bioRxiv preprint doi: https://doi.org/10.1101/288225; this version posted September 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

csuWGCNA: a combination of signed and unsigned WGCNA to capture negative correlations

Rujia Dai[1], Yan Xia[1], Chunyu Liu[1,2*], Chao Chen[1,3*]

1Center for Medical Genetics, School of Life Sciences, Central South University, Changsha, China
2Department of Psychiatry, SUNY Upstate Medical University, NY, USA
3National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University

* Corresponding author
E-mail: chenchao@sklmg.edu.cn
liuch@upstate.edu

Abstract

In original Weighted Gene Co-expression Network Analysis (WGCNA), the signed network considers the sign of correlation and only positive correlations make sense in the network. The unsigned network regard both highly positive and negative correlations as connected. This design results in loss of negative correlation in the signed network and moderate negative correlations in the unsigned network. To avoid these limitations, we provided a modified method of WGCNA named Combination of Signed and Unsigned WGCNA (csuWGCNA). We created networks for signed, unsigned and csuWGCNA on two gene expression datasets of the human brain from Stanley Medical Research Institute (SMRI) and BrainGVEX. The results obtained from our investigation indicate that our method is better than signed and unsigned WGCNA in capturing negatively correlated gene pairs. Especially for the relationship between miRNA, lncRNA and their target genes.

Introduction

Gene co-expression analysis is a tool for identifying important gene relationships[1]. WGCNA is the common method used in co-expression analysis[2].

bioRxiv preprint doi: https://doi.org/10.1101/288225; this version posted September 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The original WGCNA utilize the correlation between genes. Considering a gene expression matrix G_{mxn}, where m is the number of genes and n is the number of samples. The original WGCNA procedure generates a correlation matrix S between genes in G with two methods: Pearson and bicor at first. A parameter β is chosen to enable the network show a Scale-Free Topology (SFT) property. After that, the adjacency matrix A is constructed from S depending on whether the adjacency is signed or unsigned. In signed adjacency matrix correlations in the [−1, 1] interval is scaled into the [0, 1] interval and negative correlations are made positive in the unsigned adjacency matrix. The adjacency is defined as following for signed, unsigned and signed hybrid respectively. The adjacency a_{ij} for gene i and j is:

$$signed\ a_{ij} = |(1 + \mathrm{cor}(x_i, x_j))/2|^{\beta}$$

$$unsigned\ a_{ij} = |\mathrm{cor}(x_i, x_j)|^{\beta}$$

$$signed\ hybrid\ a_{ij} = \begin{cases} |\mathrm{cor}(x_i, x_j)|^{\beta} & \mathrm{cor}(x_i, x_j) > 0 \\ 0 & \mathrm{cor}(x_i, x_j) > 0 \end{cases}$$

From the adjacency matrix, a new matrix with the same dimension is created. The Topological Overlap Matrix (TOM) is created in to make networks less sensitive to spurious connections or to connections missing due to random noise. Once TOM is built, the hierarchical clustering is performed on the matrix 1-TOM. A dynamic tree-cut function is applied to the dendrogram to get the module of highly co-expressed genes. Module eigengene is the first principal component of the gene expression of genes clustered into this module which summarizes the whole module. By looking the correlation between module eigengene with traits, we can get modules linked to biological meaning such as disease, age[3], sex[4], cell type[5,6] and disease state[7] etc.

The signed and unsigned network created by original WGCNA has distinct features. Unsigned networks just consider the correlation of two genes and the sign of the correlation doesn't matter. This treats the positive and negative correlations fairly. In reverse, the strongly negative correlation is considered as no connection. The creator of the WGCNA recommends signed network for following two reasons. First, more often than not, direction does matter. Second, negatively correlated nodes often belong to different categories. Moreover, a study[8] of embryonic stem (ES) cell recommended signed network because the analysis shows that signed WGCNA identifies modules with more specific expression patterns than unsigned WGCNA. However, the signed network just ignores negative correlation directly. For the study focusing on negative gene correlation, using signed network will result in loss of information. Therefore, there is a need to create a method combining signed and unsigned WGCNA.

miRNA and lncRNA are two types of non-coding RNA which were reported a to

bioRxiv preprint doi: https://doi.org/10.1101/288225; this version posted September 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

be negatively correlated with target genes[9]. miRNAs can regulate the gene transcription and inhibit the translation of mRNA[10-12]. Brain-specific miRNA miR-134[13] was reported to inhibit Limk1 translation and in this way may contribute to synaptic development. The function of lncRNA has been implicating in post-transcriptional regulation, splicing regulation, regulate protein localization. BDNF-AS is the natural antisense transcript to BDNF[14], itself a key contributor to synaptic function. By dynamically repressing BDNF expression in response to neuronal depolarization, BDNF-AS modulates synaptic function. Apparently, miRNA and lncRNA played important roles in gene repression program.

In this study, we point out the disadvantages of signed and unsigned WGCNA and create a new method combining their advantages. We used two gene expression profiles from the human brain to prove our method csuWGCNA can capture more negative miRNA-target and lncRNA-gene pairs. Also, the results indicate that csuWGCNA can found more validated negative gene pairs and more significant gene/pathway enrichment.

