Graphic Lasso: Data analysis

Jiaxin Hu

March 12, 2021

1 Network Analysis

Here is the network analysis for the gtex data with top var-var gene selection and with tuning parameters $r = 3, \rho = 1500$. The membership result is in the note 030221 Table 1.

Since we only have substantia nigra in group 3, we focus on the network analysis on the first two groups, i.e., Θ_1, Θ_2. Note that there are 3 (excluding the symmetric enties) non-zero entries in Θ_1 and 19 (excluding the symmetric enties) non-zero entries in Θ_2. Combing the genes related to these non-zero entries, we have 32 strong genes.

Figure 1 are the heat-maps of the estimated sub-precision matrices with strong genes and without the global connections, which is equal to $\sum_{k=1}^{3} u_{ik} \tilde{\Theta}^k$ and $\tilde{\Theta}^k = \hat{\Theta}^k[\text{strong genes}, \text{strong genes}]$. Figure 2 are the corresponding heatmaps for the global patter and substantia nigra.

For better visualization, we show the strong relationship through networks.
In cortex and other tissues led by hippocampus, the featured connection is "SERPINA3-SLC32A1" and "LINC01007 - SERTM1", where

1. SEPRINA3: The protein encoded by this gene is a plasma protease inhibitor and member of the serine protease inhibitor class. Variations in this protein's sequence have been implicated in Alzheimer's disease, and deficiency of this protein has been associated with liver disease.

2. SLC32A1: The encoded protein is a member of amino acid/polyamine transporter family.

3. LINC01007: is a non-protein coding RNA gene.

4. SERTM1: is a protein coding gene related to Serine rich and transmembrane domain containing.

The connection "SERPINA3-SLC32A1" is enhanced in cortex while "LINC01007 - SERTM1" is depressed in cortex. Contrary pattern is shown in hippocampus.
In cerebellum and basal gangalia, the featured connections are "RP11-438B23 - LINGO2" and "FOSB - EGR1", where

1. RP11-438B23: is a non-protein coding RNA gene.

2. LINGO2: is a protein-coding gene related to Leucine Rich Repeat.

3. FOSB: encodes leucine zipper proteins and has been implicated as regulators of cell proliferation, differentiation, and transformation.
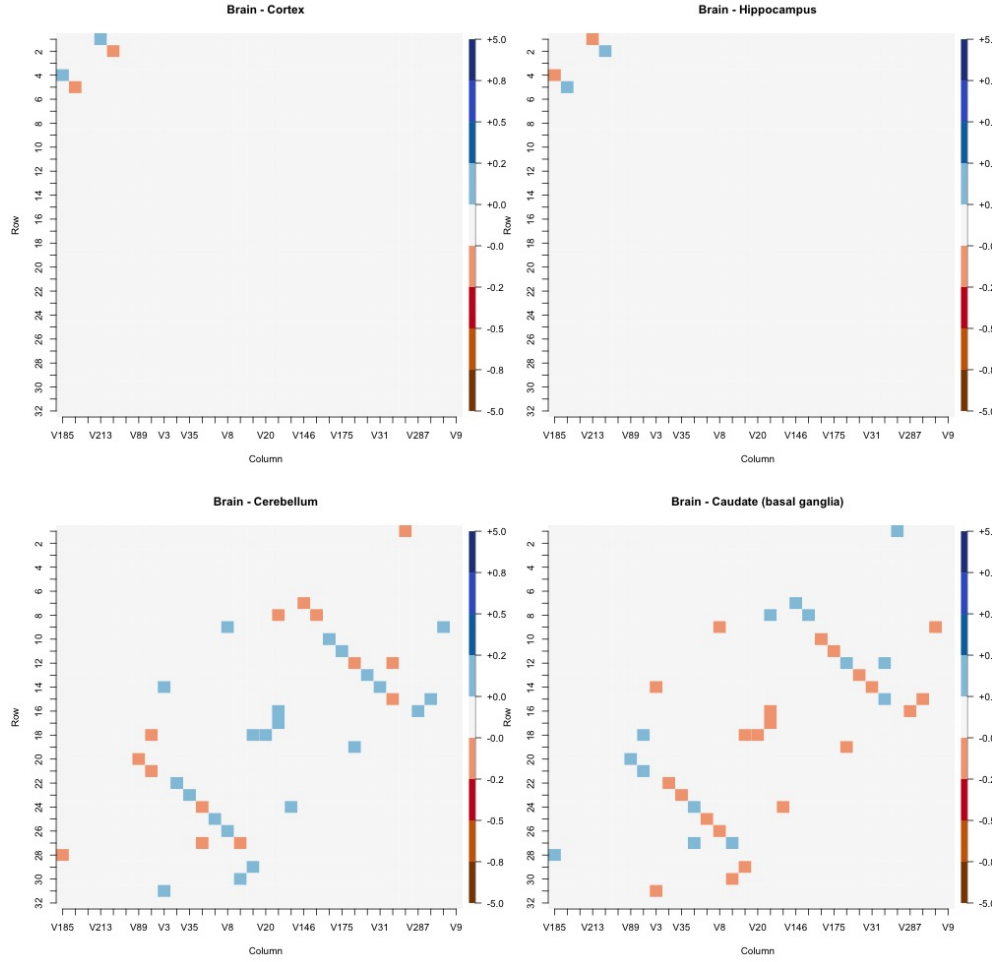
Figure 1: Sub-precision matrices with strong genes and without global correlations for Cortex, Hippocampus, basal ganglia, substania nigra.
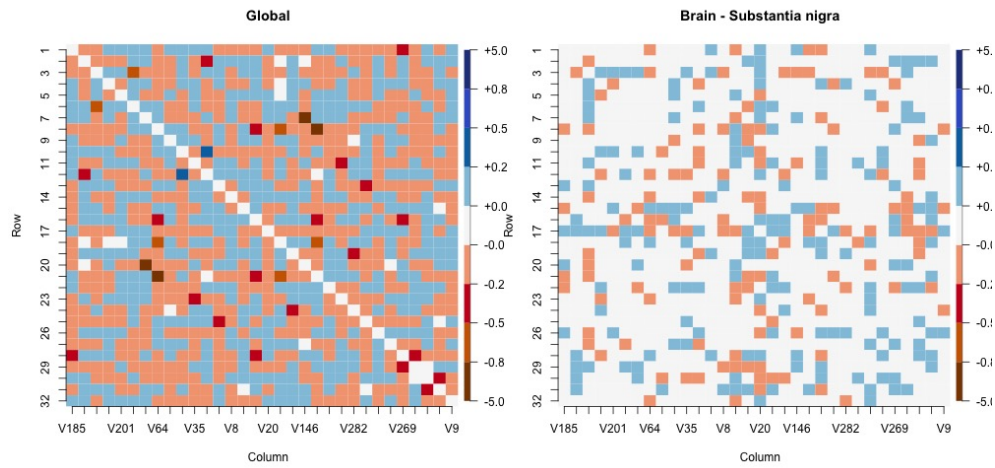

