



Identification and prioritization of gene sets associated with schizophrenia risk by co-expression network analysis in human brain

Eugenia Radulescu¹ · Andrew E. Jaffe^{1,2,3,4} · Richard E. Straub¹ · Qiang Chen¹ · Joo Heon Shin¹ · Thomas M. Hyde^{1,5,6} · Joel E Kleinman^{1,5} · Daniel R. Weinberger^{1,6,7,8}

Received: 7 April 2018 / Revised: 7 August 2018 / Accepted: 29 October 2018 / Published online: 26 November 2018
© Springer Nature Limited 2018

Abstract

Schizophrenia polygenic risk is plausibly manifested by complex transcriptional dysregulation in the brain, involving networks of co-expressed and functionally related genes. The main purpose of this study was to identify and prioritize co-expressed gene sets in a hierarchical manner, based on the strength of the relationships with clinical diagnosis and with polygenic risk for schizophrenia. Weighted Gene Co-expression Network Analysis (WGCNA) was applied to RNA-quality-adjusted DLPFC RNA-Seq data from the LIBD Postmortem Human Brain Repository (90 controls, 74 schizophrenia cases; all Caucasians) to construct co-expression networks and detect “modules” of co-expressed genes. After multiple internal and external validation procedures, modules of selected interest were tested for enrichment in biological ontologies, for association with schizophrenia polygenic risk scores (PRSs) and with diagnosis, and also for enrichment in genes within the significant GWAS loci reported by the Psychiatric Genomic Consortium (PGC2). The association between schizophrenia genetic signals and modules of co-expression converged on one module showing not only a significant association with both diagnosis and PRS but also significant overlap with 36 PGC2 loci genes, deemed the strongest candidates for drug targets. This module contained many genes involved in synaptic signaling and neuroplasticity. Fifty-three PGC2 genes were in modules associated only with diagnosis and 59 in modules unrelated to diagnosis or PRS. Our study highlights complex relationships between gene co-expression networks in the brain and clinical state and polygenic risk for SCZ and provides a strategy for using this information in selecting and prioritizing potentially targetable gene sets for therapeutic drug development.

Introduction

Schizophrenia (SCZ), the rubric for the most severe psychiatric syndromes in the psychoses spectrum, continues to

be an enigmatic human condition. Large-scale genomic [1], transcriptomic [2], and epigenomic [3] studies have started to reveal not only the multifactorial biological substrate of SCZ but also new challenges especially from the clinical translational perspective. While it has long been clear that genes and their proteins do not act in isolation to build brain circuitries and maintain their functionality [4], the subtle

Electronic supplementary material The online version of this article (<https://doi.org/10.1038/s41380-018-0304-1>) contains supplementary material, which is available to authorized users.

✉ Daniel R. Weinberger
drweinberger@libd.org

- ¹ Lieber Institute for Brain Development, Johns Hopkins Medical Campus, Baltimore, MD 21205, USA
- ² Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA
- ³ Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA
- ⁴ Center for Computational Biology, Johns Hopkins University,

Baltimore, MD 21205, USA

- ⁵ Department of Neurology, Johns Hopkins School of Medicine, Baltimore, MD 21205, USA
- ⁶ Department of Neuroscience, Johns Hopkins School of Medicine, Baltimore, MD 21205, USA
- ⁷ Department of Psychiatry and Behavioral Sciences, Johns Hopkins School of Medicine, Baltimore, MD 21205, USA
- ⁸ McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD 21205, USA

and complicated genetic and environmental interactions that influence the development and function of the brain remain largely a mystery. Consequently, translating new discoveries into efficient therapies is probably the most difficult and frustrating endeavor in contemporary psychiatry.

Taking into account that transcriptional regulation plays a major role in neurodevelopment and neuronal activity [5], a promising approach to study genetic interactions and their implications in risk for SCZ is gene co-expression analysis. The logic behind this approach is that functional gene assemblies probably require a co-regulated transcriptional profile. Consequently, networks of co-expressed genes from post-mortem brain gene expression data could mirror such functional gene assemblies. A popular bioinformatics tool for constructing and studying gene co-expression networks is Weighted Gene Co-Expression Network Analysis (WGCNA) [6]. This approach has been used to characterize patterns of co-expression in the normal brain [7], in autism spectrum disorders [8], in SCZ (human and animal modeling studies) [9–12], and across mental disorders [13]. Importantly, co-expression network analysis has also been used to understand the partitioning of polygenic SCZ risk in the brain transcriptome.

For example, in a recently published study, Fromer et al. took a stepwise approach of combining transcriptomics and genetics techniques, including gene co-expression analysis, and identified a sub-network of co-expressed genes with roles in synaptic transmission that was highly enriched for SCZ genetic associations [14]. In another study, Pergola et al. used a multi-modal approach including co-expression analysis and found that a co-expression profile including *DRD2* and other SCZ risk genes was associated with clinical intermediate phenotypes of SCZ [15]. While these earlier studies provide potentially important evidence for gene network associations in SCZ, they focused on rather selected aspects of genetic risk integration with co-expression in the brain transcriptome.

In the present study, we have taken a more global and stepwise approach to characterize and prioritize networks of co-expressed genes in the brain and their association with the clinical state of SCZ and with polygenic risk of SCZ. We first perform a systematic characterization of gene co-expression in postmortem dorsolateral prefrontal cortex (DLPFC) tissue from controls (CTRL) and patients with SCZ. We also critically address the potential influence of RNA quality on network association, which has not been specifically considered in earlier work. This is an important potential confounder as co-expression may be subsumed by co-degradation. Most importantly, we then develop a pipeline to identify the gradual convergence between the DLPFC co-transcriptome and SCZ genetic signals in order to select and prioritize gene sets as potential therapeutic drug targets for SCZ.

Materials and methods

General pipeline of data processing

Human postmortem tissue

We used postmortem human brain tissue from the Lieber Institute for Brain Development (LIBD) Human Brain Repository for testing and the CommonMind Consortium (CMC) brain collection for validation [2, 16]. DLPFC gray matter tissue from both collections was used for RNA extraction. The protocol of brain acquisition (location, legal authorizations, informed consent, clinical review/diagnosis), pre-processing, and tissue quality check is detailed elsewhere [2, 16]. Briefly, the samples selected included tissue from adults (age of death = 16–80 years), healthy CTRL ($N = 90$; male/female (M/F) = 72/18) and patients with SCZ ($N = 74$; M/F = 52/22) donors, Caucasians (CAUC), all with RNA integrity number (RIN) ≥ 7 . Chi-squared test and general linear model (t test) were applied to test for potential differences in sex distribution and age of death differences between CTRL and SCZ samples, respectively.

RNA sequencing (RNA-Seq) data processing

For the LIBD dataset, RNA was extracted from DLPFC gray matter (BA9/46) and RNA-Seq libraries were constructed with the Illumina poly A+ Kit; the resulting sequencing reads were aligned to the human genome (UCSC hg 19 build) with TopHat (v2.0.4) [17]; following alignment, the expression for genes and exons was summarized in counts based on Ensembl v75 [18], then converted to RPKM (reads per kilobase of transcript per million mapped reads) and normalized by \log_2+1 transformation. Normalized expression data from all samples were adjusted to remove unwanted variance potentially explained by RNA quality (i.e., technical or biological artifacts) (details below and in supplementary material). All analyses were performed on expression data quantified at the gene level; consequently only genes with sufficient abundance (median RPKM ≥ 0.1 across all samples) were retained for analysis. This selection yielded 22,945 genes for the LIBD dataset and 27,779 genes for the CMC data.

Processing CMC gene expression data is described elsewhere [2]. CMC library preparation utilized the Illumina Ribozero Kit. We downloaded CMC BAM files from Synapse (<https://www.synapse.org/>); the BAM files were aligned with TopHat2 and expression of genes was quantified in counts relative to Ensembl v75 and subsequently converted to RPKM and normalized by \log_2+1 transformation.