Results

Signed WGCNA captured more specific module but the negative correlations were lost

We first compared the similarity definition of signed and unsigned network. In the unsigned network, the similarity between two genes was defined as the absolute value of Pearson correlation between expression of genes. However, the definition of signed network reflected the sign of the correlation (Figure 2). The signed network was recommended for the reason that using the absolute value of the correlation may obfuscate biologically relevant information since no distinction was made between gene repression and activation. However, the distinction in the signed network was based on positive correlation taking precedence over the negative correlation. The strong repression between two genes was regarded as no similarity.

To confirm this point, we reanalyzed the data from embryonic stem (ES) cell[15] which used in explaining signed network is better than unsigned network before. As the original study reported[8], signed network identified the pluripotency-related module (Figure 1A). This small module was hidden in a large module in the unsigned network. We reanalyzed the data and we do observe the module relationship reported in the original study. A core group of ES-related transcription factors (TFs) is enriched signed brown module (Figure 1B). In an unsigned network, this TFs are scattered in the blue module that larger than the signed brown module. However, we found that the negative correlation in the unsigned blue module was lost in the two-separate signed module (Figure 1C).

bioRxiv preprint doi: https://doi.org/10.1101/288225; this version posted September 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The unsigned blue module found 139129 negative pairs and signed brown module and turquoise module only found 22 pairs in total. The result indicates that signed WGCNA identifies modules with more specific expression patterns than unsigned WGCNA but it lost a lot of negative correlations.

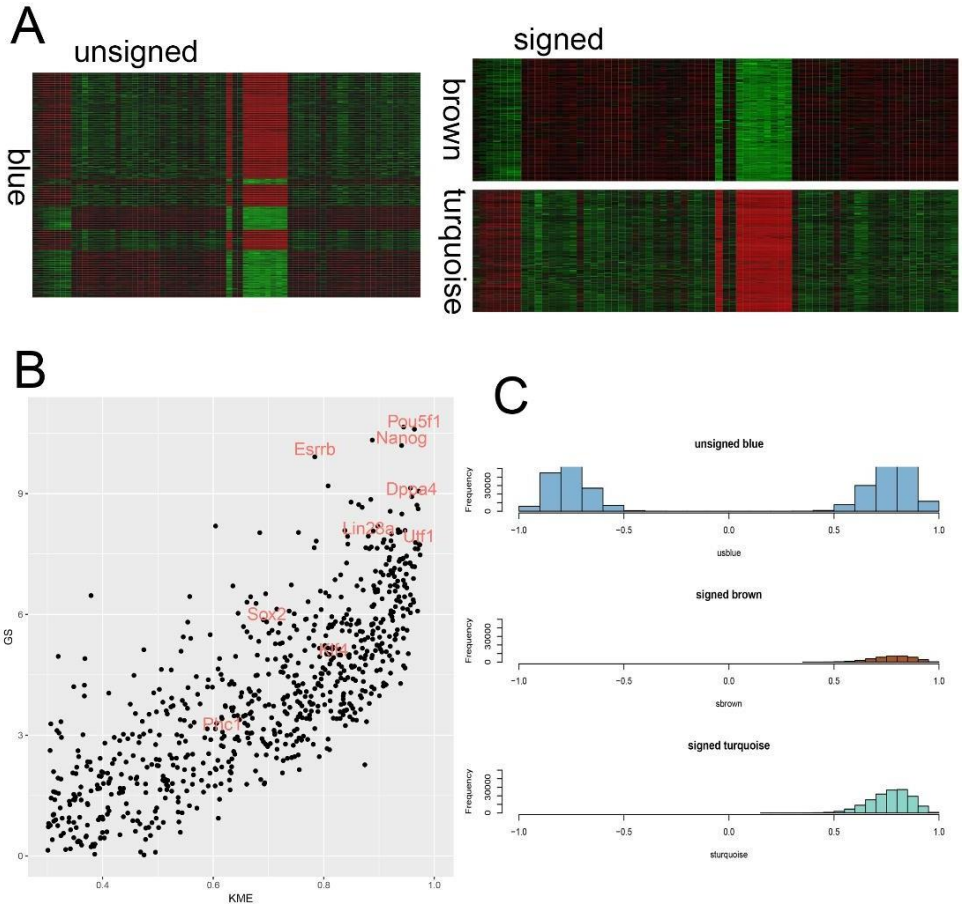

Figure 1 Signed and unsigned ES-related module identified Ivanova et al. A. heat map for visualizing standardized gene expressions (rows) across samples (columns) for genes in the blue module in the unsigned network, brown and turquoise module in the signed network. B. plot of module membership, kME, (x-axis) plotted against gene significance, GS, (y-axis) for the brown in the signed network with known ES-related TF labeled. C. The distribution of gene correlation in the unsigned blue module and signed brown turquoise module.

