Graphic Lasso: Data analysis

Jiaxin Hu

February 25, 2021

1 Clustering results

With the pre-processed brain tissue data, the clustering result is relatively stable and reasonable. I list the clustering results under different parameters.

$A_0 : u_{kl} = 0$	Brain - Cerebellum Brain - Cerebellar Hemisphere
$A_{11} : u_{k1} > 0$	Brain - Cortex Brain - Anterior cingulate cortex (BA24) Brain - Frontal Cortex (BA9)
$A_{12} : u_{k1} < 0$	Brain - Hippocampus, Brain - Hypothalamus Brain - Spinal cord (cervical c-1) Brain - Amygdala
$A_2 : u_{k2} \neq 0$	Brain - Caudate (basal ganglia) Brain - Nucleus accumbens (basal ganglia) Brain - Putamen (basal ganglia)
$A_3 : u_{k3} \neq 0$	Brain - Substantia nigra

Table 1: Membership result for 13 brain tissues with $r = 3, \rho = 1000$.

$A_0 : u_{kl} = 0$	Brain - Substantia nigra Brain - Putamen (basal ganglia)
$A_{11} : u_{k1} > 0$	Brain - Cortex Brain - Anterior cingulate cortex (BA24) Brain - Frontal Cortex (BA9)
$A_{12} : u_{k1} < 0$	Brain - Hippocampus, Brain - Hypothalamus Brain - Spinal cord (cervical c-1)
$A_2 : u_{k2} \neq 0$	Brain - Nucleus accumbens (basal ganglia)
$A_{31} : u_{k3} > 0$	Brain - Cerebellum Brain - Cerebellar Hemisphere
$A_{32} : u_{k3} < 0$	Brain - Amygdala Brain - Caudate (basal ganglia)

Table 2: Membership result for 13 brain tissues with $r = 3, \rho = 500$.

Based on the listed results, the choice $\rho = 1000$ seems perform the best both in $r = 2$ and $r = 3$, and with $r = 3$ the substantia nigra is able to be separated from cerebellum. Under other settings, there exist some unreasonable clustering. For example, when $r = 3, \rho = 500$, basal ganglia tissues are separated; when $r = 3, \rho = 1500$ and $r = 2, \rho = 500$, cerebellum is mixed with cortex.

Overall, the setting $r = 3, \rho = 1000$ gives the best clustering results. Therefore, I implement the correlation analysis under this case.

$A_{11} : u_{k1} > 0$	Brain - Cortex Brain - Anterior cingulate cortex (BA24) Brain - Frontal Cortex (BA9) Brain - Cerebellar Hemisphere Brain - Amygdala
$A_{12} : u_{k1} < 0$	Brain - Hippocampus, Brain - Hypothalamus Brain - Spinal cord (cervical c-1)
$A_2 : u_{k2} \neq 0$	Brain - Substantia nigra Brain - Nucleus accumbens (basal ganglia)
$A_3 : u_{k3} \neq 0$	Brain - Caudate (basal ganglia) Brain - Putamen (basal ganglia)

Table 3: Membership result for 13 brain tissues with $r = 3, \rho = 1500$.

$A_0 : u_{kl} = 0$	Brain - Cerebellum Brain - Substantia nigra Brain - Cerebellar Hemisphere
$A_{11} : u_{k1} > 0$	Brain - Cortex Brain - Anterior cingulate cortex (BA24) Brain - Frontal Cortex (BA9) Brain - Amygdala
$A_{12} : u_{k1} < 0$	Brain - Hippocampus, Brain - Hypothalamus Brain - Spinal cord (cervical c-1)
$A_2 : u_{k2} \neq 0$	Brain - Caudate (basal ganglia) Brain - Nucleus accumbens (basal ganglia) Brain - Putamen (basal ganglia)

Table 4: Membership result for 13 brain tissues with $r = 2, \rho = 1000$.

Others Here are few points need to be noticed

1. Sometimes, less iterations give better results than more iterations. At first, I set the max iteration as 50 but the algorithm gives bad results under most cases. Then, I set the max iteration as 30, and the get the listed results.

2. The original Gtex Brain tissue does not work well in clustering. Though we choose the same genes in the pre-processed tensor data, the performance of Gtex data is worse than listed results.

2 Correlation analysis

Here we consider the setting $r = 3, \rho = 1000$. Since group A_3 only contains one tissue, I picked some genes have strong correlations in the first two groups (cortex and basal ganglia groups).