Genotype data processing and polygenic risk score (PRS) calculation

DNA extracted from cerebellar tissue was processed and normalized as previously described [2, 16]. Genotype imputation and quality check was performed with IMPUTE2 [19] and Shape-IT [20]. Only common single-nucleotide polymorphisms (SNPs) in Hardy–Weinberg equilibrium (at $p > 1e-6$) with minor allele frequency $>5\%$ were retained for analysis [2].

PRSs were calculated for each sample by summing the imputation probability of the reference allele of the clumped SNPs using PLINK v1.07 (<http://pngu.mgh.harvard.edu/purcell/plink/>) [21] and weighted by the natural log of the odds ratio from Psychiatric Genomic Consortium (PGC2) genome-wide association study (GWAS) results [1]. We used PRS based on 10 clinical SNP sets, corresponding to GWAS p values of $p = 5e-8$ (PRS1), $p = 1e-6$ (PRS2), $p = 1e-4$ (PRS3), $p = 0.001$ (PRS4), $p = 0.01$ (PRS5), $p = 0.05$ (PRS6), $p = 0.1$ (PRS7), $p = 0.2$ (PRS8), $p = 0.5$ (PRS9), and $p = 1$ (PRS10) [22].

Selection and prioritization of gene sets associated with risk for SCZ based on gene co-expression analysis

A schematic representation of the pipeline for identification and prioritization of gene sets associated with diagnosis and with genetic risk for SCZ in co-expression networks is provided in Fig. 1a. Processing the gene expression data to remove unwanted variability associated with sequencing and tissue confounders is presented in Fig. 1b and supplementary material. The adjusted/“cleaned” expression data were then used as input for weighted gene co-expression analysis performed with functions implemented in the WGCNA package [23, 24].

WGCNA is a method to calculate correlated gene expression in the form of adjacency matrices (networks). In essence, co-expression networks are pairwise gene expression correlations raised to a power (beta; β) in order to maximize the differences between correlations (i.e., to give “weight” to strong correlations at the expense of weak correlations). Selection of β power is based on the “scale-free network topology” criterion that is considered characteristic of “biological” networks (i.e., the degree/connectivity distribution follows a power law). Co-expression network construction is followed by the detection of sub-networks/“modules” (modules consist of “nodes,” which are represented by the genes quantified in mRNA abundance that share similar expression and patterns of connectivity) [25].

Details about the next WGCNA steps are presented in Fig. 1b and supplementary methods. In brief:

1. The co-expression network based on empirical Bayes-adjusted expression data (see below) was created for the combined sample of CTRL+SCZ with WGCNA. Modules of co-expression were then detected by hierarchical clustering and summarized by the first principal component (i.e. the “eigengene”) for each module (module eigengene (ME)) [7, 23, 24]. The MEs can be regarded as expression profiles that best characterize the gene correlations within modules. Biological inference can be drawn from the genes in the constructed modules by using gene set enrichment analyses and by correlating MEs with biological covariates. Likewise, intramodular analysis can be used to assess the degree of connectivity for the genes within modules and the gene-wise significance relative to association with traits of interest or diagnosis. The major advantage of MEs’ is in dimensionality data reduction, which makes them particularly suitable for correlation with traits of interest by eliminating the problem of multiple comparison corrections.

2. Internal validation

Additional diagnostics of the co-expression modules was also performed (Fig. 1b). Previous studies in our group have highlighted the importance of RNA quality in gene expression analysis, especially in detection of differential diagnosis effects [26]. However, selecting the best method for controlling for these effects is not straightforward and relies much on the type of analysis performed. Methods using empirical Bayes (empB) moderated regression (e.g., ComBat [27] or *empiricalBayesLM* function used by us in this study) are stringent but they perform correction based only on “known” sources of unwanted variance. One possibility to remove variability from “unknown” sources of technical error can be performed by modeling “unknown” (latent) variables with “quality” surrogate variables (qSVs) [26]. This approach is based on data from a human brain RNA degradation experiment. Here, in addition to our main network construction with “cleaned” data after removing unwanted variance for “known” sources using empB, we performed a complementary post hoc analysis (Fig. 1b) with the purpose of confirming the modular pattern of the main network after adjusting also for unknown sources of variance represented by qSVs calculated from the “degradation matrix” of the entire sample (90 CTRL + 74 SCZ) described in Jaffe et al. [26].

3. External validation

- a. Module preservation analysis was performed in the CMC expression data selected and treated similarly with LIBD data (same parameters for RIN, age, and application of empB adjusting for the unwanted variation) (Fig. 1b). This approach strengthened the comparability between the two datasets, necessary for testing the preservation. Moreover, module preservation in the context of the different library protocols in the LIBD and CMC datasets further

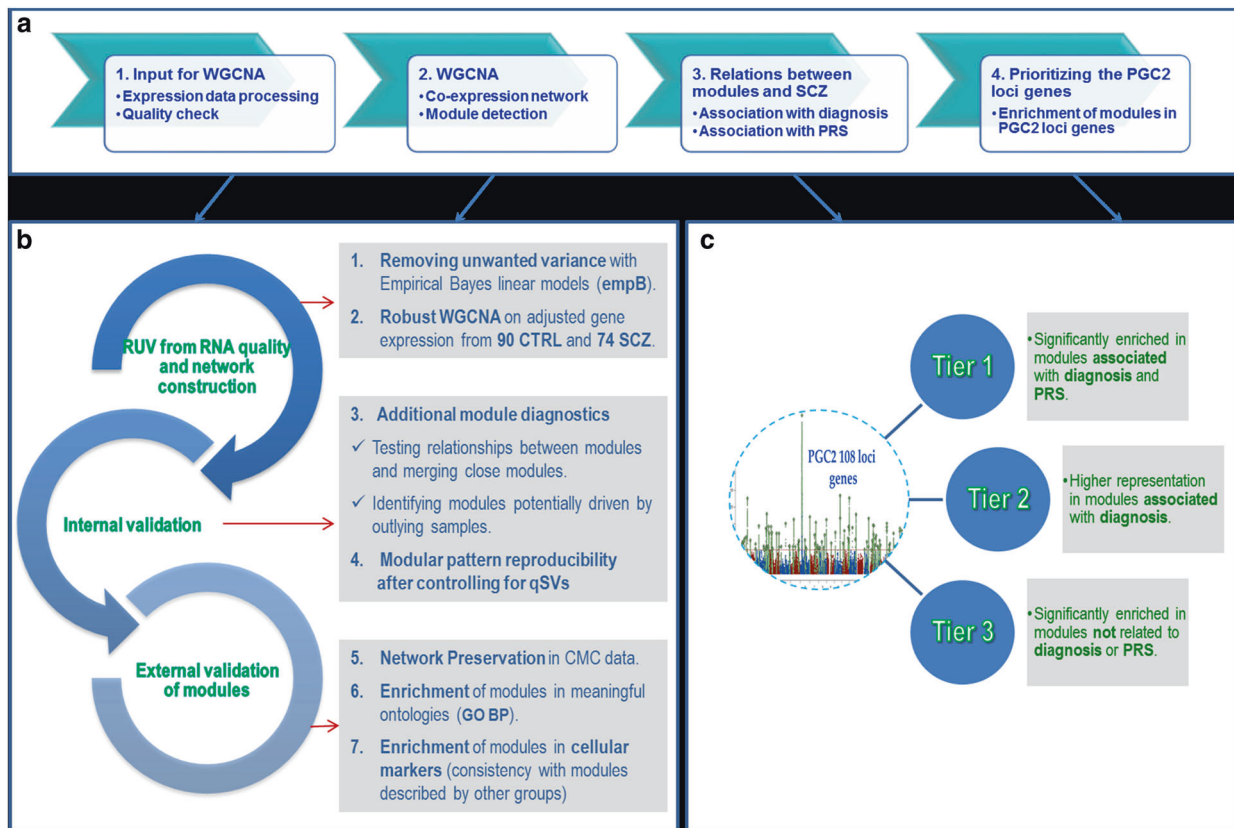


Fig. 1 Analysis pipeline and criteria for selection and prioritization of gene sets associated with genetic risk of schizophrenia. **a** Summary of the analysis pipeline. **b** (1) Prior to co-expression network analysis, expression data normalized by $(\log_2 + 1)$ transformation was adjusted to control for unwanted variation related to sequencing, tissue artifacts, and population sub-structure (i.e., RIN, postmortem interval (PMI), exonic mapping rate, alignment rate, and 10 genomic principal components—snPCs) by using the *empiricalBayesLM* function (WGCNA, version 1.61 [23]). (2) Co-expression network was constructed with WGCNA embedded routines (*blockwiseModules*) by applying the following parameters and procedures: bi-weight mid-correlation with “signed” TOM to allow for potentially non-linear correlations between genes [24]; β power = 12 selected with the “Soft Thresholding” function and applied to the gene correlations prior to network construction; modules of co-expression detection with hierarchical clustering using a measure of dissimilarity (the topological overlap); (3) Modules of co-expression were inspected through heatmaps of module specific gene expression across samples. In these heatmaps, well-defined modules are considered those displaying characteristic band structures, whereas the corresponding genes are highly correlated across samples; (4) Post hoc WGCNA limited to 9239 genes organized in nine modules of interest after the primary network construction was performed. (5) External validation by using the *modulePreservation* function in WGCNA [23, 24], which

