

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

Eugenia Radulescu M.D., Ph.D.^a, Andrew E Jaffe Ph.D.^{a,b,c,d}, Richard E Straub Ph.D.^a, Qiang Chen Ph.D.^a, Joo Heon Shin Ph.D.^a, Thomas M Hyde M.D., Ph.D.^{a,e,f}, Joel E Kleinman M.D., Ph.D.^{a,e}, Daniel R Weinberger M.D.^{a,f,g,h,l}

Affiliations:

- a. Lieber Institute for Brain Development, Johns Hopkins Medical Campus, Baltimore, MD, 21205, USA*
- b. Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, 21205, USA*
- c. Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, 21205, USA*
- d. Center for Computational Biology, Johns Hopkins University, Baltimore MD 21205 USA*
- e. Department of Neurology, Johns Hopkins School of Medicine, Baltimore, MD 21205, USA*
- f. Department of Neuroscience, Johns Hopkins School of Medicine, Baltimore, MD 21205, USA*
- g. Department of Psychiatry and Behavioral Sciences, Johns Hopkins School of Medicine, Baltimore, MD 21205, USA*
- h. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD 21205, USA*

¹**Address correspondence to:** Daniel R. Weinberger

Lieber Institute for Brain Development
Johns Hopkins Medicine
855 North Wolfe Street
Baltimore, Maryland 21205
410-955-1000
(drweinberger@libd.org)

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 the 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; Caucasians) to construct co-expression networks and detect modules of co-expressed genes. After internal and external validation, modules of selected interest were tested for enrichment in biological ontologies, association with schizophrenia polygenic risk scores (PRS), with diagnosis and 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 a significant association with diagnosis, PRS and significant overlap with 36 PGC2 loci genes, deemed as tier 1 (strongest candidates for drug targets). Fifty-three PGC2 loci genes were in modules associated only with diagnosis (tier 2) and 59 in modules unrelated to diagnosis or PRS (tier 3). In conclusion, 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.

INTRODUCTION

Schizophrenia, 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 schizophrenia (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 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 schizophrenia 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 postmortem 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 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 schizophrenia (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 by 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 intermediate phenotypes of schizophrenia [15]. While these earlier studies provide potentially important evidence for gene network associations in SCZ, they focused on rather particular 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 brain networks of co-expressed genes and their association with the clinical state of schizophrenia and with polygenic risk of SCZ. We first perform a systematic characterization of gene co-expression in postmortem DLPFC tissue from controls (CTRL) and patients with schizophrenia (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 drug targets in schizophrenia.

MATERIALS AND METHODS

General pipeline of data processing

Human Postmortem Tissue

We used postmortem human brain tissue from the LIBD Human Brain Repository (HBR) for testing and the CommonMind Consortium (CMC) brain collection for validation [16, 2]. Dorsolateral prefrontal cortex grey 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 in [16, 2]. Briefly, the samples selected included tissue from adults (age of death=16-80), healthy controls (CTRL; N=90; M/F=72/18) and patients with schizophrenia (SCZ; N=74; M/F=52/22) donors, Caucasians (CAUC), all with $RIN \geq 7$.

RNA-Seq data processing

For the LIBD dataset, RNA was extracted from DLPFC gray matter (BA9/46) and RNA sequencing 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 CMC data.

Processing CMC gene expression data is described in [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 Scores (PRS) calculation

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

Polygenic risk scores (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 PGC2 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), $p = 1$ (PRS10) [22].

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

Processing the gene expression data to remove unwanted variability associated with sequencing and tissue confounders is presented in **figure 1-A1** 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].

Details about the next WGCNA steps are presented in **figure 1 (A2-7)** and **supplementary methods**. In brief:

1. The co-expression network based on empB-adjusted expression data was created for the combined sample of CTRL+SCZ with WGCNA. This method uses correlation between pairs of genes to construct co-expression modules. The modules can then be summarized by the first

principal component (i.e. the “eigengene”) for each module (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 module eigengenes 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

a. Additional diagnostics of the co-expression modules was also performed (**figure 1-A3**). Previous studies in our group have highlighted the importance of RNA quality in gene expression analysis, especially in detection of differential diagnosis effects [25]. 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 moderated regression (e.g. ComBat [26] 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 modelling “unknown” (latent) variables with “quality” surrogate variables (qSVs) [25]. 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, we performed a complementary post-hoc analysis (**figure 1-A4**) 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” [25] of the entire sample (90 CTRL + 74 SCZ).