Combination of signed and unsigned WGCNA

To combined the features of signed and unsigned network, we proposed a new method termed csuWGCNA. The core modification of csuWGCNA is the

bioRxiv preprint doi: https://doi.org/10.1101/288225; this version posted September 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

definition of adjacency network, which integrates the advantages of signed network and unsigned network (Figure 2). The adjacency of csuWGCNA is calculated as follows:

$$\text{csu } a_{ij}=((1+|cor(x_i,x_j)|)/2)^{\beta}$$

With this calculation of adjacency, the strong and weak negative correlations are taken into account. Meanwhile, the positive correlation remains the same as they in signed network. We modified two functions for picking soft thresholding power and calculating network adjacency. The whole process of csuWGCNA includes adjacency calculation based on similarity matrix, the topological overlap Matrix (TOM) construction, hierarchical clustering, dynamic tree cutting and module merging.

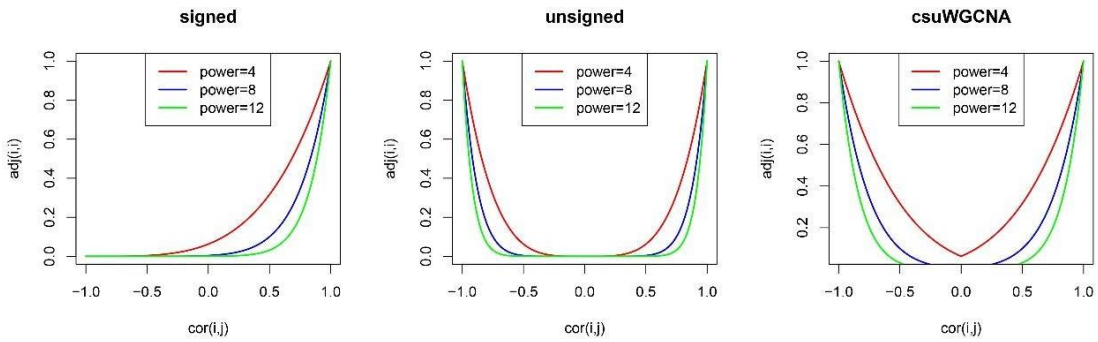

Figure 2 Network adjacency versus gene expression correlation. Network adjacency (y-axis) versus correlation (x-axis) for weighted networks in the signed network, unsigned network, and csuWGCNA network. The color of the line denotes the power used. Note that correlation=-1 leads to adjacency = 0 in the signed network and adjacency =1 in the unsigned and csuWGCNA network.

The csuWGCNA can detect modules containing genes with negative correlations, which may be more useful when lncRNAs and miRNAs are included in the network. We applied the csuWGCNA on the one miRNA datasets from SMRI and another lncRNA data sets from BrainGVEX to compare it with signed and unsigned WGCNA (Table 1).

miRNA-mRNA

We performed signed WGCNA, unsigned WGCNA and csuWGCNA on the mRNA and miRNA expression in parietal cortex tissues from the SMRI samples of patients with schizophrenia (SCZ), bipolar disorder (BD) and healthy controls. Signed WGCNA detected 13 modules, unsigned WGCNA detected 20 modules and csuWGCNA detected 15 modules.

Firstly, csuWGCNA identified more informative gene pairs. The Pearson correlation between genes detected in this data set is calculated (Figure 3A).

bioRxiv preprint doi: https://doi.org/10.1101/288225; this version posted September 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We classed the gene pairs into negative pairs(cor<-0.3) and positive pairs (cor>0.3) by the correlation. To compare three networks, we defined informative gene pairs as they are located in the same module (SM). If gene pairs are in different modules(DM), their correlation or relationship makes no sense. The pairs contain genes in the grey module are removed. For the negative pairs, csuWGCNA captured 33% SM pairs which are better than unsigned WGCNA (Figure 3B). However, the signed network failed to detect the SM negative pairs. For the positive pairs, the signed WGCNA performed best which captured 64% SM pairs. csuWGCNA in the second place and unsigned WGCNA is the worst. Overall, csuWGCNA can capture more informative gene pairs (84%) compare to signed and unsigned WGCNA.