supports their validity (details in supplementary methods and [28]).

b. Further external validation was represented by enrichment in putatively meaningful ontologies and comparisons with modules previously identified by other groups. We tested for enrichment in Gene Ontology–biological processes (GO-BP) with functions implemented in

computes network-based, pair-wise module preservation statistics by taking as input adjacency matrices in a reference set (the LIBD networks) and a test set (CMC; $N = 210$; CTRL = 102; SCZ = 108; female_{CTRL} = 26; female_{SCZ} = 32); relevant statistics output: measures of preservation for density and connectivity summarized as individual composite Z scores (Z_{summary}). By convention: $Z_{\text{summary}} = 0-2$ means no preservation, $Z_{\text{summary}} = 2-10$ means weak preservation and $Z_{\text{summary}} > 10$ means strong preservation [28]. (6) Enrichment in meaningful biological ontologies was tested with *enrichGO* function from the R package *clusterProfiler* [29]: a hypergeometric test was applied to test for overrepresentation of gene sets (i.e., module-specific genes) in relevant GO-BP; statistical threshold for significance was set at default values $p = 0.01$ and $q = 0.05$ with the Benjamini–Hochberg (BH) multiple comparisons correction method, using as background only the genes from the network construction, annotated by Entrez Gene IDs (which are largely protein-coding) (annotation performed with the Bioconductor package *org.Hs.eg.db* [https://www.bioconductor.org/packages/release/data/annotation/html/org.Hs.eg.db.html]); (7) Enrichment in cellular markers was tested by using the hypergeometric test and gene lists embedded in the *userListEnrichment* function [7, 24, 30]. **c** Modules’ enrichment in PGC2 loci genes was also tested with the hypergeometric test (*userListEnrichment* function [24]); WGCNA intramodular analysis measures and functions were used to test associations between modules and polygenic scores and diagnosis [24]

clusterProfiler R package [29, https://www.bioconductor.org/packages/release/data/annotation/html/org.Hs.eg.db.html] (Fig. 1b).

c. As a further approach to module validation, we sought to see whether our modules are similar to those reported in the study by Oldham et al., which provided the initial characterization of the gene co-expression relationships in

the human brain [7]. For this analysis, we tested for enrichment of our modules in cellular markers based on the cortex modules reported in Oldham et al. [7] and additionally on a transcriptome database including markers for neurons, astrocytes and oligodendrocytes by Cahoy et al. [30] (Fig. 1b). Likewise, we reasoned that, notwithstanding methodological differences, our modules should bear some overlap with modules of cortical gene co-expression reported in recent studies based on samples from donors with a history of neuropsychiatric disorders, including SCZ. To test this assumption, we compared the lists of genes in our modules with the lists of genes in modules reported by Fromer et al. [14] and Gandal et al. [13]. Methodological details are provided in supplementary material.

4. Prioritizing modules of co-expression based on their relevance to SCZ

To prioritize our co-expression modules for their relevance to illness and to genetic risk for illness, we used intramodular analysis routines from the WGCNA package [23, 24]. We first defined a measure of “Gene Significance” (GS) represented by the absolute value of Pearson correlation between each gene’s expression and diagnosis status. By averaging the GS for each module, we obtained a measure of “Module Significance” (MS). A module with high significance for diagnosis would be a module with many genes strongly correlated with the diagnosis. We then plotted the measure of module significance to visualize the most relevant modules for the association with the diagnosis (modules above a conventional cutoff = 0.15 were considered significant [23, 24]).

We next identified modules whose eigengenes (MEs) were significantly correlated with the SCZ PRS calculated for various p value thresholds as previously specified. We focused on modules correlated especially but not exclusively with PRS5–PRS6 because these scores presumably contain most of the true-positive-risk PGC2 genes and explain the maximum diagnostic liability in the PGC sample. After evaluating the relationship between MEs and PRS and diagnosis, we looked at whether PGC2 protein-coding genes were significantly enriched in any of these modules, in order to select the overlapping PGC2 genes and prioritize them according to criteria presented in Fig. 1c. In addition to associations with PRS, we also performed sensitivity analyses (linear models) to assess age and sex as potential confounders that biased the results of MEs–PRS correlations (Methods in supplementary material).

5. Prioritizing genes associated with genetic risk for SCZ within the co-expression networks

Figure 1c describes our hierarchical approach for prioritizing genes associated with genetic risk for SCZ from the latest GWAS based on the co-expression modules that they inhabit. We consider three tiers of gene sets consisting of protein-coding genes within the 108 loci with GWAS

significant genetic signal for risk of SCZ identified by the PGC2 [1] and their presence within modules potentially relevant for association with SCZ pathogenesis as determined by the convergence of biologically relevant function, clinical state, and genetic risk as explained above. Tier 1, i.e., the highest priority, comprises PGC2 loci genes that are enriched in modules found to be significant for association with both diagnosis and PRSs. Tier 2 includes genes enriched in modules associated with diagnosis but not with PRS. Finally, tier 3 includes only the PGC2 genes overrepresented in any module that was not associated with diagnosis or with the PRS.

We calculated the enrichment of PGC2 loci protein-coding genes (obtained from supplementary table 2 in ref. [1]) in the entire co-expression network (Fig. 1c). Two groups of genes were compared: one represented by all protein-coding genes used for the network construction annotated by gene symbol ($N = 15,359/22,945$) and labeled by the color of their corresponding module and one represented by 309/349 PGC2 GWAS significant loci protein-coding genes according to supplementary table 2 from ref. [1]. We selected only 309 protein-coding genes because 40 had especially low abundance in our data set ($\text{RPKM} < 0.1$) and therefore were not analyzed. For every pair of lists overlaps tested, the output was represented by uncorrected and familywise error (FWE)-corrected p values and by overlapping genes, respectively, the PGC2 GWAS significant loci protein-coding genes represented in a module.

Results

Demographic characteristics of the brain samples

All samples, CTRL and SCZ, were from CAUC donors. The two diagnostic groups did not significantly differ by age of death ($\text{mean}_{\text{age}} = 43.07$ years, $\text{SD} = 14.95$ years, $\text{min}_{\text{age}} = 16.17$ years, $\text{max}_{\text{age}} = 77.94$ years, t -statistic = -1.2267 , $p = 0.2217$). There was no statistically significant difference in sex distribution between CTRL and SCZ samples ($\text{female}_{\text{total}} = 40$, $\text{female}_{\text{SCZ}} = 22$; $\text{male}_{\text{total}} = 124$, $\text{male}_{\text{SCZ}} = 52$; χ -squared = 1.5905, $\text{DF} = 1$, $p = 0.2073$). At the time of death, 32/74 SCZ were recorded as “negative” for antipsychotics, 41/74 were under antipsychotic treatment, and 1 had no available data with respect to treatment. Also, PRS did not correlate with sex or age at death (the detailed sample description is presented in supplementary table 1).

