

Graphic Lasso: Data analysis

Jiixin Hu

February 20, 2021

1 Brain Tissue

The membership now makes sense if we consider the sign of the coefficients in \mathbf{U} . Below is the result with $r = 3, \rho = 500$ with 30 iterations.

$A_0 : u_{kl} = 0$	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 - Caudate (basal ganglia) Brain - Amygdala

Table 1: Membership result for 13 brain tissues.

The cortex, cerebellum, and other tissues are separated. The only problem is basal ganglia. See Figure 1. According to the figure, the covariance matrices for Putamen and substantia are very similar, and they are both in group A_0 . Amygdala is more similar to cortex, and in the case $r = 2$ Amygdala is clustered with cortex. So, the group A_{32} is not reasonable. Covariance matrix for Cerebellum is different with other five covariance matrices.

2 Gtex Tissue

I tried different ways to select the genes might have strong correlations.

1. (Top variance genes) I combined the all the gene expression data into a matrix with dimension 18468×8555 , where 18468 is the number of genes and 8555 is the total sample size. I calculated the variance of each gene cross the sample size, and selected the genes with m top variance.

I firstly chose $m = 500$. However, I met problems when submitting the jobs to CHTC.

Then, I chose $m = 350$ and ran on my computer. With $r = 6, \rho = 1000$ and 30 iterations, we obtain the membership in Table 2.

Except the group A_{21} is led by Brain tissues, others seems not very make sense.

