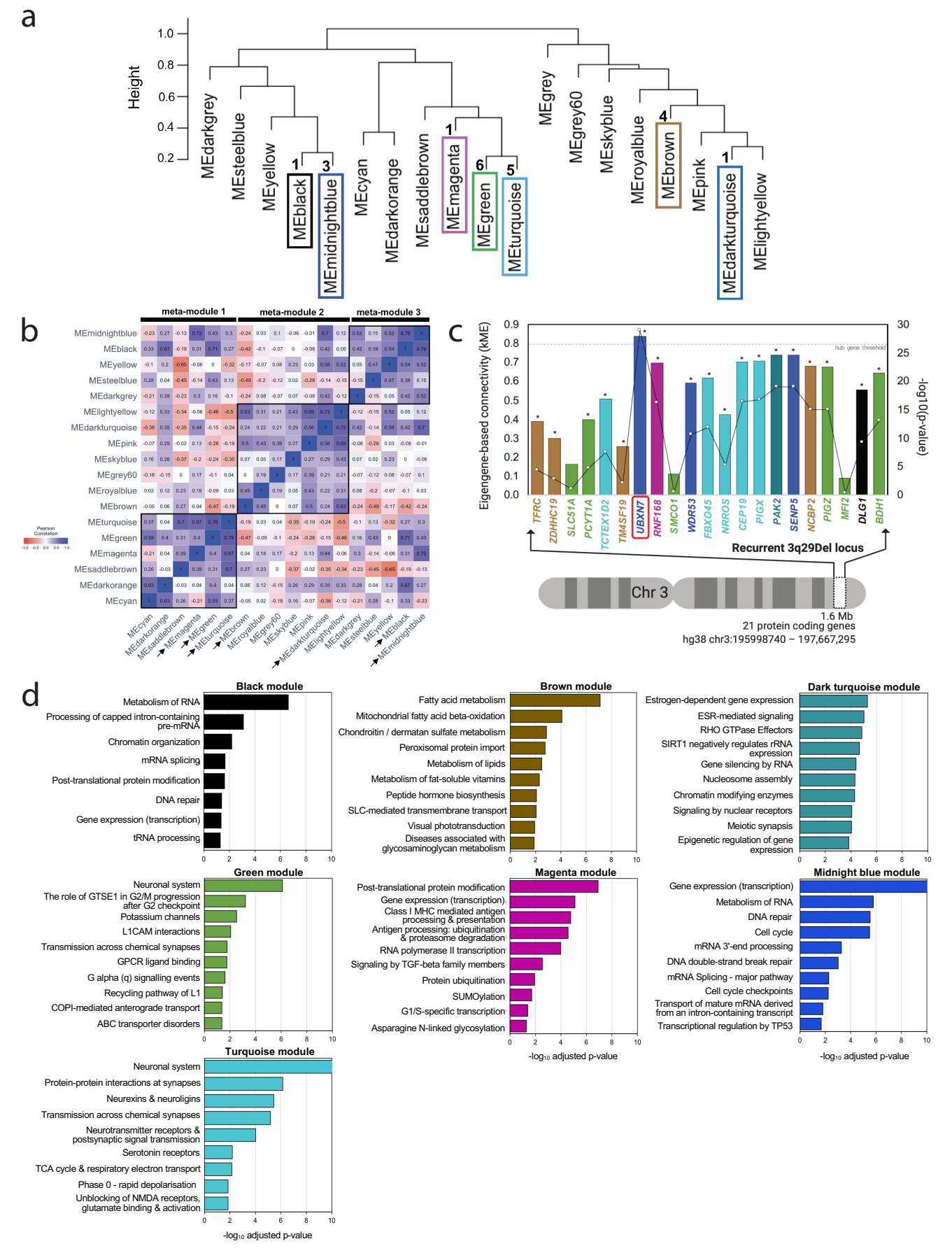
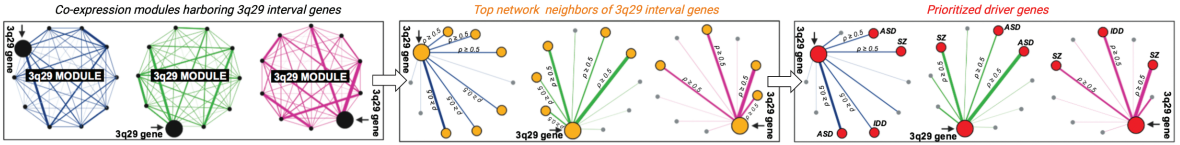