WGCNA results

After expression data pre-processing, RNA-quality correction, and network construction, 12 gene co-expression

Table 1 Summary of “Gene Ontology Biological Processes” (GO-BP) and cell markers enriched in modules of co-expression

Modules and top 10 GO-BP enrichments	
Black- Astrocytes	
sensory organ development	astrocyte differentiation
eye development	carboxylic acid catabolic process
small molecule catabolic process	camera-type eye development
gliogenesis	organic acid catabolic process
glial cell differentiation	renal system development
Blue- Neuron 1	
neurotransmitter secretion	vesicle-mediated transport in synapse
signal release from synapse	regulation of neurotransmitter levels
presynaptic process involved in chemical synaptic transmission	regulation of neuron projection development
modulation of synaptic transmission	neurotransmitter transport
neuron projection morphogenesis	synaptic vesicle transport
Brown- Neuron 2	
single-organism behavior	learning
learning or memory	potassium ion transport
cognition	regulation of ion transmembrane transport
behavior	regulation of membrane potential
homophilic cell adhesion via plasma membrane adhesion molecules	modulation of synaptic transmission
Green- Mitochondria 1	
protein targeting to ER	ribonucleoprotein complex biogenesis
translational initiation	protein localization to endoplasmic reticulum
establishment of protein localization to endoplasmic reticulum	mitochondrial ATP synthesis coupled electron transport
SRP-dependent cotranslational protein targeting to membrane	ATP synthesis coupled electron transport
cotranslational protein targeting to membrane	oxidative phosphorylation
Magenta- Mitochondria 2	
nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	establishment of protein localization to endoplasmic reticulum
SRP-dependent cotranslational protein targeting to membrane	mRNA catabolic process
cotranslational protein targeting to membrane	protein localization to endoplasmic reticulum
nuclear-transcribed mRNA catabolic process	viral transcription
protein targeting to ER	viral gene expression
Red- Oligodendrocytes	
ensheathment of neurons	negative regulation of neurogenesis
axon ensheathment	negative regulation of cell development
myelination	gliogenesis
oligodendrocyte differentiation	negative regulation of nervous system development
glial cell differentiation	negative regulation of cell differentiation
Salmon- Microglia	
lymphocyte activation	immune response-activating signal transduction
positive regulation of immune response	adaptive immune response
activation of immune response	regulation of leukocyte activation
immune effector process	regulation of cell activation
immune response-regulating signaling pathway	leukocyte cell-cell adhesion

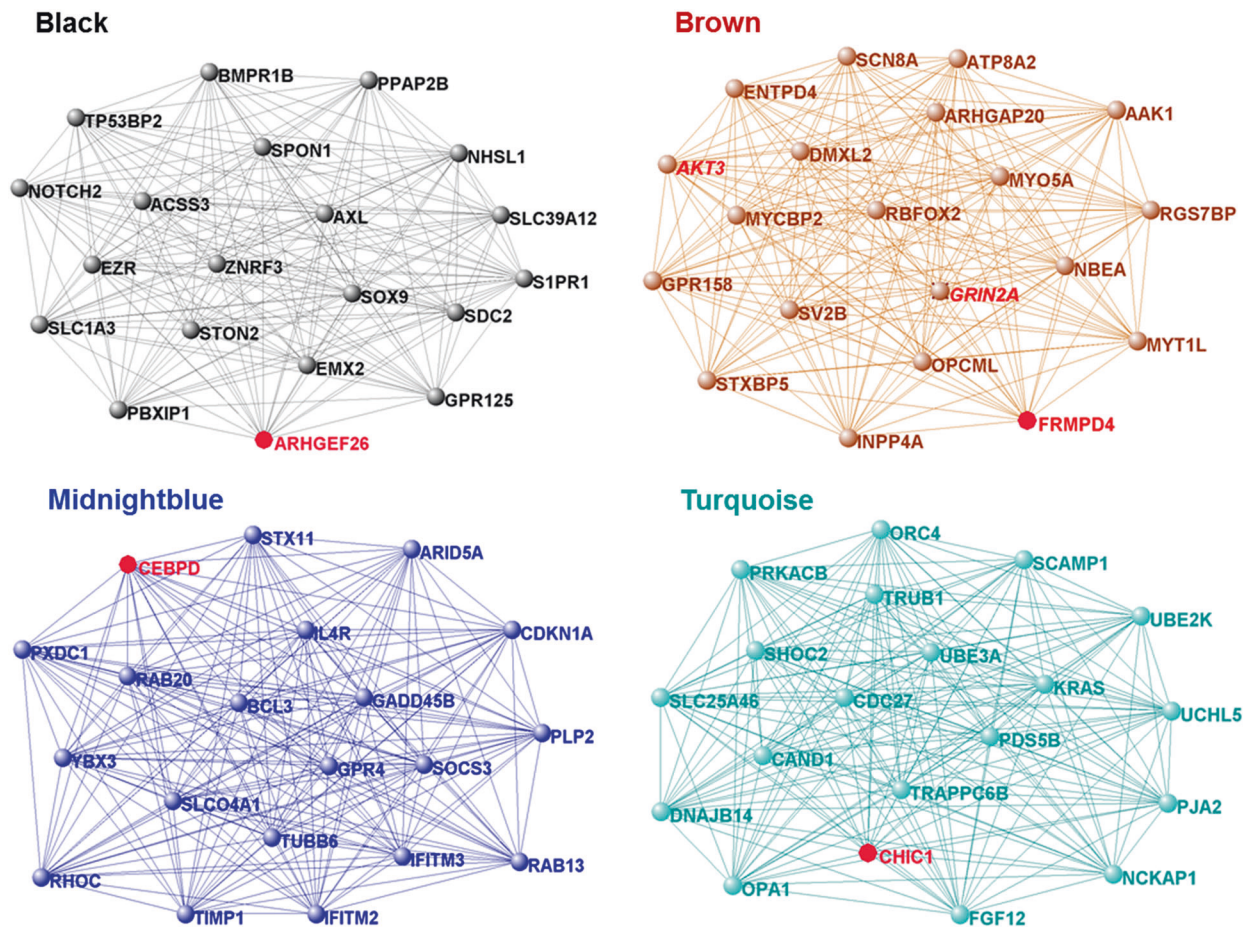


Fig. 2 Top twenty most connected genes in modules significantly associated with diagnosis and polygenic scores. Modules are represented by their colors assigned through co-expression network construction and module detection. Nodes colored in red = modules’

hubs (the most connected gene in the module; for brown module, 2 PGC2 loci genes (*AKT3* and *GRIN2A* labeled in red italics) are among the top 20 most connected genes. Modules’ visualization was created with VisAnt (<http://visant.bu.edu/>)

modules were identified in the overall sample with size between 40 and 1813 genes ($N = 12,475$, 54%, were “gray” genes, not assigned to a module). The module-specific gene distribution is reported in supplementary table 2.

The post hoc WGCNA performed on the expression of selected genes after controlling also for hidden qSVs as described in supplementary material yielded also a pattern of 12 modules, 7 of which were enriched for ontologies and cellular markers similar with the “core” modules identified in the primary co-expression network analysis (supplementary table 3). In general, we found a significant degree of overlap between the modules of interest defined from the entire gene set network and the post hoc qSV analysis-constructed modules based on the selected genes (supplementary table 3 and supplementary figure 1). We note, however, that qSV correction, a well-validated method for differential expression and expression quantitative trait loci analysis [26], has not been critically examined in the context of network analysis, and this conservative approach results in a higher proportion of

genes unassigned to modules (“gray” genes) (unpublished observations).

Additional results of module diagnostics for the primary co-expression network are also reported in supplementary figures 2-5, which show for the majority of modules (except cyan) characteristic band structures suggestive of well-defined modules, consistent across samples as previously described [23, 25]. Preservation analysis performed for the primary network constructed from 22,945 genes indicated that all 12 modules were preserved in the CMC data (supplementary figure 6). The preservation statistics varied between $Z_{\text{summary}} = 10$ (“midnightblue” module) and $Z_{\text{summary}} = 60$ (“yellow” module), all of which are considered “strong” preservation results.