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) (**figure 1-A5**). 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 supports their validity (details in **supplementary methods** and [27]).

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 *clusterProfiler* R package [28] (**figure 1-A6**).

c. As a further approach to module validation, we sought to see if our modules are similar to those reported in the study by Oldham et al (2008), 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 [7] and additionally, on a transcriptome database including markers for neurons, astrocytes and oligodendrocytes by Cahoy et al (2008) [30] (**figure 1-A7**).

4. Identification of gene sets associated with the genetic risk for SCZ within the co-expression networks

Figure 1-B describes a hierarchical approach of grouping gene sets within the co-expression modules by association with genetic risk for SCZ informed by the Psychiatric Genomic Consortium Genome Wide Association Study (PGC2 GWAS) [1].

We consider three tiers of gene sets relevant for association with SCZ biology, determined by the gradual convergence of biologically relevant function, clinical state and genetic risk. All the genes in the three tiers are PGC2 protein-coding genes within the 108 loci with GWAS significant genetic signal for risk of SCZ. **Tier 1**, i.e. the highest priority, comprises PGC2 loci gene sets that are enriched in modules significant for the association with both diagnosis and polygenic risk scores. **Tier 2** includes genes enriched in modules associated with the diagnosis but not with PRS. Finally, **tier 3** includes only the PGC2 genes over-represented in any module that was not associated with diagnosis or with the PRS.

We calculated the enrichment of PGC2 protein-coding genes (obtained from supplementary table 2 in [1]) in the entire co-expression network (**figure 1-B**). 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 the supplementary table 2 from [1]. We selected only 309 protein-coding genes because 40 had especially low abundance in our data set (RPKM<0.1) and therefore were not analyzed. For every pair of lists overlaps tested, the output was represented by uncorrected and FWE corrected p-values and by overlapping genes, respectively the PGC2 GWAS significant loci protein-coding genes represented in a module.

Prioritizing the PGC2 loci genes distributed in the modules of co-expression

To prioritize the PGC2 protein-coding genes enriched in our co-expression modules, 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 cut-off=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 if the PGC2 protein-coding genes enrichment was significant in any of these modules, in order to select the overlapping PGC2 genes and prioritize them according to our criteria presented in **figure 1-B**.

RESULTS:

After expression data pre-processing, RNA-quality correction and network construction, 12 gene co-expression modules were identified in the overall sample with size between 40 and 1813 genes (N=12,475, 54%, were “grey” genes, not assigned to a module). The module specific gene distribution is reported in **supplementary table 1**.

The post-hoc WGCNA performed on the expression of selected genes after controlling also for hidden quality surrogate variables (qSVs) as described in **supplementary material**, yielded also

a pattern of 12 modules, seven of which were enriched for ontologies and cellular markers similar with the “core” modules identified in the primary co-expression network analysis (**supplementary table 2**). 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 qSVA constructed modules based on the selected genes (**supplementary table 2** and **supplementary figure 1**). We note, however, that qSV correction, a well validated method for differential expression and eQTL analysis [25], 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 (“grey” genes) (unpublished observations).

Additional results of modules diagnostics for the primary co-expression network are also reported in the **supplementary figures 2-5**, which show for the majority of modules (except cyan) characteristic band structures suggestive for well-defined modules, consistent across samples as described previously [23, 26]. 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 $Z_{summary}=10$ (“midnightblue” module) and $Z_{summary}=60$ (“yellow” module).

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

Figure 2 and **table 2** summarize the significant and biologically plausible modules of gene 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 dorsolateral prefrontal cortex grey matter. Modules reflecting similar biological processes have been observed in earlier work in 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 with GO-BP enrichment for the empB modules is presented in **supplementary table 2**. Importantly, the organization of our co-expression modules significantly overlapped with the cell type specific modules reported in cortex by Oldham et al [7] (**figure 2, table 2, supplementary table 3**).

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 3**).

Identification of gene sets associated with genetic risk (PRS) for SCZ within the co-expression networks

Overall, we found that almost half of GWAS significant PGC2 loci protein-coding genes (148/309) were distributed in modules across the co-expression networks; 36 of them (which we designate as **tier 1 genes**) were in modules related to both diagnosis and polygenic risk score, 53 of them were in modules related only to diagnosis (**tier 2**) and 59 were in modules unrelated to diagnosis or PRS (**figure 3C**).