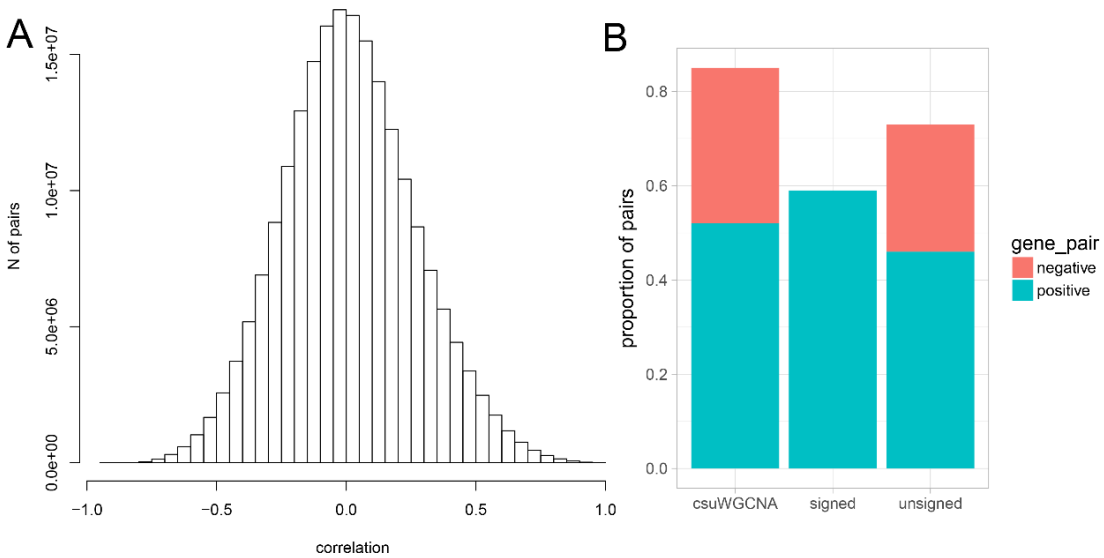

Figure 3 Correlation of gene expression of SMRI data. A. The distribution of gene expression in SMRI data. B. The proportion positive and negative gene pairs located in the same module in csuWGCNA, signed and unsigned network.

Secondly, more miRNA and target gene are captured by csuWGCNA. To examined the capability of WGCNA for capturing miRNA-target interaction (MTI), we downloaded the MTIs for human from miRTarBase[16]. In total, 101493 MTIs were involved in the analysis. We divided the MTI into two classes according to the sign of correlation: positive MTI and negative MTI. Firstly, the distribution of correlation of MTIs is symmetric and most of the correlation are weak (Figure 4A). To compare three networks, we counted the SM MTIs for both positive and negative correlation. The result showed that no matter negative or positive, csuWGCNA captured the most SM MTIs (Figure 4B). The overlap of negative MTIs between three networks showed that even csuWGCNA detected the most MTIs, the number of MTIs both unsigned WGCNA and csuWGCNA found in common is less than either of their own (Figure 4C). So, we suppose whether the csuWGCNA and unsigned WGCNA

bioRxiv preprint doi: https://doi.org/10.1101/288225; this version posted September 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

captured different types of negative MTIs. We then classified the negative MTIs into three classes: common, csuWGCNA_specific, unsigned_specific. The highly negative MTIs (cor<-0.5) are captured by both unsigned and csuWGCNA. Meanwhile, the csuWGCNA captured a great deal of weakly negative MTIs while unsigned WGCNA tends to capture moderate negative MTIs (Figure Figure 4D).

Thirdly, csuWGCNA finds more known repression relationship from KEGG database[17]. We derived 10198 gene pairs from KEGG which is repression or inhibition relationship. We counted the pairs in SM for three networks. Figure 4E showed that the csuWGCNA captured 20% repression/inhibition gene pairs while signed and unsigned WGCNA only captured ~13% pairs.