General biological characterization of DLPCF modules: enrichment in GO-BP and cellular markers

Table 1 and supplementary figure 7 summarize the significant and biologically plausible modules of gene

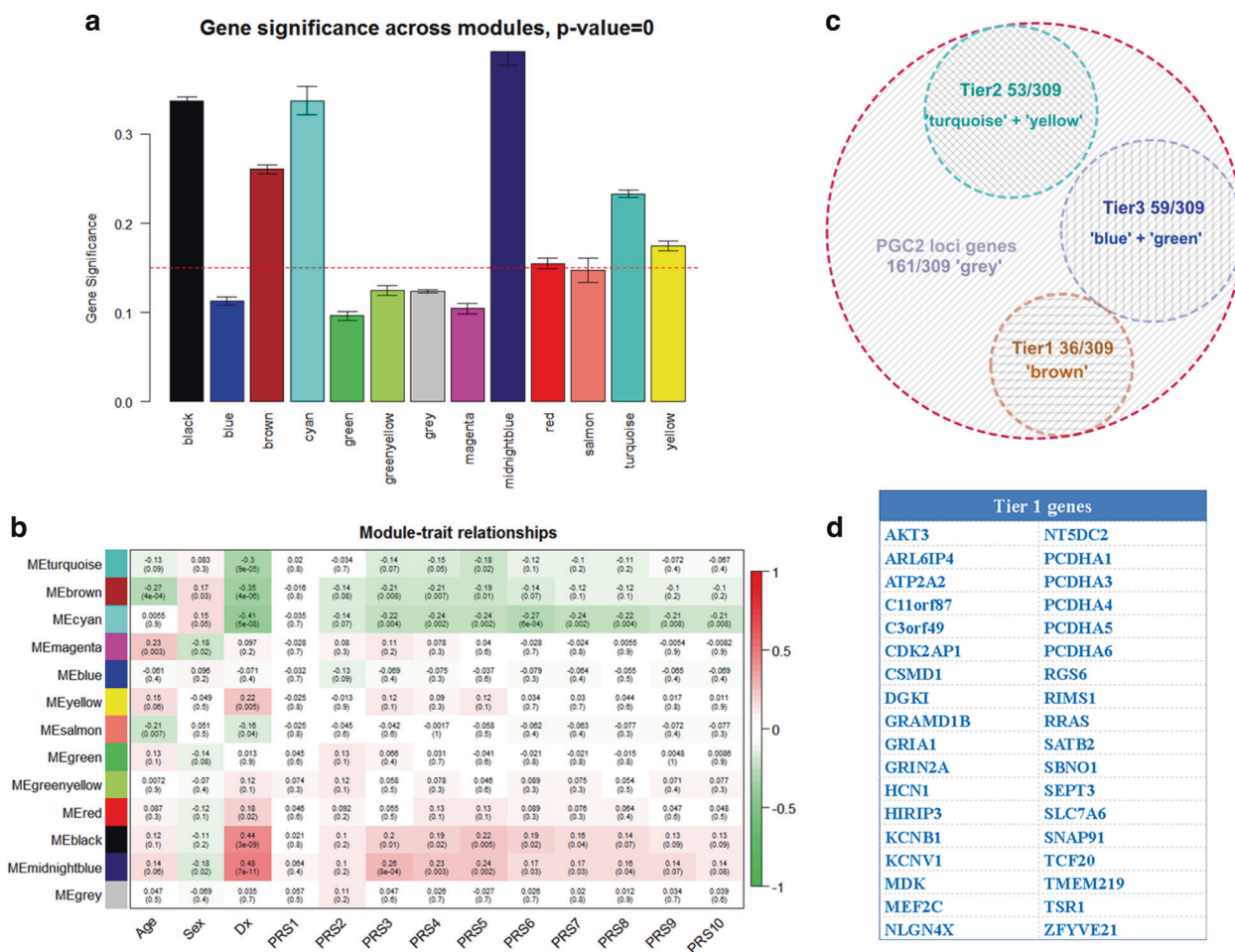


Fig. 3 Prioritization of PGC2 loci genes based on the distribution in modules associated with polygenic risk score and diagnosis of SCZ. **a** Modules significantly correlated with the diagnosis of SCZ;

b Modules' eigengenes (MEs) correlated with diagnosis and PRS; **c** Partitioning of PGC2 loci genes in tiers 1–3; **d** Tier 1 genes

co-expression identified in our networks based on CTRL and SCZ DLPFC RNA-Seq data. In general, these modules are not diagnostically specific and presumably reflect patterns of gene co-expression in human adult DLPFC gray matter. Modules reflecting similar biological processes have been observed in earlier work in the human brain [7]. Notably, we found seven “core” modules reproducing fundamental processes for nervous system development (i.e., neuronal differentiation and migration, synaptogenesis, gliogenesis, and myelination) and functionality, metabolic processes critical for cellular survival and function, including nervous cells, specific neuronal processes linked to neuronal excitability and synaptic activity, immune system functions, mechanisms of transcription and translation, etc. A complete list of GO-BP enrichment for the empB modules is presented in supplementary table 3. Importantly, the organization of our co-expression modules significantly overlapped with the cell-type-specific modules originally reported in the cortex by

Oldham et al, [7] (supplementary figure 7, Table 1, supplementary table 4).

From our empB modules not enriched for GO-BP but significantly enriched for cellular markers, we note: turquoise (overlapping with “neuron-M16_CTX,” M18_CTX, M11_CTX, “interneurons-M17_CTX”), yellow (overlapping with “glutamatergic synaptic function-M10, M18_CTX, M19_CTX”), and greenyellow (“oligodendrocytes-M9_CTX,” “glutamatergic synaptic function-M10,” and “astrocytes-M15_CTX”). We also note that enrichment in cellular markers for these modules was more mixed than for the seven “core” modules (supplementary table 4).

Identification of co-expression networks related to SCZ diagnosis and to genetic risk

We found that the MEs of four modules were significantly correlated with both PRS and diagnosis: the black module

(enriched for astrocytes markers), the midnightblue module (mixed cellular markers), the brown (enriched for neuronal markers), and the cyan module (mixed cellular markers) (Figs. 2 and 3b). Turquoise module was correlated with diagnosis and only marginally with PRS5 (Figs. 2 and 3b). Brown and midnightblue MEs were also significantly correlated with age and sex (Fig. 3b). For the brown module that was the source for tier 1 genes (see below), subsequent adjustment for age and sex resulted in consistent effects estimates and *p* values for the PRS effects on ME brown. In addition, we show that age and sex do not have confounding effects on the association between the ME brown and diagnosis (supplementary results).

Modules correlated only with diagnosis were the yellow module (enriched for markers of glutamatergic synaptic function) and, marginally, the salmon module (enriched for microglial markers) and the red module (enriched for oligodendrocytes markers) (Fig. 3a). Only the brown module, however, showed a significant overrepresentation of PGC2 loci genes (*p* corrected = 0.00327) (Table 2). A synthesis of the associations between modules and diagnosis and polygenic scores, combined with the information regarding the PGC2 loci genes distribution within modules is presented in Table 2.

Overall, we found that almost half of the GWAS significant PGC2 loci protein-coding genes (148/309) were

Table 2 Summary of the relationships between co-expression modules and diagnosis, PRS, and intramodular PGC2 loci gene distribution

Modules	Association with dx	Association with PRS	Enrichment in PGC2 loci genes
Black	yes	yes	N.S.
Blue ³	no	no	yes
Brown ¹	yes	yes	yes
Cyan ^{**}	yes	yes	N.S.
Green ³	no	no	yes
Greenyellow	no	no	N.S.
Magenta	no	no	N.S.
Midnightblue [*]	yes	yes	N.S.
Red	±	no	N.S. (<i>n</i> ⁰ = 24)
Salmon	±	no	N.S.
Turquoise ²	yes	±	N.S. (<i>n</i> ⁰ = 32)
Yellow ²	yes	no	N.S. (<i>n</i> ⁰ = 21)

*n*⁰ number of PGC2 loci genes in the modules used for tier 2 genes selection

¹Modules used for tier 3 gene selection

²Module used for tier 1 gene selection

³Module with no GO-BP enrichment

⁴Small-sized module with no GO-BP enrichment and weakly preserved in CMC data

⁵Modules used for tier 2 gene selection

distributed in modules across the co-expression networks: 36 of them (which we designate as tier 1 genes below) were in modules related to both diagnosis and PRS, 53 of them were in modules related only to diagnosis (tier 2), and 59 were in modules unrelated to diagnosis or PRS (Fig. 3c).