We found that MEs of four modules were significantly correlated with both PRS sets 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) (**figure 3B**).

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) (**figure 3A**). Modules with a significant over-representation of PGC2 loci genes were blue, brown, green (p corrected values <0.05) and red, turquoise, yellow (p uncorrected values <0.05) (**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**.

Tier 1 genes: PGC2 loci genes over-represented in modules associated with both diagnosis and PRS

We selected as **tier 1** 36 PGC2 loci protein-coding genes that were represented in the brown module (**figure 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].

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

We selected 53 PGC2 genes as tier 2 genes, with a trend of over-representation in turquoise and yellow modules (**supplementary table 4**). However, we specify that PGC2 enrichment did not reach the levels of significance after FWE correction for multiple comparisons (uncorrected values were $p=0.023$ for yellow, respectively $p=0.01$ for turquoise (**supplementary table 4**). As previously mentioned these modules represented mixtures of cellular markers (**supplementary table 3**) and were not significantly enriched for GO-BP. However, we considered them interesting modules based on the PGC2 genes present in these modules. One of the notable tier 2 genes in this module is ZNF804A [33]. Other interesting members of tier 2 genes are NDUFA13, NDUFA4L2, NOSIP, and NRGN.

Tier 3 genes: PGC2 genes over-represented across networks

59 PGC2 GWAS significant loci genes belonged to **tier 3**: they were over-represented 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 (**figure 2**). Interesting tier 3 genes are CACNA1C, CHRM4, FURIN, TSNARE1 (possibly implicated in neurotransmitter release by regulating the SNARE- Synaptotagmin complex), 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 5**.

Finally, 161/ 309 PGC2 locus genes were randomly distributed in modules that did not pass the threshold of significance for PGC2 enrichment or were “grey” genes (not assigned to a module in any network). A notable “grey” 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, because of their association at least with illness diagnosis, are more attractive candidates for experimental studies to decipher pathophysiological mechanisms of SCZ or drug development.

DISCUSSION:

We have performed an extensive analysis of gene co-expression architecture in adult postmortem DLPFC from CAUC unaffected donors (CTRL) and from donors diagnosed with schizophrenia (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 the polygenic risk for schizophrenia. 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 abnormalities in SCZ (i.e. DLPFC) with both the illness state and with genomic risk for the illness is a more optimal strategy to isolate potentially co-functional gene sets that could be investigated as harboring novel drug targets in 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 over-represented in modules of co-expression

by the modules' relationship with polygenic risk score and diagnosis of schizophrenia. 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.

Tier 1 genes

We identified 36 PGC locus genes that were distributed in one module (brown), significant for both diagnosis and association with genomic risk, i.e. PRS (**figure 3**). In principle, these genes and this network should bear an especially close relationship to schizophrenia pathogenesis and pathobiology. These genes are members of putatively relevant signaling pathways, such as PIK3/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 Ca²⁺ signaling pathway with numerous functions, including energetic metabolism [32] that was underscored in our study mainly by ATP2A2; a RAS/ERK signaling pathway represented by RRAS (RAS related small GTPase) implicated in cell adhesion and axon guidance [32].

Interestingly, when we interrogated STRING, a database of known and predicted protein-protein interaction [35] we 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 serotonergic receptors modulated by GSK3B (i.e. HTR2A). Another PGC2 locus gene, ATP2A2 (ATPase Sarcoplasmic/ Endoplasmic Reticulum Ca²⁺ Transporting), and its interactors (CALM1, RYR2, ITPR1) indicate potential dysfunctions on the Ca²⁺ Signaling Pathway in relation to energetic metabolism [32, 35]. NLGN4X (Neurologin 4 X linked) previously implicated in autism and some of the interactors also

members of the brown module (NRXN3, DLGAP1, DLG2, GRM1, GRM5) are constituents of the post-synaptic density and regulators of glutamatergic signaling [32, 35]. We highlight also RIMS1, co-expressed in 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) 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 intra-modular 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 which showed cross-talk between PIK3/AKT and RAS/ERK signaling pathways that regulates neurodevelopmental processes and synaptic plasticity [34]. This expected cross-talk between signaling pathways has been highlighted in a recent proposal about the underlying polygenic architecture of complex clinical syndromes [36].

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 implicated 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 (Eukariotic 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 one 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), COMT, and RGS4. The inconsistent results of previous studies on some of these genes, coupled with the lack of confirmation by GWAS, prompted a 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 genetic risk 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 black 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.

