Graphic Lasso: Data analysis

Jiaxin Hu

February 28, 2021

1 Higher r leads to degeneracy

The Table 1 concludes the situations for rank degeneracy with pre-processed data.

Rank	ρ	Degeneracy
$r = 3$	$\rho = 1000$	No degeneracy
$r = 3$	$\rho = 1500$	No degeneracy
$r = 3$	$\rho = 2000$	$\hat{r} = 1$
$r = 4$	$\rho = 1000$	No degeneracy
$r = 4$	$\rho = 1500$	$\hat{r} = 3$
$r = 4$	$\rho = 2000$	$\hat{r} = 2$
$r = 5$	$\rho = 1000$	$\hat{r} = 3$
$r = 5$	$\rho = 1500$	$\hat{r} = 2$
$r = 5$	$\rho = 2000$	$\hat{r} = 2$
$r = 6$	$\rho = 1000$	$\hat{r} = 3$
$r = 7$	$\rho = 1000$	$\hat{r} = 3$

Table 1: Rank degeneracy under different settings.

Based on the results, higher rank and larger sparsity parameters will lead to rank degeneracy.

2 Gtex data with top variance genes

Picking top variance genes is not a good way to select representative genes. Bellow is the clustering results.
The results imply that we need to use other ways to find the representative genes.

Here are some graphical checks.