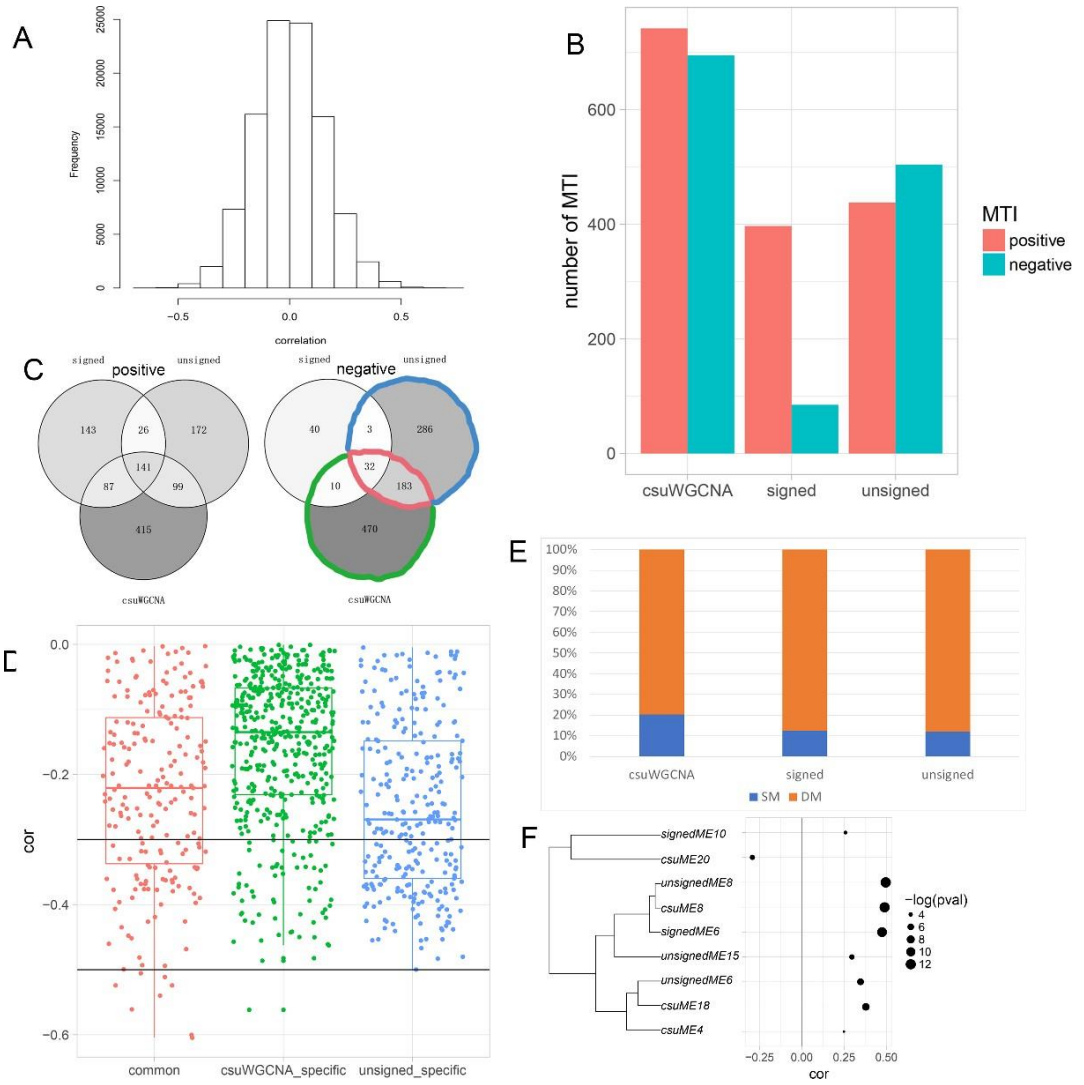

Figure 4 Summary of MTIs in the co-expression of SMRI data. A. the distribution of correlation of MTIs from miRTarBase. B. The number of positive and negative MTIs located in the same module in csuWGCNA, signed and unsigned network. C. The positive and negative MTIs captured by three networks. D. The correlation of negative MTIs captured by unsigned and csuWGCNA. According to the overlaps in C, the MTIs were divided into three classes:

bioRxiv preprint doi: https://doi.org/10.1101/288225; this version posted September 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

common MTI, csuWGCNA specific MTI and unsigned WGCNA specific. E. The proportion of repression or inhibition gene pairs from KEGG located in the same module or different module in three types of network. F. The disease-related module identified in three networks. The left panel is the cluster of disease-related modules according to the module eigengenes. The right panel is the correlation value and p-value of disease module eigengene with the disease state.

Fourthly, disease-related miRNAs were identified by both unsigned and csuWGCNA. Due to this dataset included BD and SCZ, we found the disease modules (Pearson correlation $p<0.05$). miRNA-320 which was reported to involve in putative regulation in psychiatric disorder in our previous work. We found that miRMA-320b, miRNA-320c, miRMA-320d, and miRMA-320e were captured by disease module in the unsigned WGCNA (ME8) and csuWGCNA (ME8). But signed network failed to capture these important miRNAs (Figure 4F).

Finally, the signed WGCNA and csuWGCNA enriched more significant GO term and pathway than unsigned. We annotated all non-grey modules in three networks with Gene Ontology database and KEGG database. We then aggregate the p-value of all significant terms (FDR<0.05) for a network in a single measurement of significance as follows. For each network, n is the number of significant terms, p is p-value for the term.

$$S_{network} = (\Sigma_1^n(-log10(p)))/n$$

The significance of signed network is 2.55, and the significant of csuWGCNA and unsigned WGCNA is 2.34 and 2.24 separately.

In summary, our analysis on miRNA and mRNA indicates that csuWGCNA is better than both signed and unsigned WGCNA in capturing MTIs and validated repression gene pairs. In the aspect of module enrichment, csuWGCNA is slightly worse than signed WGCNA but better than unsigned WGCNA.

lncRNA-mRNA

LncRNA is another type of non-coding RNA which reported involved in gene repression. We performed three types WGCNA on data including lncRNA from BrainGVEX. The data contain both healthy control and psychiatry patients (SCZ and BD). Signed WGCNA detected 20 modules, unsigned WGCNA detected 25 modules and csuWGCNA detected 16 modules.

In total, 1132 lncRNAs in this dataset including four types: lincRNA, antisense, sense_intronic and sense_overlapping. We correlated the expression of

bioRxiv preprint doi: https://doi.org/10.1101/288225; this version posted September 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

lncRNA with the expression of other genes and picked up the lncRNA-gene pairs which Pearson correlation value lower than -0.3. In total, 273360 lncRNA-gene pairs matched the condition. The number of pairs in the same non-grey module for csuWGCNA and unsigned WGCNA is 7657 and 7391 (Figure 5A). None of the pairs located in the same nodule in signed WGCNA. Now that csuWGCNA found more SM lncRNA-gene pairs, is csuWGCNA capable of finding pairs with stronger correlation? The figure 5B is the boxplot of correlations of negative SM pairs in csuWGCNA and unsigned WCGNA. The plot showed that the csuWGCNA captured lncRNA-genes pairs that more negative (p-value <2.2e-16, t-test).