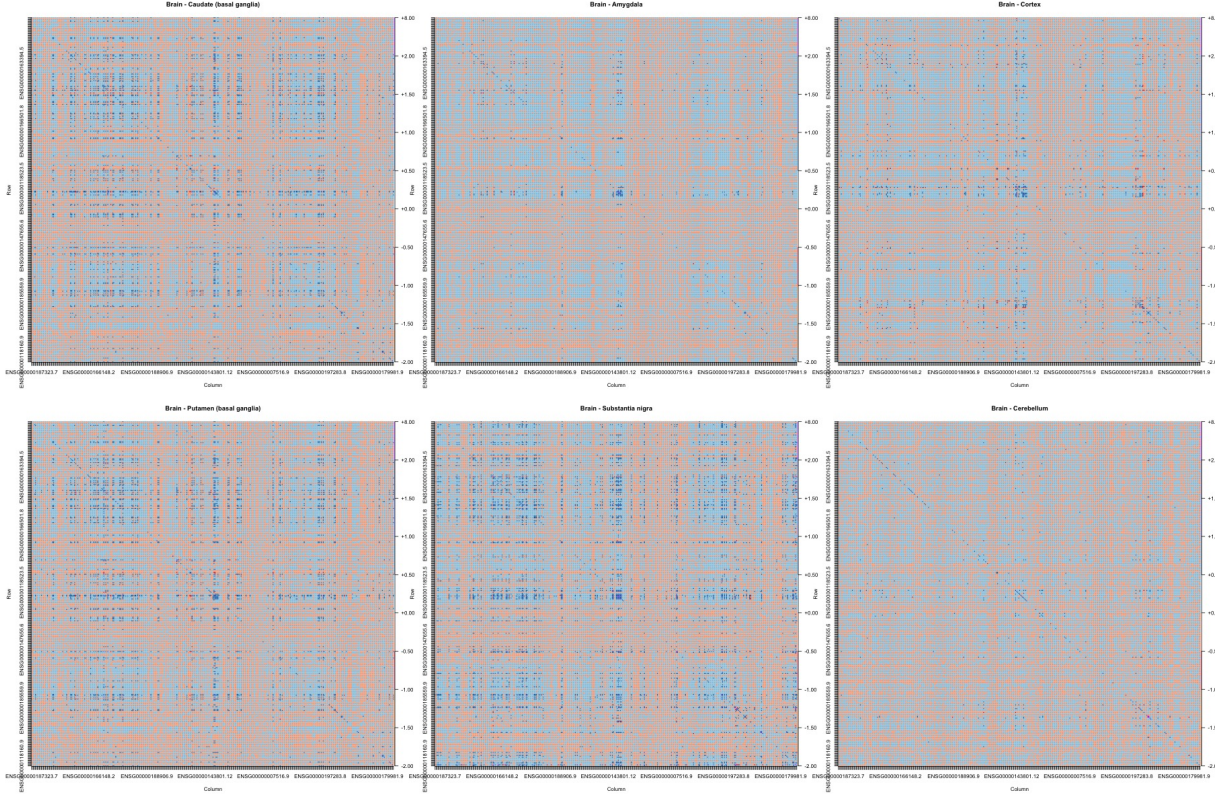


Figure 1: Sample covariance matrix for basal ganalia, substantia, amygdala, cortex and cerebellum.

2. (Gene with large mean difference) Based on the previous results, I want to reselect the genes which may represent the difference between tissues. According to the AOAS paper, I assume there are four different kinds: brain, blood, artery and adipose, and skin. I combine the expression data of these four kinds respectively, and find the genes in each kind with large mean difference compared with total mean. Combining the strong genes from four kinds, the resulted precision matrices are of dimension 277×277 .

I consider 20 typical tissues from these four kinds. With $r = 2, \rho = 1000$ with 25 iterations. The membership results are in Table 3. The result implies the brain tissues can be distinguished with these genes. However, other kinds of tissues are hard to distinguish.

$A_0 : u_{kl} = 0$	Adipose - Subcutaneous Artery - Coronary Heart - Atrial Appendage Adipose - Visceral (Omentum) Uterus Adrenal Gland Thyroid Lung Spleen Pancreas Esophagus - Muscularis Esophagus - Mucosa Colon - Sigmoid Small Intestine - Terminal Ileum Colon - Transverse Prostate Testis Skin - Sun Exposed (Lower leg) Nerve - Tibial Heart - Left Ventricle Pituitary Brain - Cerebellum Artery - Aorta Cells - EBV-transformed lymphocytes Liver Kidney - Cortex Brain - Cerebellar Hemisphere Brain - Putamen (basal ganglia) Brain - Hypothalamus Brain - Spinal cord (cervical c-1) Brain - Amygdala Fallopian Tube Cervix - Ectocervix Cervix - Endocervix
$A_{11} : u_{k1} > 0$	Brain - Cortex
$A_{12} : u_{k1} < 0$	Whole Blood Stomach
$A_{21} : u_{k1} < 0$	Muscle - Skeletal Artery - Tibial Vagina Breast - Mammary Tissue Esophagus - Gastroesophageal Junction Cells - Transformed fibroblasts
$A_{22} : u_{k1} > 0$	Brain - Hippocampus Brain - Substantia nigra Brain - Anterior cingulate cortex (BA24) Brain - Frontal Cortex (BA9) Brain - Caudate (basal ganglia) Brain - Nucleus accumbens (basal ganglia)
$A_3 : u_{k3} < 0$	Bladder
$A_4 : u_{k4} \neq 0$	Ovary
$A_5 : u_{k5} \neq 0$	Minor Salivary Gland
$A_6 : u_{k6} \neq 0$	Skin - Not Sun Exposed (Suprapubic)

Table 2: Membership result for 53 tissues.

$A_{11} : u_{k1} < 0$	Brain - Cortex Brain - Cerebellum Brain - Substantia nigra Brain - Anterior cingulate cortex (BA24) Brain - Frontal Cortex (BA9) Brain - Cerebellar Hemisphere Brain - Caudate (basal ganglia) Brain - Nucleus accumbens (basal ganglia) Brain - Putamen (basal ganglia) Brain - Hypothalamus Brain - Spinal cord (cervical c-1) Brain - Amygdala
$A_{12} : u_{k1} > 0$	Brain - Hippocampus Skin - Not Sun Exposed (Suprapubic) Skin - Sun Exposed (Lower leg) Whole Blood Cells - EBV-transformed lymphocytes Artery - Tibial Artery - Coronary Artery - Aorta

Table 3: Membership result for 20 typical tissues.