Tier 1 genes: PGC2 loci genes overrepresented in modules associated with both diagnosis and PRS

We selected as our tier 1 gene set 36 PGC2 loci protein-coding genes that were represented in the brown module (Fig. 3d) because this module's eigengene is associated both with diagnosis and with PRS and brown is the only module with ME association with both of these features and also with PGC2 locus gene overrepresentation. Notable tier 1 genes are AKT3, essential in brain development [31], ATP2A2 (encoding for a magnesium-dependent enzyme that catalyzes ATP's hydrolysis [32], genes involved in neural circuits development via cell adhesion processes (protocadherins alpha cluster—PCDHA1,3,4,5,6, ZFYVE21, RRAS), transcription factors (TCF20, MEF2C), regulators of G-protein signaling (RGS6), RRAS—a small GTPase involved in cell adhesion and axon guidance, and TSR1 (Ribosome Maturation Factor), which is important for fundamental functions related to protein synthesis and gene expression in all cells [32]. We also note among tier 1 genes two glutamatergic ionotropic receptors (GRIA1 and GRIN2A) and several potassium ion channels receptors (HCN1—Hyperpolarization Activated Cyclic Nucleotide Gated Potassium Channel 1, KCNB1—Potassium Voltage-Gated Channel Subfamily B Member 1 associated with epilepsy, and KCNV1—Potassium Voltage-Gated Channel Modifier Subfamily V Member 1, essentially expressed in the brain) [32].

Interestingly, the two voltage-gated potassium channels—KCNB1 and KCNV1 that are tier 1 genes—are also co-expressed with the voltage-gated sodium channels SCN8A, SCN2A, and SCN3B. Of note, SCN8A is essential for the rapid membrane depolarization during the formation of the action potentials in neurons [32]. This observation is illustrative of the ability of co-expression networks to give a more coherent view on possible functional complexes implicated in higher-order neuronal function in comparison with gene-by-gene analyses. It is also noteworthy that a number of these genes are “druggable” targets.

Tier 2 genes: PGC2 loci genes represented in modules significant for the diagnosis of SCZ but not PRS

We note 53 PGC2 genes as tier 2 genes, clustered in turquoise and yellow modules (supplementary table 5). However, PGC2 loci genes enrichment did not reach the levels of significance after FWE correction for multiple

comparisons (uncorrected values were $p = 0.52$ for yellow, $p = 0.62$ for turquoise (supplementary table 5). As previously mentioned, these modules represented mixtures of cellular markers (supplementary table 4) and were not significantly enriched for GO-BP. However, they are potentially interesting modules based on selected PGC2 genes present in them that have been the subject of prior discussion in the literature, including ZNF804A [33], NDUFA13, NDUFA4L2, NOSIP, and NRG1.

Tier 3 genes: PGC2 genes overrepresented across networks

Fifty-nine PGC2 GWAS significant loci genes belonged to tier 3: they were overrepresented in blue and green modules that were neither associated with PRS nor with the diagnosis. Of note, blue is one of the neuronal modules and green is enriched for mitochondria markers (Fig. 2). Interestingly, tier 3 genes based on prior data related to SCZ include CACNA1C, CHRM4, FURIN, TSNARE1 (possibly implicated in neurotransmitter release by regulating the SNARE—Synaptotagmin complex), and FXR1 (FMR1 Autosomal homolog, associated with Fragile X syndrome) [32]. The complete list of the tier 1–3 PGC2 loci genes is presented in supplementary table 6.

Finally, 161/309 PGC2 locus genes were randomly distributed in modules that did not pass the threshold of significance for PGC2 enrichment or were “gray” genes (not assigned to a module in any network). A notable “gray” PGC2 locus gene was DRD2, the best-established drug target for SCZ. While not excluding the possible implication of any PGC2 gene in the etiopathogenesis of SCZ, based on our results, we believe that tier 1 and tier 2 genes, and importantly, the other genes in their specific co-expression modules, which are not PGC2 risk associated loci genes, are more attractive candidates for experimental studies to decipher the pathophysiological mechanisms of SCZ and therapeutic drug development.

Discussion

We have performed an extensive analysis of gene co-expression architecture in adult postmortem DLPFC from Caucasian unaffected donors (CTRL) and from donors diagnosed with SCZ. The main purpose of the study was to identify and prioritize co-expressed gene sets in a hierarchical manner, based on the strength of the relationships with clinical diagnosis and with polygenic risk for SCZ. For this purpose, we focused on co-expression modules that included PGC2 protein-coding genes, i.e., the genes within the significant 108 loci reported in the latest published GWAS of SCZ [1]. The rationale of our approach was that finding a convergence between the co-expression

architecture in a region with known molecular deviations in SCZ (i.e., DLPFC) with both the illness state and with genomic risk for the illness is a more optimal strategy to identify co-functional gene sets of potential pathogenic relevance that could be investigated as harboring novel drug targets for SCZ. Importantly, we found seven “core” modules enriched for meaningful ontologies and significantly overlapping with modules reported by other groups. We further selected and hierarchized PGC2 loci genes overrepresented in modules of co-expression by the modules’ relationship with PRS and diagnosis of SCZ. This additional step adds confidence that modules so identified are not likely to be based solely on illness state phenomena, many of which (e.g., treatment, chronicity effects) may be epiphenomena, or on technical artifacts.

The tier 1 genes

We identified 36 PGC2 locus genes that were distributed in one module (brown), a module significant for both diagnosis and association with genomic risk, i.e., PRS (Fig. 3). In principle, these genes and this network should bear an especially close relationship to SCZ pathogenesis and pathobiology. These genes are members of putatively relevant signaling pathways, such as phosphoinositide-3 kinase (PI3K)/AKT signaling, which has numerous functions in neurodevelopment and adult brain and has been implicated in a variety of neurological and mental disorders, including SCZ [34]; a Ca^{2+} signaling pathway with numerous functions, including energetic metabolism [32] that was underscored in our study mainly by ATP2A2; a RAS/extracellular signal-regulated kinase (ERK) signaling pathway represented by RRAS (RAS-related small GTPase) implicated in cell adhesion and axon guidance [32].

Interestingly, we explored our tier 1 genes with the informatic tool STRING, a database of known and predicted protein–protein interactions [35] and noticed that several of the tier 1 genes are co-expressed in the same module with some of their interactors. For example, the PGC2 locus gene AKT3 is co-expressed with at least eight of its predicted interactors (ADCY2, CREB3L4, EIF4EBP1, GNB5, GSK3B, PHF20, PHLPP2, PIK3R1) [32, 35] and with a serotonergic receptor modulated by GSK3B (i.e., HTR2A). Another PGC2 locus gene, ATP2A2 (ATPase Sarcoplasmic/ Endoplasmic Reticulum Ca^{2+} Transporting), and its interactors (CALM1, RYR2, ITPR1) indicate potential dysfunctions on a Ca^{2+} signaling pathway in relation to energetic metabolism [32, 35]. NLGN4X (Neurigin 4 X linked) previously implicated in autism and some of its interactors also members of the brown module (NRXN3, DLGAP1, DLG2, GRM1, GRM5) are constituents of the postsynaptic density and regulators of glutamatergic signaling [32, 35]. We highlight also RIMS1,

co-expressed in the brown module with RIMBP2 (RIMS Binding Protein 2) and synaptotagmins (SYT10, SYT11, SYT13, SYT16) [32, 35]. Some of the tier 1 genes in the brown module are in the RAS/ERK signaling pathway: RRAS (RAS related small GTPase) is implicated in cell adhesion and axon guidance together with its interactors [35] and co-expression partners, i.e., PAK3 (role in dendritic development and synaptic plasticity), BRAF, RASGRP1, RASSF5, PRKCB (important for GABA-ergic synapse), RASAL2, RAPGEF2 (involved in neuritogenesis, neuronal migration), and RASGRF2.