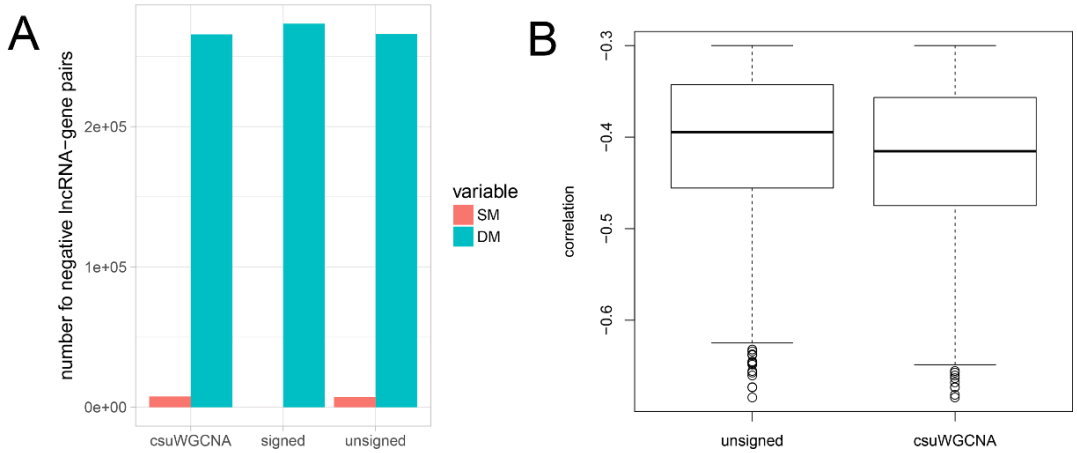

Figure 5 The negatively correlated lncRNA-gene pairs in BraiGVEX data. A. The number of negative lncRNA-gene pairs located in the same module and the different module in csuWGCNA signed and unsigned WGCNA. B. The correlation of negative lncRNA-gene pairs in the same module in the unsigned and csuWGCNA.

The csuWGCNA enriched more significant GO term and pathway than signed and unsigned WGCNA in lncRNA dataset. We annotated all non-grey modules in three networks with Gene Ontology database and KEGG database. We then calculated the significance described in the last section for three networks. The significance of csuWGCNA network is higher than both signed and unsigned WGCNA. (csuWGCNA=2.38, signed=2.25, unsigned=2.19)

Methods

data collection

Table 1 Summary of dataset used

dataset	sample size	brain region	platform
ES cell dataset	70	-	Affymetrix microarrays
SMRI	75	parietal	Affymetrix Human

bioRxiv preprint doi: https://doi.org/10.1101/288225; this version posted September 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

		cortex	Gene 1.0 ST Array, Illumina sequencing
BrainGVEX	413	frontal cortex	Illumina sequencing

samples and quality control

SMRI data: Parietal cortex tissue specimens from the Stanley Medical Research Institute (SMRI) Neuropathology Consortium and Array collections included SCZ, BD and control samples. The non-Europeans, replicates, and samples missing any of the mRNA, miRNA and genotyping results were removed. After filtering, we retained 75 samples, yielding data for 19,984 mRNAs and 470 miRNAs.
BrainGVEX data: Frontal cortex samples were collected from the PsychENCODE[18] project. The sample including 260 health control, 76 BD samples and 94 SCZ samples.

gene profiling and data pre-processing

ES cell: The raw data were downloaded from Ivanova et al DataSet. MAS5 was used to process raw data and then the data were log2 transformed. We removed duplicated genes according to variance. Finally, 13627 genes and 70 samples are kept.

SMRI: Total RNA was extracted from PC tissue using the RNeasy Mini kit (Qiagen, Hilden, Germany). The concentration and A260/A280 ratio were measured on the NanoDrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA). The 28S:18S rRNA ratio and RNA Integrity Number (RIN) were measured using an RNA LabChip kit on the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA). Only RNA samples with a RIN > 6 were used for expression profiling. Total RNA was extracted from tissues using the mirVana miRNA Isolation Kit (Ambion, Austin, TX) according to the manufacturer's instructions. RNA quality and the presence of small RNAs were inspected on a 2100 Bioanalyzer (Agilent Technologies). After strict RNA quality assurance, 15 µg of total RNA was used for small RNA library creation using Illumina's DGE small RNA sample prep kit per the manufacturer's instructions. Purified cDNA was quantified with the Quant-iT PicoGreen dsDNA Kit (Thermo Fisher Scientific) and diluted to 3 pM for sequencing on the Illumina 1G Genome Analyzer (University of Houston). Each library was sequenced in a single lane.

Affymetrix Human Gene 1.0 ST Array (Affymetrix, Santa Clara, CA) was used for whole-genome transcriptome profiling at the NIH Neuroscience Microarray

bioRxiv preprint doi: https://doi.org/10.1101/288225; this version posted September 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Consortium facility at Yale University. Single nucleotide polymorphisms (SNPs) in probe regions can affect probe hybridization efficiency. For the Robust Multichip Average (RMA) preprocessing steps[19]: background correction, quantile normalization, and gene level summarization6. Afterward, for the convenience of comparison, only genes with Entrez IDs were kept.