a
Height
1.0
0.8
0.6
0.4
0.2
MEdarkgrey
MEsteelblue
MEyellow
MEblack
MEmidnightblue
MEcyan
MEdarkorange
MEsaddlebrown
MEmagenta
MEgreen
MEturquoise
MEgrey
MEgrey60
MEskyblue
MEroyalblue
MEbrown
MEpink
MEdarkturquoise
MElightyellow
b
meta-module 1
meta-module 2
meta-module 3
Pearson Correlation
c
Eigengene-based connectivity (kME)
-log10(p-value)
hub gene threshold
TFRC
ZDHHC19
SLC51A
PCYT1A
TCTEX1D2
TM4SF19
UBXN7
RNF168
SMCO1
WDR53
FBXO45
NRROS
CEP19
PIGX
PAK2
SENP5
NCBP2
PIGZ
MFI2
DLG1
BDH1
Recurrent 3q29Del locus
Chr 3
1.6 Mb
21 protein coding genes
hg38 chr3:195998740 – 197,667,295
d
Black module
Metabolism of RNA
Processing of capped intron-containing pre-mRNA
Chromatin organization
mRNA splicing
Post-translational protein modification
DNA repair
Gene expression (transcription)
tRNA processing
Brown module
Fatty acid metabolism
Mitochondrial fatty acid beta-oxidation
Chondroitin / dermatan sulfate metabolism
Peroxisomal protein import
Metabolism of lipids
Metabolism of fat-soluble vitamins
Peptide hormone biosynthesis
SLC-mediated transmembrane transport
Visual phototransduction
Diseases associated with glycosaminoglycan metabolism
Dark turquoise module
Estrogen-dependent gene expression
ESR-mediated signaling
RHO GTPase Effectors
SIRT1 negatively regulates rRNA expression
Gene silencing by RNA
Nucleosome assembly
Chromatin modifying enzymes
Signaling by nuclear receptors
Meiotic synapsis
Epigenetic regulation of gene expression
Green module
Neuronal system
The role of GTSE1 in G2/M progression after G2 checkpoint
Potassium channels
L1CAM interactions
Transmission across chemical synapses
GPCR ligand binding
G alpha (q) signalling events
Recycling pathway of L1
COPI-mediated anterograde transport
ABC transporter disorders
Magenta module
Post-translational protein modification
Gene expression (transcription)
Class I MHC mediated antigen processing & presentation
Antigen processing: ubiquitination & proteasome degradation
RNA polymerase II transcription
Signaling by TGF-beta family members
Protein ubiquitination
SUMOylation
G1/S-specific transcription
Asparagine N-linked glycosylation
Midnight blue module
Gene expression (transcription)
Metabolism of RNA
DNA repair
Cell cycle
mRNA 3'-end processing
DNA double-strand break repair
mRNA Splicing - major pathway
Cell cycle checkpoints
Transport of mature mRNA derived from an intron-containing transcript
Transcriptional regulation by TP53
Turquoise module
Neuronal system
Protein-protein interactions at synapses
Neurexins & neuroligins
Transmission across chemical synapses
Neurotransmitter receptors & postsynaptic signal transmission
Serotonin receptors
TCA cycle & respiratory electron transport
Phase 0 - rapid depolarisation
Unblocking of NMDA receptors, glutamate binding & activation
-log10 adjusted p-value