The presence of intramodular sub-clusters from different signaling pathways in our co-expression networks may indicate higher-order inter-network interactions. This scenario is plausible given previous studies that have shown crosstalk between PI3K/AKT and RAS/ERK signaling pathways that regulates neurodevelopmental processes and synaptic plasticity [34]. This expected crosstalk between signaling pathways has been highlighted in a recent article about the underlying polygenic architecture of complex clinical syndromes [36]. The co-expression of the voltage-gated potassium channels KCNB1 and KCNV1 (both PGC tier 1 genes) with voltage-gated sodium channels, in particular with SCN8A (the initiator of action potentials), a gene not identified by GWAS, is particularly noteworthy. Their pattern of co-expression underscores the potential higher-order functional coherence of the “Brown” module in terms of synaptic activity and suggests that a critical target of drug treatment may be to optimize neuronal membrane potential.

Our identification of a module of gene co-expression that is associated with SCZ risk and also with illness state, enriched with GWAS significant genes, implicates a set of interacting signaling networks that may represent a “core” pathogenic network. A recent proposal suggests that complex traits are “omnigenic” with GWAS loci genes influencing diverse cell signaling networks only some of which are core networks fundamental to pathogenesis [36]. The “Brown” gene module may thus represent a core biological pathway as suggested by the omnigenic hypothesis, and the genes in this network, including but in addition to the GWAS loci genes, may be insights to therapeutic intervention.

Tier 2 genes

We identified 53 potentially tier 2 PGC2 loci genes that were distributed in modules significant for diagnosis only, turquoise and yellow. A notable tier 2 gene in the turquoise module is ZNF804A, the first gene identified using a GWAS approach to SCZ [33]. However, the mechanism by which ZNF804A is implicated in SCZ etiopathogenesis has yet to be determined, and earlier reports suggest association with a novel isoform only during fetal life [37]. Recent

studies support the hypothesis that ZNF804A has multiple and important roles in neuronal physiology, including transcription regulation of interacting genes involved in cell adhesion, neurite outgrowth, and synapse formation [33]. Interestingly, we found that several genes which demonstrated transcriptional variation in studies based on ZNF804A knockdown were ZNF804A partners of co-expression in the turquoise module: C2Orf80 (unknown function), EIF4A2 (Eukaryotic Translation Initiation Factor), and ATP1B1 (ATPase Na⁺/K⁺ responsible for maintaining the Na-K gradients across plasma membrane) [38, 39]. It is also interesting to note that, in this large module, ZNF804A is just 1 of the 42 co-expressed transcription factors from the ZNF family. This implicates the formidable transcriptional regulation machinery that is putatively mobilized during various neuronal functions.

It seems worthy of comment that several historic candidate genes for SCZ are members of the modules containing tier 1 and tier 2 genes. Most notable examples are dopaminergic receptors (DRD4), receptor tyrosine kinases (i.e. ERBB4, receptor for neuregulins), NRG3 (growth factor that mediates cell–cell signaling and has multiple roles in neurodevelopment and has been previously associated with SCZ [40]), GABA receptors and glutamate decarboxylases involved in GABA synthesis (GAD1, GAD2), glutamatergic receptors, ionotropic and metabotropic (GRIA2, 3, 4, GRM1, GRM5, GRM8), serotonergic receptors (i.e., HTR2A), RGS4, and COMT. The inconsistent results of previous studies on some of these genes, coupled with the lack of confirmation by GWAS, prompted a gratuitous rebuke of the role played by historic candidate genes in the genetic risk for SCZ [41]. However, our results based on DLPFC co-transcriptome architecture suggest a complex distribution of relevant genetic factors possibly organized in a modular fashion and indicative of vast gene–gene interactions that entrain new and previous genes associated with SCZ.

Tier 3 PGC2 genes

We found that 59/148 PGC2 genes were enriched in two modules (blue and green) not related to the diagnosis of SCZ or to PRS. One of the stand-out tier 3 genes in the blue module is CACNA1C, strongly associated with the risk for SCZ and bipolar disorder [42]. Other notable tier 3 PGC2 genes are CHRM4 (a drug target for SCZ [43]), FURIN and TSNARE1. The latter two, FURIN and TSNARE1 were recently highlighted by Fromer et al. [14] who showed in an experimental model of zebrafish neurodevelopment that overexpression of TSNARE1 and suppression of FURIN were associated with decrease of head size. While these genes are related to genetic risk, their association with modules not associated with either risk or illness suggest

that their potential roles in pathogenesis will not be delineated in adult postmortem DLPFC.

Comparisons with previous work

Notwithstanding methodological differences, we found that our cortical gene expression networks were roughly consistent with previous similar studies. Of note, modules in our CAUC (CTRL+SCZ) network were significantly overlapping with modules originally reported by Oldham et al. [7] (supplementary figure 7 and supplementary table 4). Likewise, our results were in reasonable agreement with Fromer et al. [14], considering the methodological differences related to mRNA processing, network construction, and demographic characteristics of the samples (supplementary table 7). Interestingly, we found that 19 of the 31 PGC2 genes overrepresented in their module of interest—M2C—were members of at least one of our tier genes. Moreover, 9 of these 19 genes were actually tier 1 PGC2 genes in our study (Fig. 3d): SBNO1, TCF20, KCNB1, GRIA1, ATP2A2, HCN1, CSMD1, GRIN2A, and NLGN4X. We also found a significant overlap between the modules of co-expression recently reported by Gandal et al. [13] and our modules of co-expression (supplementary table 8).

Methodological considerations

Although these results are intriguing in many respects, we cannot rule out the possibility that they represent at least in part coincidental events, spurious associations, or effects of ongoing epiphenomena. Some of the limitations of our study are related to the network construction based on expression data at the gene level, which most probably obscures even more complex correlation patterns at the transcript and isoform level; likewise, we used RNA extracted from a tissue with a heterogeneous cell composition, which may not capture the cell-type-specific co-expression architectures. One known limitation of approaches associating genes with genetic risk variants based on genomic location is that significant SNPs do not always tag the nearest gene [44], due to linkage disequilibrium and other epigenetic factors related to gene expression regulation. Therefore, our prioritization strategy of using genes physically localized within GWAS hits may not reveal the exact subset of genes implicated in SCZ risk. However, we believe that the significant enrichments of these modules with these positional overlaps likely still capture the subset of genes that do represent SCZ risk, and future work can refine our network-level enrichments.

The finding that association between the ME of our top module of interest (brown) and polygenic scores is less strong than its correlations with age and sex raised the

possibility of confounding effects explained by these covariates. However, we believe this is not the case as correlations between brown ME and PRS remained significant after controlling for age and sex effects. We also cannot rule out the role of treatment exposure of SCZ samples or other epiphenomena in module construction in contrast to primary illness mechanisms. For example, animal studies have indicated some overlap between haloperidol regulation and co-expression networks enriched for SCZ genetic signals [12]. Further, while we have endeavored to pay special attention to the role of RNA quality as a confounder in co-expression, we cannot rule out this factor, and similarities to earlier work do not exclude a shared artifact. We have stressed genes in tier 1 because of the convergence of association with illness state and also with genetic risk, the latter not likely related to potential epiphenomena and confounders. In principle, genetic risk association obviates state only factors, but this is still conjecture. Our and others' work have previously suggested that much of the genetic risk for SCZ has been largely orthogonal to the illness state-related differences in gene expression (Fromer et al. [14], Jaffe et al. [2]). This study using co-expression analysis indeed suggests some overlap in gene networks between genetic risk and illness state, which was missed at the level of individual genes.