Sequence reads with 36-nt read length were picked for miRNA mapping. Reads that did not pass the Illumina chastity and no-calls filter were removed. FastQC (v0.11.2) was used to check for homopolymers, adapters, and distribution of base quality. After trimming for adapters, sequences that read length < 10nt, copy number < 4, or more than 10 consecutive, repetitive nucleotides were discarded. The miRBase database release was used to identify miRNAs, and Bowtie 2 was used for mapping. Average valid sequence reads were 15M in each sample, and the total count was used for sample-wise normalization. ComBat, a batch effect adjustment program was used to remove batch effects from both miRNA and mRNA data sets.

BrainGVEX:

Total RNA was isolated at the University of Illinois at Chicago and the University of Chicago with the Qiagen miRNeasy mini kit. Approximately 50mg fresh-frozen brain tissue was homogenized by the FastPrep-24 system in QIAzol Lysis Reagent with Lysing Matrix D, then mixed well with chloroform. The separated aqueous layer was recovered, mixed with ethanol and applied to a miRNeasy mini column. Columns were treated with Qiagen RNase-free DNase digestion set, then washed with the appropriate miRNeasy mini kit buffers. Total RNA was eluted with RNase-free water. Total RNA was quantified by either Qubit 2.0 RNA BR assay kit or Xpose spectroscopy; the quality of total RNA was assayed by Agilent RNA 6000 Nano Kit on the Agilent Bioanalyzer. Total RNA samples that pass QC to library generation have a concentration of >= 100ng/uL assayed by Qubit 2.0 RNA BR Assay or Xpose, and RIN score >= 5.5 assayed by Agilent Bioanalyzer RNA 6000 Nano assay kit.

All total RNA from BSHRI collections were processed into rRNA-depleted stranded libraries for sequencing on the Illumina HiSeq2000 using the TruSeq Stranded Total RNA sample prep kit with Ribo Zero Gold HMR. For some libraries, 2ul of 1:100 ERCC RNA ExFold Spike-In Mix 1 was added to total RNA starting material before ribo-depletion step as an internal way of tracking library prep and sequencing quality. Libraries are PCR amplified for 12 cycles and cleaned with 0.60X Ampure XP beads.

Libraries Quality Control were processed at the University of Chicago HGAC by quantification with the Qubit 2.0 dsDNA HS assay kit and quantification and quality check with the Agilent Bioanalyzer DNA HS assay kit. Libraries were sequenced on Illumina's HiSeq2000 on a high output flow cell for 100bp PE sequencing. Libraries are 3-plexed per lane to reach 40M paired-end reads per library.

bioRxiv preprint doi: https://doi.org/10.1101/288225; this version posted September 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fastq files go through adapter removal using cutadapt, then the resulting adapter-trimmed fastq files are checked for quality using FastQC. A subset of 10,000 reads is used to estimate insert mean size and standard deviation for use with Tophat. Tophat is used to aligned trimmed reads to the GENCODE19 reference. Expression level is then calculated using HTSeq and Cufflinks with custom scripts used to summarize proportion of reads assigned to each RNA type. The genes that FPKM lower than 1 in more than 60% samples were dropped. We did co-expression on the samples and sample that z-score normalized connectivity with other samples lower than -2 were removed. Finally, 413 sample and 14865 genes were kept. Then, FPKM was log2 transformed. The linear regression was used to remove the effect of covariates including age, sex, RIN, PMI, brain bank, batches, principal components of sequencing (seqPC) except group. The seqPCs were top 10 principal components of PCA on sequencing statistics. The interaction between covariates was calculated.

network construction

ES cell: The normalized data were used to construct signed and unsigned network. The corType is Pearson correlation. The soft power for signed and unsigned is 12 and 7. Other parameters are as follows: TOMtype is signed, deepSplit is 2, minimum module size is 30 and mergeCutHeight is 0.15.

SMRI: We performed signed, unsigned and csuWGCNA on the miRNA-mRNA data. We applied bicor function to calculate the correlation between genes. The soft power picked up for signed, unsigned, and csuWGCNA is 12, 5 and 12 separately. The parameters are as follows: TOMtype is signed, deepSplit is 2, minimum module size is 30 and mergeCutHeight is 0.15, pamStage is true.

BrainGVEX: We performed signed, unsigned and csuWGCNA on the lncRNA-mRNA data. We applied bicor function to calculate the correlation between genes. The soft power picked up for signed, unsigned, and csuWGCNA is 12, 4 and 10 separately. The parameters are as follows: TOMtype is signed, deepSplit is 4, minimum module size is 40 and mergeCutHeight is 0.2, pamStage is false. cutreeHybrid function was used to cut the gene tree.

repression/inhibition gene relationships from KEGG

The KGML files for human species were downloaded from KEGG website. The R package KEGGgraph[20] was used to operate the KGML file and extract the gene relationship. We chose the gene pairs which subtype of relationship is

bioRxiv preprint doi: https://doi.org/10.1101/288225; this version posted September 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

'inhibition' or 'repression'.