Figure 2: Sub-precision matrices with strong genes for global pattern and substanita nigra.
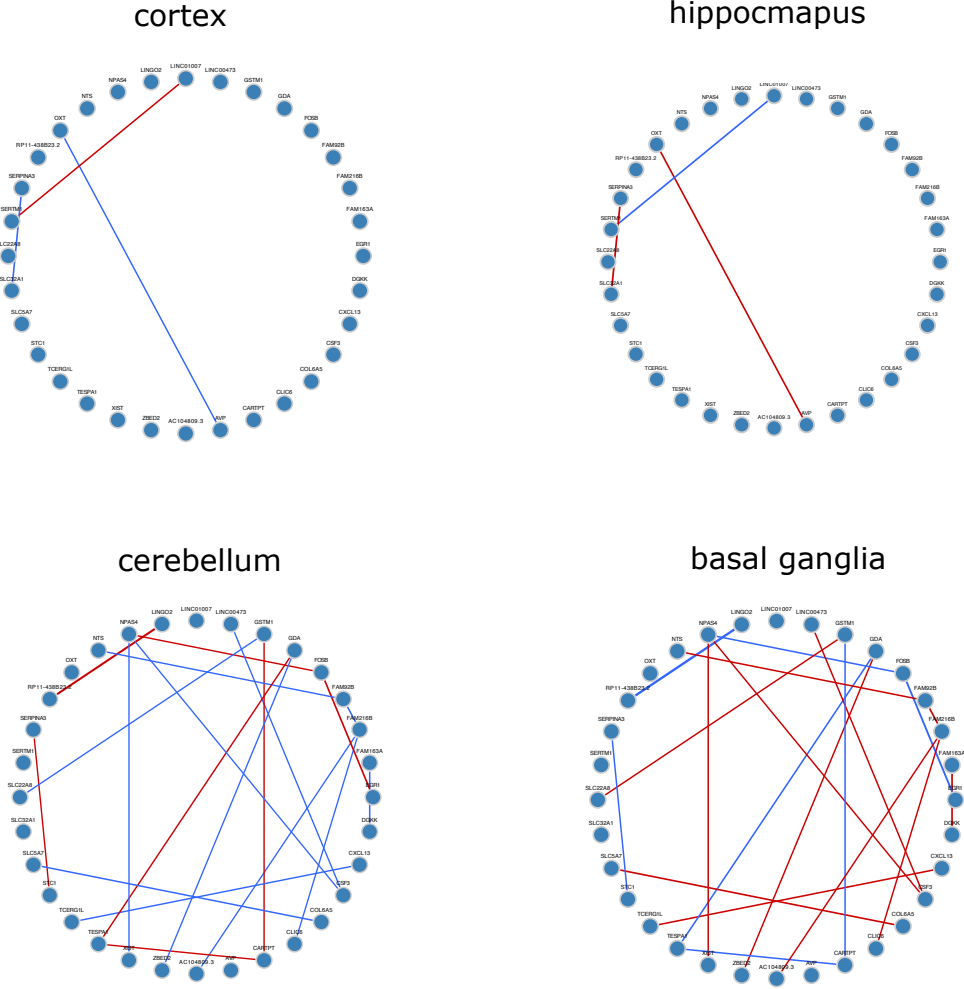
Figure 3: Precision networks with strong genes and without global correlations for Cortex, Hippocampus, cerebellum, and basal ganglia. Red edge refers to negative correlation, and blue edge refers positive correlation.

4. EGR1: encodes a nuclear protein, who functions as a transcriptional regulator. The products of target genes it activates are required for differentitation and mitogenesis. Studies suggest this is a cancer suppressor gene.

Also, genes "NPAS4" is recognized as the genes has most connections with other genes. NPAS4 encodes transcriptional regulators, which are involved in a wide range of physiologic and developmental events. Genes include "GDA", "TESPA1" and "CARTPT" have strong connections with other genes. Note that "GDA" gene is also recognized in the pre-processed data network analysis (note 022421).

Overall, the interpretability of the network is not as good as pre-processed the data (in note 022421). We have recognized few connections between RNA genes and protein-coding genes the network. However, the specific functions of these RNA genes and protein-coding genes are not very clear.