After combine the genes have strong performance in Θ_1, Θ_2, we have 28 genes (strong genes) which

$A_0 : u_{kl} = 0$	Brain - Cortex Brain - Cerebellum Brain - Substantia nigra Brain - Cerebellar Hemisphere Brain - Caudate (basal ganglia) Brain - Nucleus accumbens (basal ganglia) Brain - Putamen (basal ganglia) Brain - Amygdala
$A_1 : u_{k1} \neq 0$	Brain - Spinal cord (cervical c-1)
$A_{21} : u_{k2} > 0$	Brain - Hippocampus Brain - Hypothalamus
$A_{22} : u_{k2} < 0$	Brain - Anterior cingulate cortex (BA24) Brain - Frontal Cortex (BA9)

Table 5: Membership result for 13 brain tissues with $r = 2, \rho = 500$.

may represent the correlation differences of each tissue compared with the global pattern.

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 $\tilde{\Theta}^k = \hat{\Theta}^k[\text{strong genes}, \text{strong genes}]$. Note that cerebellums only have the global pattern, and thus only have blank heat-maps. Also, the tissues in the same clustering groups have similar patterns, and thus we omit them here.

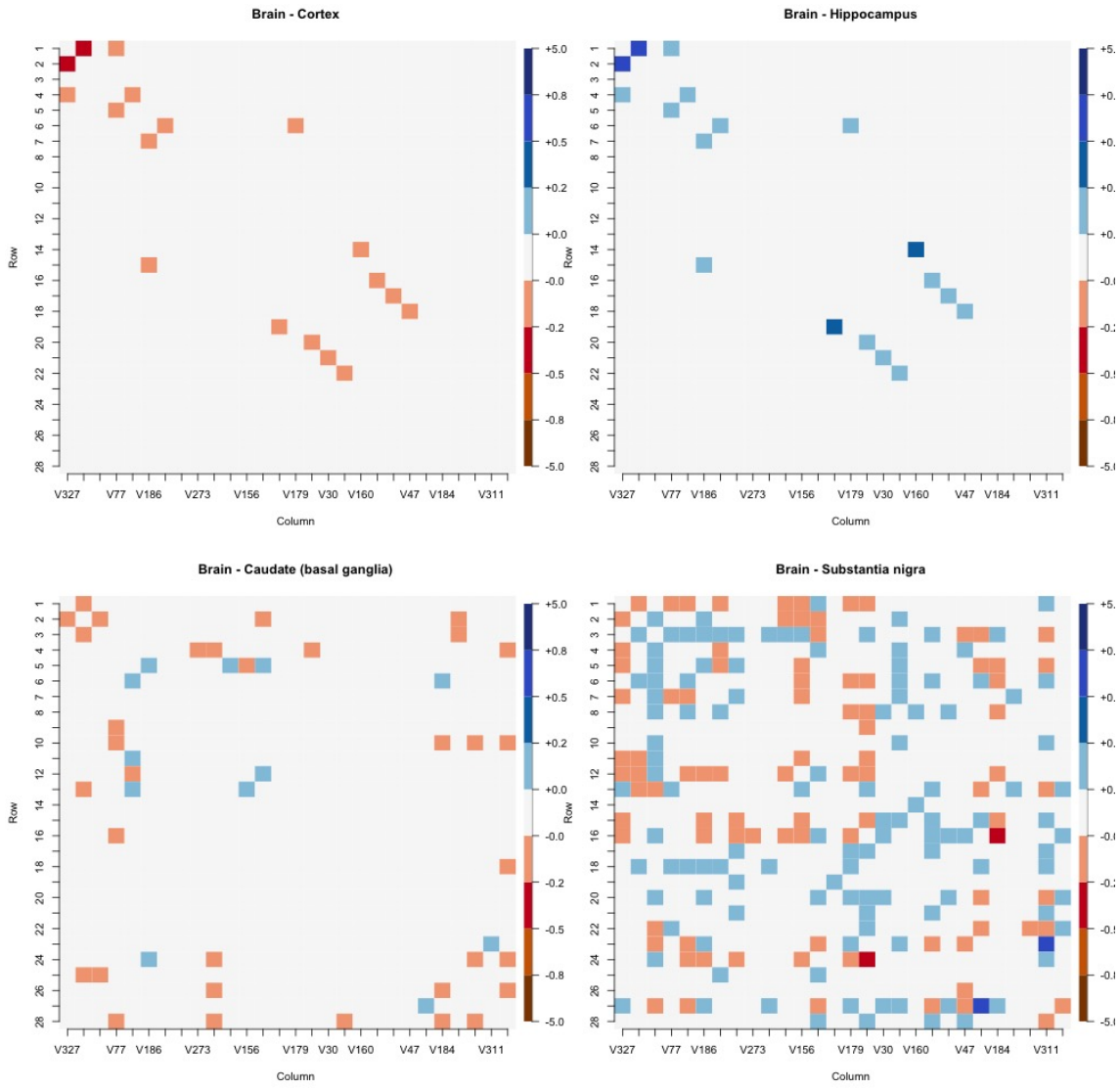

Figure 1: Sub-precision matrices with strong genes and without global correlations for Cortex, Hippocampus, basal ganglia, substania nigra.