GO and KEGG annotation

The annotation of modules was achieved with goProfileR[21]. The terms used for annotation from both GO and KEGG database. The parameter setting as follows: max_set_size=500, correction_function='fdr', hier_filtering='strong'.

Acknowledgement

This work was supported by NSFC grants 81401114, 31571312, the National Key Plan for Scientific Research and Development of China (2016YFC1306000), and Innovation-Driven Project of Central South University (No. 2015CXS034，2018CX033) (to C.Chen), and NIH grants 1U01 MH103340-01, 1R01ES024988 (to C.Liu). All the data contributors are sincerely thanked for data supported. The authors thank members of the Chunyu and Chao laboratory for critical reading of this manuscript.

Reference

1 Parikshak, N. N., Gandal, M. J. & Geschwind, D. H. Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders. *Nature reviews. Genetics* **16**, 441-458, doi:10.1038/nrg3934 (2015).

2 Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics* **9**, 559, doi:10.1186/1471-2105-9-559 (2008).

3 Forabosco, P. *et al.* Insights into TREM2 biology by network analysis of human brain gene expression data. *Neurobiology of aging* **34**, 2699-2714, doi:10.1016/j.neurobiolaging.2013.05.001 (2013).

4 Oldham, M. C. *et al.* Functional organization of the transcriptome in human brain. *Nature neuroscience* **11**, 1271-1282, doi:10.1038/nn.2207 (2008).

5 Miller, J. A., Woltjer, R. L., Goodenbour, J. M., Horvath, S. & Geschwind, D. H. Genes and pathways underlying regional and cell type changes in Alzheimer's disease. *Genome medicine* **5**, 48, doi:10.1186/gm452 (2013).

6 Winden, K. D. *et al.* The organization of the transcriptional network in specific neuronal classes. *Molecular systems biology* **5**, 291, doi:10.1038/msb.2009.46 (2009).

7 Miller, J. A., Horvath, S. & Geschwind, D. H. Divergence of human and mouse brain transcriptome highlights Alzheimer disease pathways. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 12698-12703, doi:10.1073/pnas.0914257107 (2010).

bioRxiv preprint doi: https://doi.org/10.1101/288225; this version posted September 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

8 Mason, M. J., Fan, G., Plath, K., Zhou, Q. & Horvath, S. Signed weighted gene co-expression network analysis of transcriptional regulation in murine embryonic stem cells. *BMC genomics* **10**, 327, doi:10.1186/1471-2164-10-327 (2009).

9 Yoon, J. H., Abdelmohsen, K. & Gorospe, M. Functional interactions among microRNAs and long noncoding RNAs. *Seminars in cell & developmental biology* **34**, 9-14, doi:10.1016/j.semcdb.2014.05.015 (2014).

10 Valencia-Sanchez, M. A., Liu, J., Hannon, G. J. & Parker, R. Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes & development* **20**, 515-524, doi:10.1101/gad.1399806 (2006).

11 Filipowicz, W., Bhattacharyya, S. N. & Sonenberg, N. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nature reviews. Genetics* **9**, 102-114, doi:10.1038/nrg2290 (2008).

12 Kim, D. H., Saetrom, P., Snove, O., Jr. & Rossi, J. J. MicroRNA-directed transcriptional gene silencing in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 16230-16235, doi:10.1073/pnas.0808830105 (2008).

13 Schratt, G. M. *et al.* A brain-specific microRNA regulates dendritic spine development. *Nature* **439**, 283-289, doi:10.1038/nature04367 (2006).

14 Modarresi, F. *et al.* Inhibition of natural antisense transcripts in vivo results in gene-specific transcriptional upregulation. *Nature biotechnology* **30**, 453-459, doi:10.1038/nbt.2158 (2012).

15 Ivanova, N. *et al.* Dissecting self-renewal in stem cells with RNA interference. *Nature* **442**, 533-538, doi:10.1038/nature04915 (2006).

16 Hsu, S. D. *et al.* miRTarBase: a database curates experimentally validated microRNA-target interactions. *Nucleic acids research* **39**, D163-169, doi:10.1093/nar/gkq1107 (2011).

17 Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic acids research* **44**, D457-462, doi:10.1093/nar/gkv1070 (2016).

18 Psych, E. C. *et al.* The PsychENCODE project. *Nature neuroscience* **18**, 1707-1712, doi:10.1038/nn.4156 (2015).

19 Gautier, L., Cope, L., Bolstad, B. M. & Irizarry, R. A. affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**, 307-315, doi:10.1093/bioinformatics/btg405 (2004).

20 Zhang, J. D. & Wiemann, S. KEGGgraph: a graph approach to KEGG PATHWAY in R and bioconductor. *Bioinformatics* **25**, 1470-1471, doi:10.1093/bioinformatics/btp167 (2009).

21 Reimand, J., Kull, M., Peterson, H., Hansen, J. & Vilo, J. g:Profiler--a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic acids research* **35**, W193-200, doi:10.1093/nar/gkm226 (2007).