Figure 3

a



b

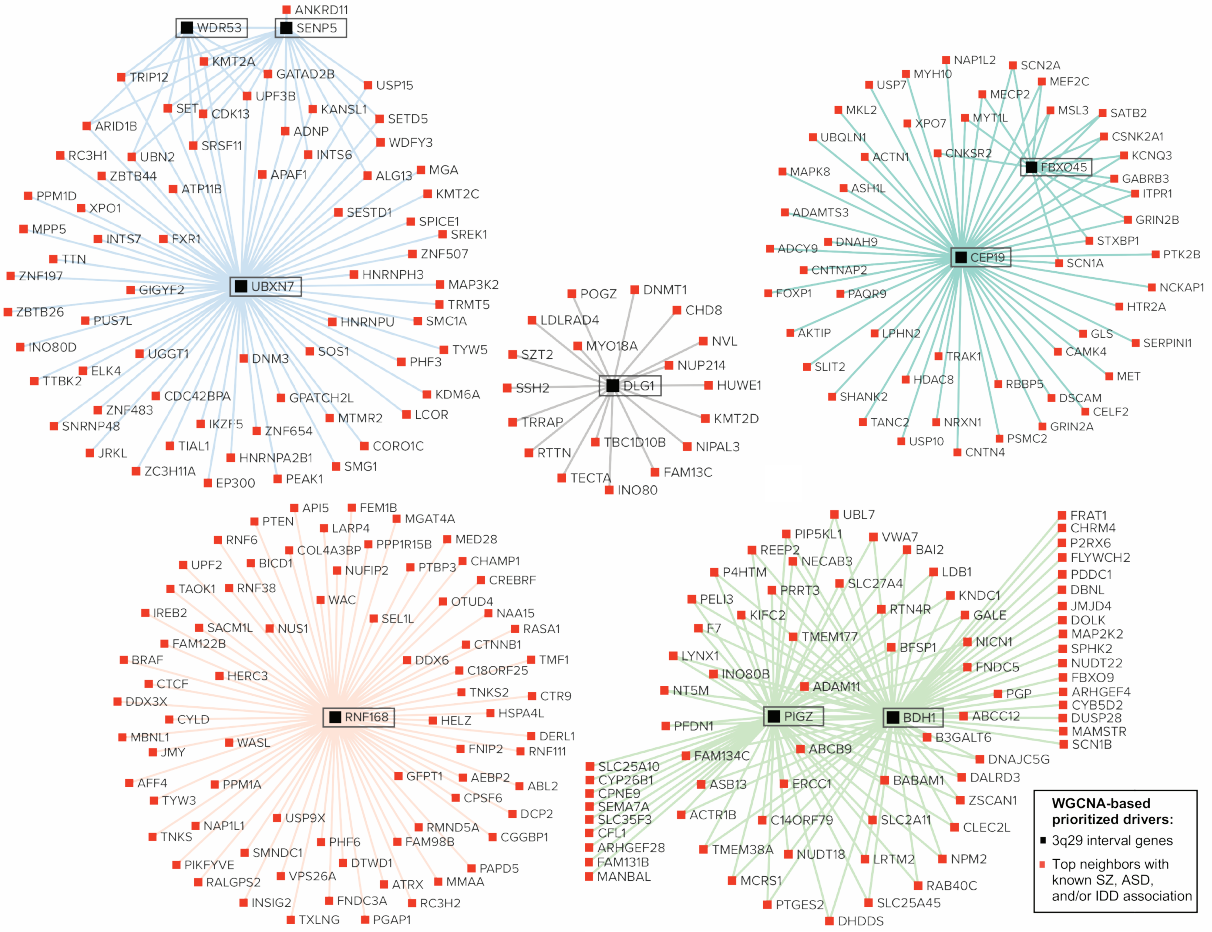
	BDH1	CEP19	DLG1	FBXO45	NCBP2	PAK2	PCYT1A	PIGX	RNF168	SENP5	TCTEX1D2	TFRC	UBXN7	WDR53	PIGZ	NRROS
IDD risk *(McRae et al., 2017)* gene-set size = 93	0.568	0.002*	0.085	0.014*	1.000	0.673	0.699	0.080	0.011*	0.005*	0.677	1.000	< 0.000*	0.015*	0.150	1.000
ASD risk *(SFARI categories 1 & 2)* gene set size = 86	0.568	0.002*	0.128	0.070	1.000	0.754	0.699	0.258	0.160	0.001*	0.991	1.000	< 0.000*	0.015*	0.849	1.000
Schizophrenia risk *(BrainGVEX case-control)* gene-set size = 651	0.769	0.665	1.000	0.940	1.000	0.673	1.000	0.635	1.000	1.000	0.420	1.000	0.807	0.875	0.849	1.000
Schizophrenia risk *(CommonMind case-control)* gene-set size = 636	< 0.000*	0.966	1.000	0.953	1.000	0.673	1.000	0.999	< 0.000*	0.060	0.991	1.000	0.004*	0.283	0.001*	1.000
Schizophrenia risk *(GWAS identified common risk variants)* gene-set size = 340	0.283	0.158	0.362	0.776	1.000	0.754	0.699	0.258	1.000	1.000	0.677	0.590	0.771	0.822	0.253	1.000
Schizophrenia risk *(Exonic de novo mutations)* gene-set size = 290	0.769	0.033*	< 0.000*	0.147	1.000	0.754	0.699	0.143	0.640	1.000	0.991	1.000	0.738	0.919	0.868	1.000
Alzheimer's disease risk *(GWAS identified common risk variants)* gene-set size = 25	0.769	0.665	1.000	0.940	1.000	0.673	1.000	0.999	1.000	1.000	1.000	0.590	0.807	1.000	0.849	1.000
Parkinson's disease risk *(GWAS identified common risk variants)* gene-set size = 67	0.568	0.514	0.493	0.278	1.000	0.754	0.699	0.143	0.640	1.000	1.000	1.000	0.807	1.000	0.849	0.020*
IBD risk *(GWAS identified common risk variants)* gene-set size = 98	0.769	0.665	1.000	0.912	1.000	1.000	0.699	0.679	1.000	1.000	0.991	1.000	0.738	1.000	0.849	1.000
Height *(GWAS identified common associated variants)* gene-set size = 479	0.175	0.771	0.128	0.940	1.000	0.702	0.699	0.999	0.640	0.024*	0.991	1.000	0.738	0.283	0.435	1.000