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] (**figure 2**). 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. Interestingly, we found that 19 of the 31 PGC2 genes over-represented in their module of interest- M2C- were members of at least one of our tier genes. Moreover, nine of these 19 genes were actually tier 1 PGC2 genes in our study (**figure 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 (2018) [13] and our modules of co-expression (**supplementary table 6**).

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 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. 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 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. 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 grey, 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 the 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 to genetic factors, but also to an environmental contribution. Most importantly, our study highlights complex relationships between gene co-expression networks in the brain and the polygenic risk for SCZ and provides a strategy for using this information in selecting potentially targetable gene sets for therapeutic drug development.

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.

Conflict of interest: The authors declare no conflict of interest.

BIBLIOGRAPHY:

1. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 2014 July 24;511(7510):421-7.
2. Jaffe AE, Straub RE, Shin JH, Tao R, Gao Y, Torres LC, et al. Developmental And Genetic Regulation Of The Human Cortex Transcriptome In Schizophrenia. *bioRxiv*. 2017; doi: <https://doi.org/10.1101/124321>.
3. 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 January 01;19(1):48-54.
4. Webber C. Epistasis in Neuropsychiatric Disorders. *Trends Genet*. 2017 April 01;33(4):256-65.
5. Birnbaum R, Weinberger DR. Genetic insights into the neurodevelopmental origins of schizophrenia. *Nat Rev Neurosci*. 2017 December 01;18(12):727-40.
6. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008 December 29;9:559.
7. 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 November 01;11(11):1271-82.
8. 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 May 25;474(7351):380-4.
9. 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 December 01;18(12):1308-14.
10. Torkamani A, Dean B, Schork NJ, Thomas EA. Coexpression network analysis of neural tissue reveals perturbations in developmental processes in schizophrenia. *Genome Res*. 2010 April 01;20(4):403-12.
11. 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 December 01;69(12):1205-13.

12. 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 May 30.
13. 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 February 09;359(6376):693-7.
14. 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 November 01;19(11):1442-53.
15. 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 January 17;7(1):e1006.
16. BrainSeq: A Human Brain Genomics Consortium. BrainSeq: Neurogenomics to Drive Novel Target Discovery for Neuropsychiatric Disorders. *Neuron*. 2015 December 16;88(6):1078-83.
17. TopHat: <https://ccb.jhu.edu/software/tophat/index.shtml>
18. Ensembl: <http://www.ensembl.org/index.html>
19. 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 June 01;5(6):e1000529.
20. Delaneau O, Coulonges C, Zagury JF. Shape-IT: new rapid and accurate algorithm for haplotype inference. *BMC Bioinformatics*. 2008 December 16;9:540.
21. 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 September 01;81(3):559-75.
22. Chen Q, Ursini G, Romer AL, Knodt AR, Mezeivitch K, Xiao E, et al. Schizophrenia polygenic risk score predicts mnemonic hippocampal activity. *Brain*. 2018 February 03.
23. WGCNA manual: <https://cran.r-project.org/web/packages/WGCNA/WGCNA.pdf>
24. WGCNA tutorials: <https://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/Rpackages/WGCNA/Tutorials/index.html>

25. 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 U S A*. 2017 July 03;114(27):7130-5.
26. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 2007 January 01;8(1):118-27.
27. Langfelder P, Luo R, Oldham MC, Horvath S. Is my network module preserved and reproducible? *PLoS Comput Biol*. 2011 January 20;7(1):e1001057.
28. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012 May 01;16(5):284-7.
29. <https://www.bioconductor.org/packages/release/data/annotation/html/org.Hs.eg.db.html>
30. 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 January 02;28(1):264-78.
31. Poduri A, Evrony GD, Cai X, Elhosary PC, Beroukhim R, Lehtinen MK, et al. Somatic activation of AKT3 causes hemispheric developmental brain malformations. *Neuron*. 2012 April 12;74(1):41-8.
32. GeneCards: www.genecards.org
33. Chang H, Xiao X, Li M. The schizophrenia risk gene ZNF804A: clinical associations, biological mechanisms and neuronal functions. *Mol Psychiatry*. 2017 July 01;22(7):944-53.
34. Manning BD, Toker A. AKT/PKB Signaling: Navigating the Network. *Cell*. 2017 April 20;169(3):381-405.
35. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*. 2015 January 01;43(Database issue):447.
36. Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell*. 2017 June 15;169(7):1177-86.
37. 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 October 01;71(10):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 March 01;21(5):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 April 23;10(4):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 March 01;174(3):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 May 01;20(5):555-62.
42. Kabir ZD, Martinez-Rivera A, Rajadhyaksha AM. From Gene to Behavior: L-Type Calcium Channel Mechanisms Underlying Neuropsychiatric Symptoms. *Neurotherapeutics.* 2017 July 01;14(3):588-613.
43. Lencz T, Malhotra AK. Targeting the schizophrenia genome: a fast track strategy from GWAS to clinic. *Mol Psychiatry.* 2015 July 01;20(7):820-6.