For better visualization, Figure 2 shows the strong correlations as networks.
In cortex, there exist significant under-expressions in the correlations "TBR1-NEUROD6" and "MBP-MOBP" while these two correlations over-express in other brain tissues lead by Hippocampus. The genes MBP, MOBP have similar functions and thus the correlation is less interesting. In mouse, TBR1 is found in cortex and hippocampus related areas and plays an important role in neuronal migration and axonal projection. NEUROD6 may be involved in the development and

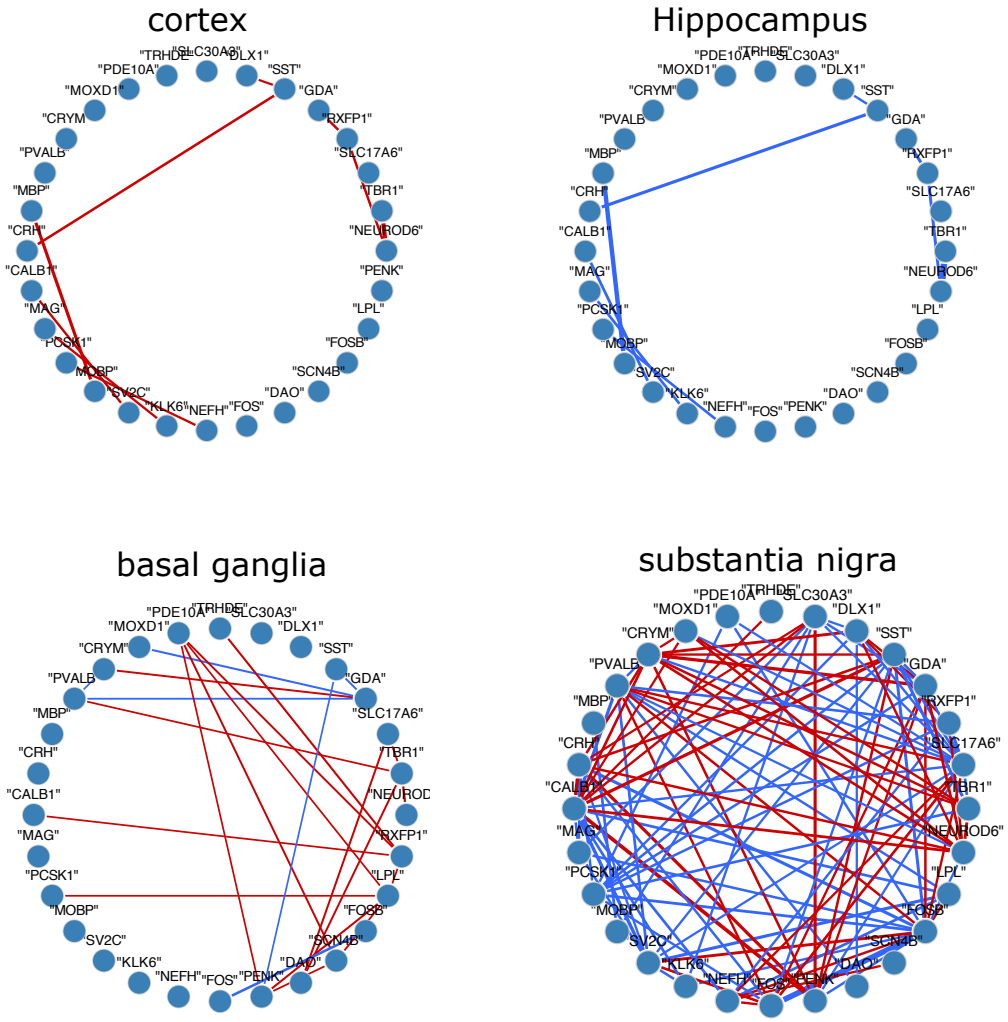

Figure 2: Precision networks with strong genes and without global correlations for Cortex, Hippocampus, basal ganglia, substania nigra. Red edge refers to negative correlation, and blue edge refers positive correlation.

differentiation of the nervous system.

In basal ganglia, the gene PDE10A, GDA have more correlations, where PDE10A plays an important role in signal transduction and GDA is related to microtubule assembly in mouse. Also, top correlation in basal ganglia is "RXFP1-TRHDE" and "TBR1-NEUROD6". RXFP1 plays an important role in sperm motility, pregnancy and parturition as a receptor for the protein hormone relaxin, and TRHDE is able to cleave and inactivate the neuropeptide neuropeptide thyrotropin-releasing hormone.

Notice that the correlation "TBR1-NEUROD6" is significant both in cortex, hippocampus and basal ganglia. This implies that "TBR1-NEUROD6" may be a special correlation does not exist in cerebellum.