Figure 1 compares the three cortex tissues in three datasets. Figure 2 compares the cortex and cerebellum in three datasets.
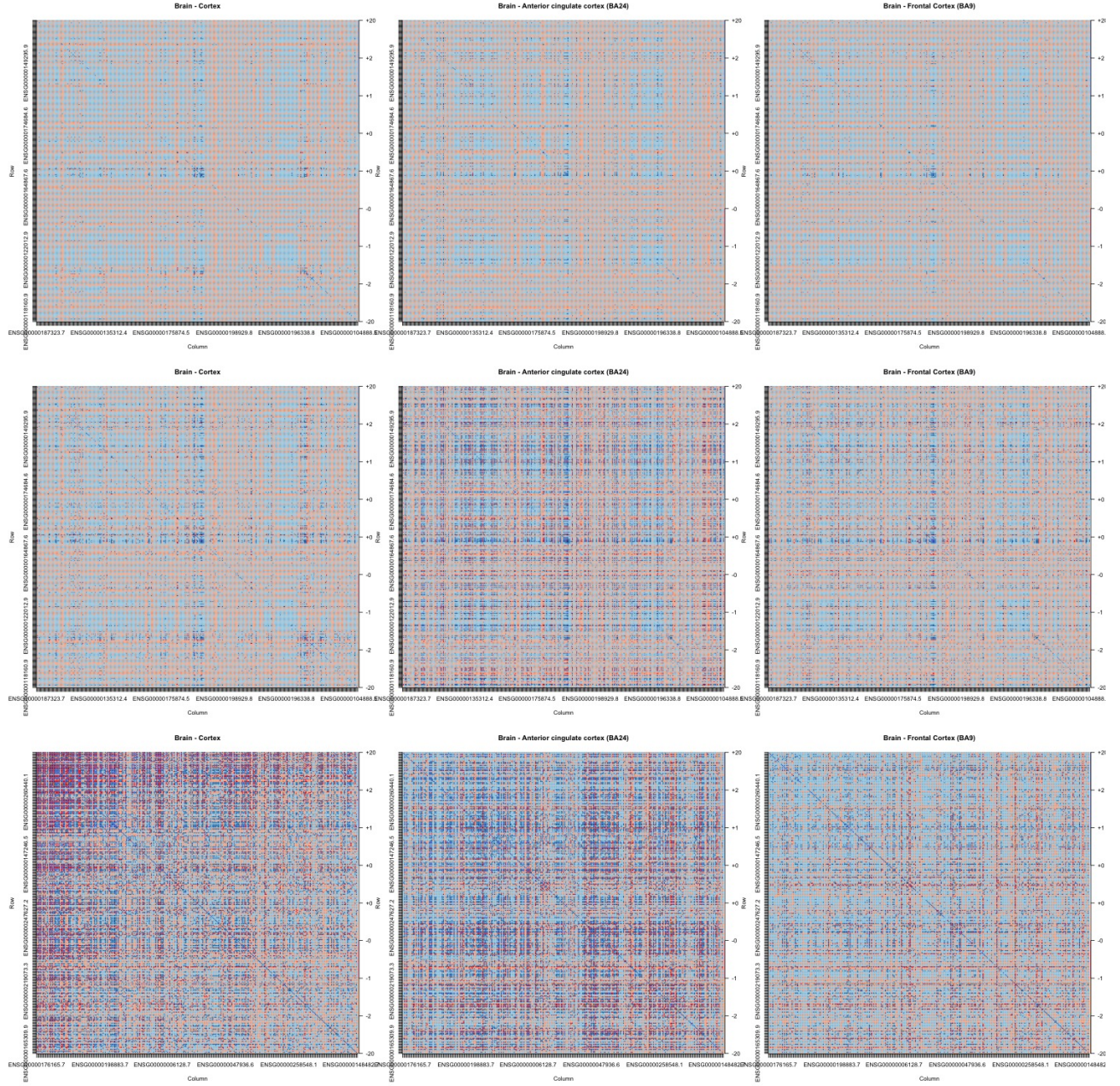
Figure 1: Covariance matrices for three cortex tissues in pre-processed data(first row), gtex data with pre-selected genes(second row), and gtex data with top variance genes(third row).
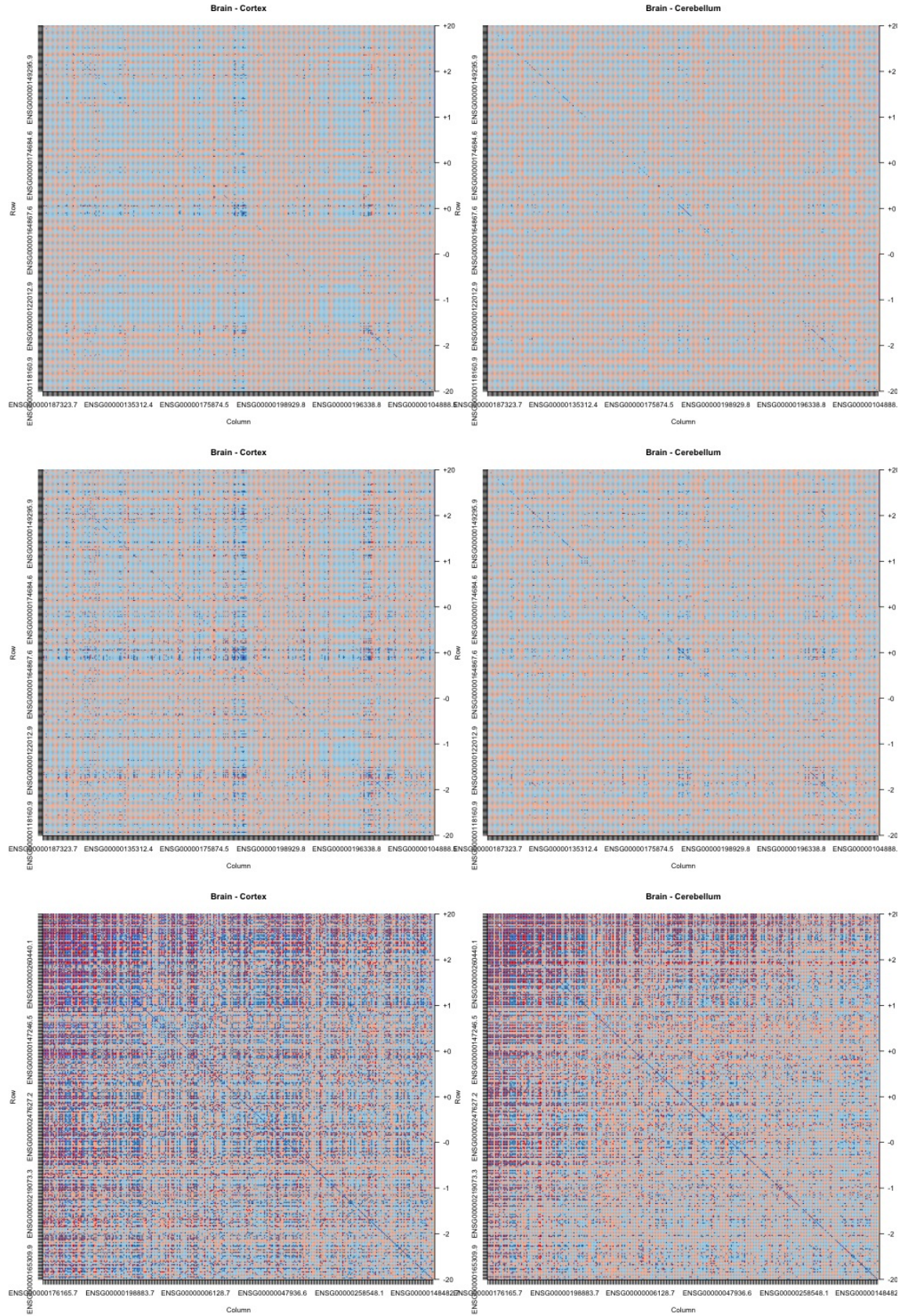
Figure 2: Covariance matrices for cortex and cerebellum in pre-processed data(first row), gtex data with pre-selected genes(second row), and gtex data with top variance genes(third row).

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

Table 2: Membership result for 13 brain tissues with $r = 3, \rho = 1000$ with top variance genes.

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

Table 3: Membership result for 13 brain tissues with $r = 2, \rho = 1000$ with top variance genes.