FIGURES LEGENDS:

Figure 1: Analysis pipeline and criteria for selection and prioritization of gene sets associated with the genetic risk of schizophrenia. **A1:** 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- snpPCs) by using *empiricalBayesLM* function (WGCNA, version 1.61 (23)). **A2:** Co-expression network was constructed with WGCNA embedded routines (*blockwiseModules*) by applying the following parameters and procedures: bi-weight mid-correlation with “signed” networks to allow for potentially non-linear correlations between genes (24); β power=12 selected with “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); **A3:** 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; **A4:** Post-hoc WGCNA limited to 9239 genes organized in nine modules of interest after the primary network construction was performed. **A5:** External validation by using the *modulePreservation* function in WGCNA (23, 24), which 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); 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 (27). **A6:** Enrichment in meaningful biological ontologies was tested with *enrichGO* function from the R package *clusterProfiler* (28): hypergeometric test was applied to test for over-representation 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* (29)); **A7:** Enrichment in cellular markers was tested by using the hypergeometric test and gene lists embedded in the *userListEnrichment* function (24, 7, 30); **B:** 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).

Figure 2: Top twenty most connected genes in seven “core” modules significantly enriched for biological processes relevant for cortex structure and functionality. Modules are represented by their colors assigned through co-expression network construction and module detection. *Legend:* * modules reported by Oldham et al (2008); modules’ representation created with VisAnt (<http://visant.bu.edu/>).

Figure 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. *Legend:* * these genes were members of two modules that shortly failed the significance for PGC2 loci genes enrichment after FWE multiple comparisons correction.

TABLES:

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

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

Modules and top 10 GO-BP enrichments	
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

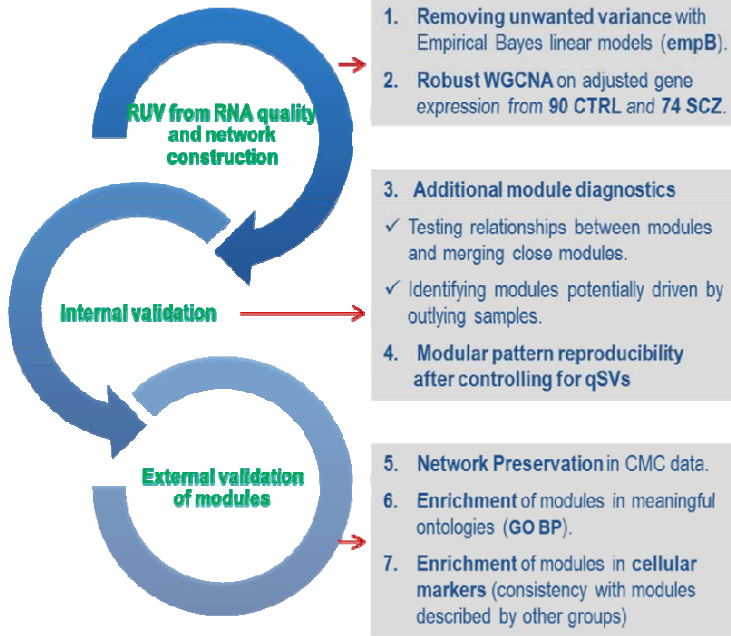
Table 2: Summary of the relationships between co-expression modules and diagnosis, PRS and intra-modular PGC2 loci genes distribution

Modules	Association with dx	Association with PRS	Enrichment in PGC2 loci genes
Black	yes	yes	no
Blue ³	no	no	yes
Brown ¹	yes	yes	yes
Cyan**	yes	yes	no
Green ³	no	no	yes
Greenyellow	no	no	no
Magenta	no	no	no
Midnightblue*	yes	yes	no
Red	±	no	p _{uncorr} <0.05
Salmon	±	no	no
Turquoise ²	yes	±	p _{uncorr} <0.05
Yellow ²	yes	no	p _{uncorr} <0.05

Legend: ¹Module used for tier 1 genes selection; ²Modules used for tier 2 genes selection; ³Modules used for tier 3 genes selection; *small-sized module with no GO-BP enrichment and weakly preserved in CMC data; **module with no GO-BP enrichment.