19 / 96 tests

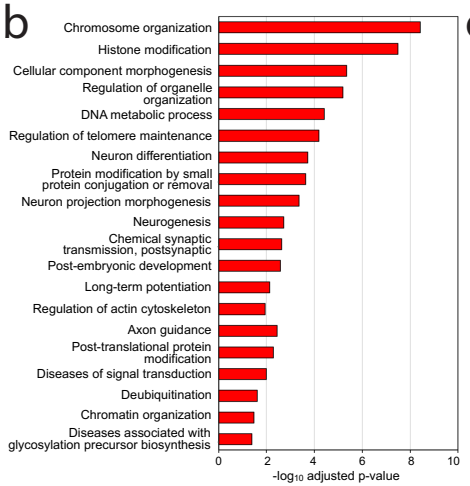
2 / 64 tests

Figure 4

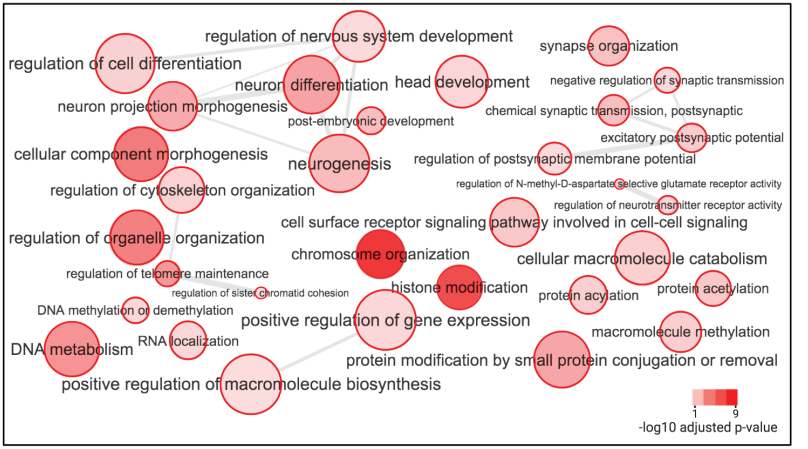
a


WDR53
SENP5
UBXN7
FBXO45
CEP19
DLG1
RNF168
PIGZ
BDH1
WGCNA-based prioritized drivers:
3q29 interval genes
Top neighbors with known SZ, ASD, and/or IDD association

b


Chromosome organization
Histone modification
Cellular component morphogenesis
Regulation of organelle organization
DNA metabolic process
Regulation of telomere maintenance
Neuron differentiation
Protein modification by small protein conjugation or removal
Neuron projection morphogenesis
Neurogenesis
Chemical synaptic transmission, postsynaptic
Post-embryonic development
Long-term potentiation
Regulation of actin cytoskeleton
Axon guidance
Post-translational protein modification
Diseases of signal transduction
Deubiquitination
Chromatin organization
Diseases associated with glycosylation precursor biosynthesis
0
2
4
6
8
10
-log10 adjusted p-value

c


regulation of nervous system development
synapse organization
regulation of cell differentiation
neuron differentiation
head development
negative regulation of synaptic transmission
neuron projection morphogenesis
post-embryonic development
chemical synaptic transmission, postsynaptic
excitatory postsynaptic potential
cellular component morphogenesis
neurogenesis
regulation of postsynaptic membrane potential
regulation of N-methyl-D-aspartate selective glutamate receptor activity
regulation of cytoskeleton organization
regulation of neurotransmitter receptor activity
cell surface receptor signaling pathway involved in cell-cell signaling
regulation of organelle organization
chromosome organization
cellular macromolecule catabolism
regulation of telomere maintenance
regulation of sister chromatid cohesion
histone modification
protein acylation
protein acetylation
DNA methylation or demethylation
positive regulation of gene expression
macromolecule methylation
DNA metabolism
RNA localization
protein modification by small protein conjugation or removal
positive regulation of macromolecule biosynthesis
1
9
-log10 adjusted p-value