While modules of gene co-expression represent potential insights toward understanding physiological and etiopathogenic mechanisms, firm evidence of functional relevance requires experimental studies. Of note, our approach to network construction, based on enhanced adjustment for RNA quality, yielded a greater percentage of genes in gray, i.e., not in an explicit module, than in earlier studies. We believe this reflects several factors, including our relatively small sample size and removal of more complete co-expression based on co-degradation.

In conclusion, our study offers an extensive characterization of the co-transcriptome in the postmortem DLPFC of non-affected individuals and individuals with SCZ. Our results indicate potentially broad interactions of PGC2 locus genes, which may represent the tip of an iceberg of multiple convergent signaling pathways associated with genetic risk of SCZ revealed through the co-transcriptome architecture in DLPFC. Interestingly, altered mechanisms suggested by these pathways span from prenatal neurodevelopmental events through brain functionality in adult life and hint not only at genetic factors but also to an environmental contribution. Most importantly, our study highlights complex relationships between gene co-expression networks in the brain and polygenic risk for SCZ and provides a strategy for using this information in selecting potentially targetable gene sets for therapeutic drug development. Targets identified because they inhabit high confidence networks related to both risk and the illness state offer therapeutic potential

beyond what might be inferred from any specific gene candidate.

Acknowledgements We are grateful for the generosity of the Lieber and Maltz families in establishing an institute dedicated to understanding the basis of developmental brain disorders. We would like to gratefully acknowledge the families of the subjects whose donations made this research possible. This research was funded by the Lieber Institute for Brain Development.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 2014;511:421–7.
- Jaffe AE, Straub RE, Shin JH, Tao R, Gao Y, Torres LC, et al. Developmental and genetic regulation of the human cortex transcriptome illuminate schizophrenia pathogenesis. *Nat Neurosci*. 2018;21:1117–25.
- Hannon E, Spiers H, Viana J, Pidsley R, Burrage J, Murphy TM, et al. Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci. *Nat Neurosci*. 2016;19:48–54.
- Webber C. Epistasis in neuropsychiatric disorders. *Trends Genet*. 2017;33:256–65.
- Birnbaum R, Weinberger DR. Genetic insights into the neurodevelopmental origins of schizophrenia. *Nat Rev Neurosci*. 2017;18:727–40.
- Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinforma*. 2008;9:559.
- Oldham MC, Konopka G, Iwamoto K, Langfelder P, Kato T, Horvath S, et al. Functional organization of the transcriptome in human brain. *Nat Neurosci*. 2008;11:1271–82.
- Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S, et al. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature*. 2011;474:380–4.
- 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;18:1308–14.
- Torkamani A, Dean B, Schork NJ, Thomas EA. Coexpression network analysis of neural tissue reveals perturbations in developmental processes in schizophrenia. *Genome Res*. 2010;20:403–12.
- Roussos P, Katsel P, Davis KL, Siever LJ, Haroutunian V. A system-level transcriptomic analysis of schizophrenia using postmortem brain tissue samples. *Arch Gen Psychiatry*. 2012;69:1205–13.
- Kim Y, Giusti-Rodriguez P, Crowley JJ, Bryois J, Nonneman RJ, Ryan AK, et al. Comparative genomic evidence for the involvement of schizophrenia risk genes in antipsychotic effects. *Mol Psychiatry*. 2017;23:708–712.
- Gandal MJ, Haney JR, Parikshak NN, Leppa V, Ramaswami G, Hartl C, et al. Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science*. 2018;359:693–7.
- 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;19:1442–53.
- Pergola G, Di Carlo P, D'Ambrosio E, Gelao B, Fazio L, Papalino M, et al. DRD2 co-expression network and a related polygenic index predict imaging, behavioral and clinical phenotypes linked to schizophrenia. *Transl Psychiatry*. 2017;7:e1006.
- BrainSeq A. Human Brain Genomics Consortium. BrainSeq: neurogenomics to drive novel target discovery for neuropsychiatric disorders. *Neuron*. 2015;88:1078–83.
- TopHat: <https://ccb.jhu.edu/software/tophat/index.shtml>
- Ensembl: <http://www.ensembl.org/index.html>
- Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*. 2009;5:e1000529.
- Delaneau O, Coulonges C, Zagury JF. Shape-IT: new rapid and accurate algorithm for haplotype inference. *BMC Bioinforma*. 2008;9:540.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559–75.
- Chen Q, Ursini G, Romer AL, Knodt AR, Mezevitch K, Xiao E, et al. Schizophrenia polygenic risk score predicts mnemonic hippocampal activity. *Brain*. 2018;141:1218–1228.
- WGCNA manual: <https://cran.r-project.org/web/packages/WGCNA/WGCNA.pdf>
- WGCNA tutorials: <https://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/Rpackages/WGCNA/Tutorials/index.html>
- Horvath, S. Networks and fundamental concepts. *Weighted network analysis. Applications in systems biology*. Springer: New York, Dordrecht, Heidelberg, London; 2011. p.1–34.
- Jaffe AE, Tao R, Norris AL, Kealhofer M, Nellore A, Shin JH, et al. qSVA framework for RNA quality correction in differential expression analysis. *Proc Natl Acad Sci USA*. 2017;114:7130–5.
- Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 2007;8:118–27.
- Langfelder P, Luo R, Oldham MC, Horvath S. Is my network module preserved and reproducible? *PLoS Comput Biol*. 2011;7:e1001057.
- Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16:284–7.
- Cahoy JD, Emery B, Kaushal A, Foo LC, Zamanian JL, Christopherson KS, et al. A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J Neurosci*. 2008;28:264–78.
- Poduri A, Evrony GD, Cai X, Elhosary PC, Beroukhi R, Lehtinen MK, et al. Somatic activation of AKT3 causes hemispheric developmental brain malformations. *Neuron*. 2012;74:41–8.
- GeneCards: www.genecards.org
- Chang H, Xiao X, Li M. The schizophrenia risk gene ZNF804A: clinical associations, biological mechanisms and neuronal functions. *Mol Psychiatry*. 2017;22:944–53.
- Manning BD, Toker A. AKT/PKB signaling: navigating the network. *Cell*. 2017;169:381–405.
- Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRINGv10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*. 2015;43(Database issue):447.
- Boyle EA, Li YI, Pritchard JK. An expanded view of complex traits: from polygenic to omnigenic. *Cell*. 2017;169:1177–86.
- Tao R, Cousijn H, Jaffe AE, Burnet PW, Edwards F, Eastwood SL, et al. Expression of ZNF804A in human brain and alterations in schizophrenia, bipolar disorder, and major depressive disorder:

- a novel transcript fetally regulated by the psychosis risk variant rs1344706. *JAMA Psychiatry*. 2014;71:1112–20.
38. Hill MJ, Jeffries AR, Dobson RJ, Price J, Bray NJ. Knockdown of the psychosis susceptibility gene ZNF804A alters expression of genes involved in cell adhesion. *Hum Mol Genet*. 2012;21:1018–24.
 39. Chen J, Lin M, Hrabovsky A, Pedrosa E, Dean J, Jain S, et al. ZNF804A transcriptional networks in differentiating neurons derived from induced pluripotent stem cells of human origin. *PLoS One*. 2015;10:e0124597.
 40. Paterson C, Wang Y, Hyde TM, Weinberger DR, Kleinman JE, Law AJ. Temporal, diagnostic, and tissue-specific regulation of NRG3 isoform expression in human brain development and affective disorders. *Am J Psychiatry*. 2017;174:256–65.
 41. Farrell MS, Werge T, Sklar P, Owen MJ, Ophoff RA, O'Donovan MC, et al. Evaluating historical candidate genes for schizophrenia. *Mol Psychiatry*. 2015;20:555–62.
 42. Kabir ZD, Martinez-Rivera A, Rajadhyaksha AM. From gene to behavior: L-type calcium channel mechanisms underlying neuropsychiatric symptoms. *Neurotherapeutics*. 2017;14:588–613.
 43. Lencz T, Malhotra AK. Targeting the schizophrenia genome: a fast track strategy from GWAS to clinic. *Mol Psychiatry*. 2015;20:820–6.
 44. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 Years of GWAS discovery: biology, function, and translation. *Am J Hum Genet*. 2017;101:5–22.