Figure 1

A. Data processing and co-expression network analysis (WGCNA)



B. Identification and prioritization of PGC2 gene sets by membership in co-expression modules

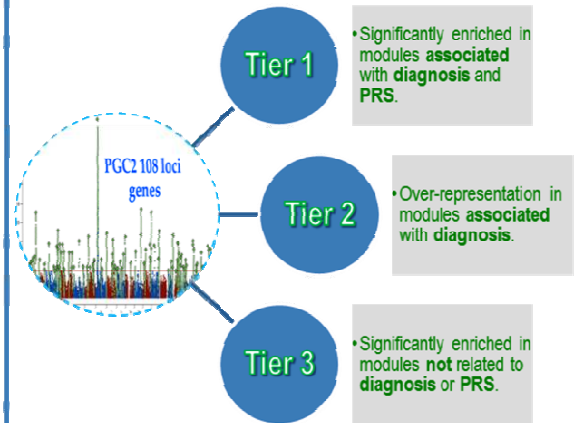
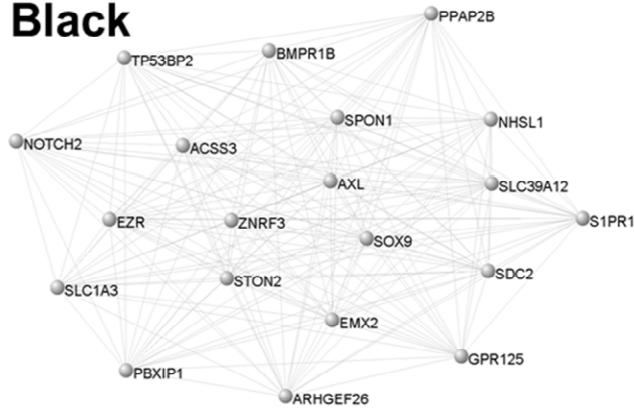
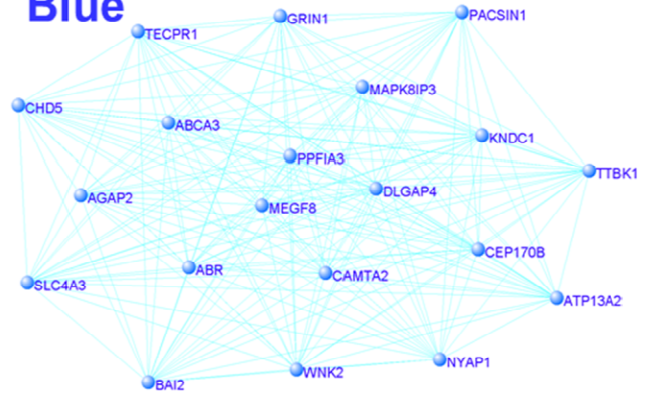


Figure 2:

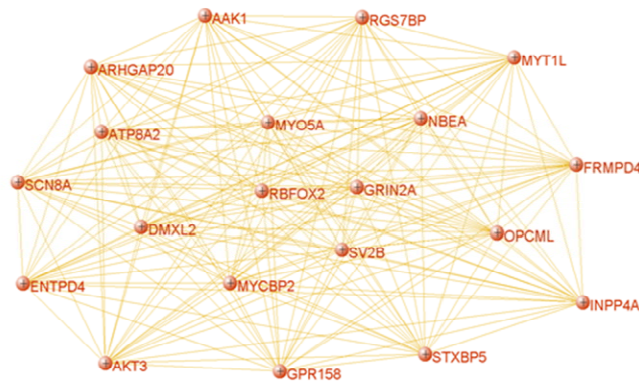
Black



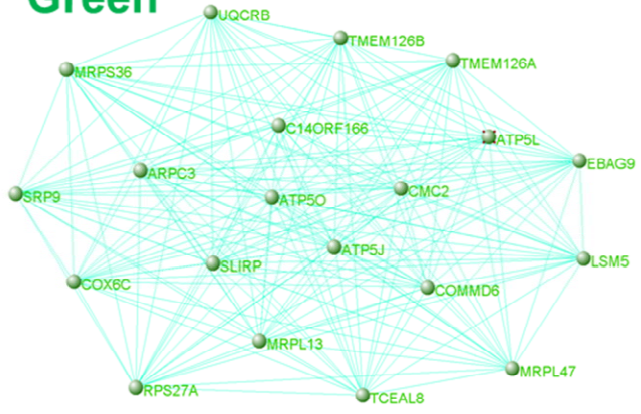
Blue



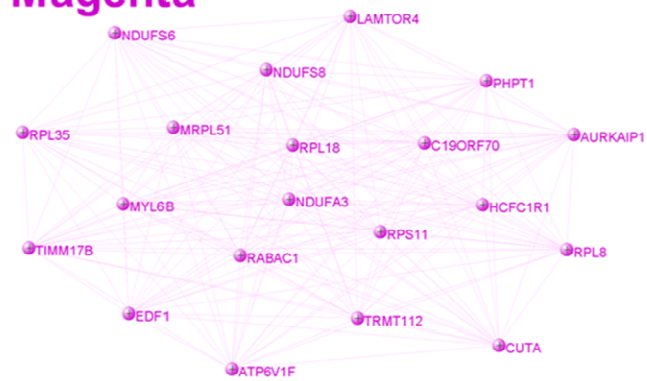
Brown



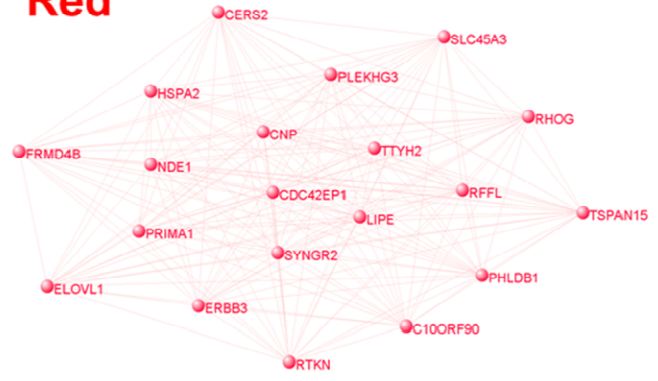
Green



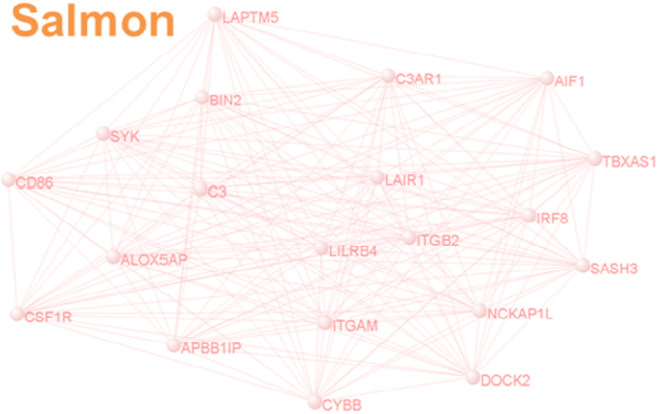
Magenta



Red



Salmon



Black: Astrocytes \leftrightarrow M15_CTX*

Blue: Neuron \leftrightarrow M10_CTX*

Brown: Neuron \leftrightarrow M16_CTX*

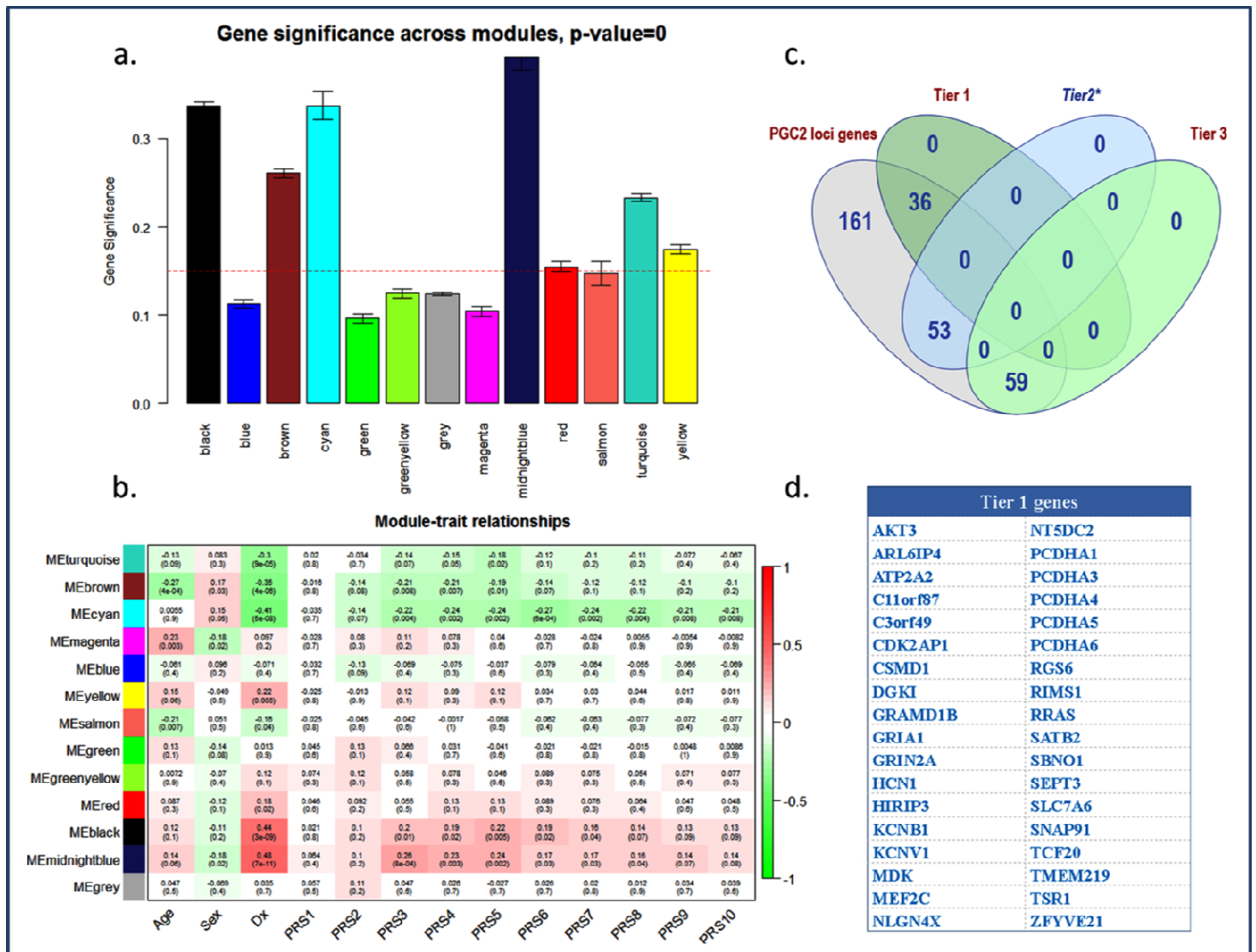
Green: Mitochondria \leftrightarrow M18_CTX*

Magenta: Mitochondria \leftrightarrow M7_CTX*

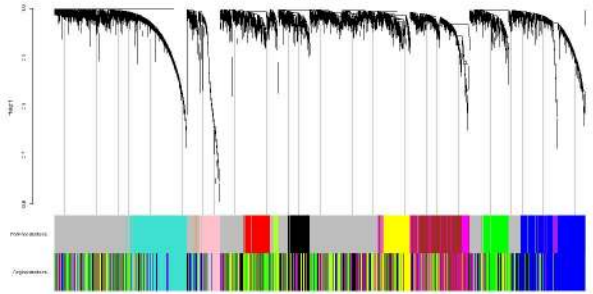
Red: Oligodendrocytes \leftrightarrow M9_CTX*

Salmon: Microglia \leftrightarrow M4_CTX*

Figure 3:



Cluster Dendrogram



20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000



Figure 1

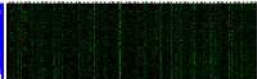
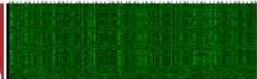
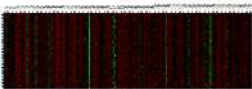
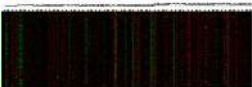


Figure 2





gms1



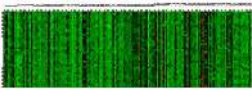
gms1b



sample



microRNAs



cell



Figure 1



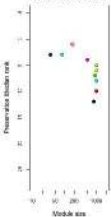
Figure 2



Figure 3



Preservation Median rank



Preservation Journey

