Joint Sparse Canonical Correlation Analysis for Detecting Differential Imaging Genetics Modules

Supplementary Data

1 Discussion on the selection of λ_v

In this section, we compare two different strategies to choose λ_v. One is the standard way that keeps λ_v the same among different classes, while the other keeps κ_v the same as defined in Sec 2.3. Fig. S1 depicts the solutions of the two cases with different λ_v and κ_v. As shown in the figures, the solutions become unstable given large λ_v. They tend to concentrate in one solution, which are therefore meaningless. When using the same κ_v, the solutions are more stable in both dense and sparse cases.

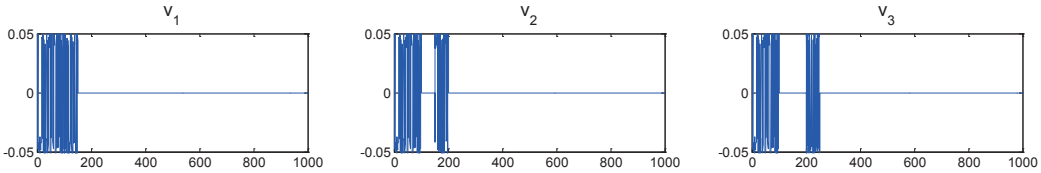
2 Stability selection and cross validation

In this section, we compare stability selection and cross validation in detecting differential correlated SNPs by ROC analysis. For cross validation, we draw the ROC using full data with $k_w = 100$ and k_v from 10 to 1000. In this way, we investigate the full solution path, which will not be affected by specific criterion. The TP and FP are calculated in the same way as introduced in Section 3. As seen in Fig. S2, stability selection is constantly better than cross-validation in detecting differential correlated SNPs, and the performance is much less sensitive to the selection of parameter τ.

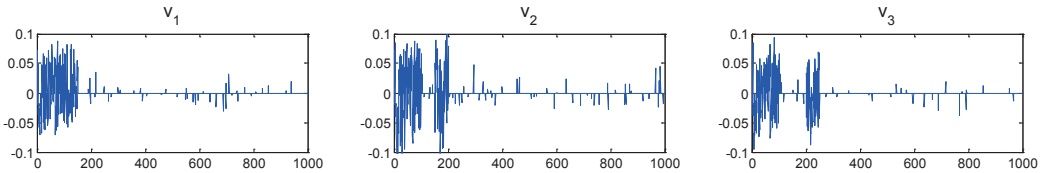
3 Simulations

In this section, we study the performance of the proposed method in different scenarios. Fig. S3-S7 show the simulation results with different combinations of sample size n, dimension p, q, and l, m_c, m_s. In each figure, we change one of them and check their effects on the performance. As seen in the figures, l, m_c does not affect much on the results. The results degrade as the sample size decreases and the dimension of the problem increases. It is very interesting that m_s has a lot of influence on the results. When $m_s = 0$, the fusion did not improve the performance, and when m_s increases, the performance becomes better. This simulation indicates that joint estimation can combine the strength of related datasets.
Fig. S8-S9 show the simulation results with different number of selected features. In particular, the top ranked 50 and 200 features are included to calculated the precision in Fig. S8 and Fig. S9, respectively. As seen in the figures, less selected features provide more reliable and stable results, which conform to application requirements.

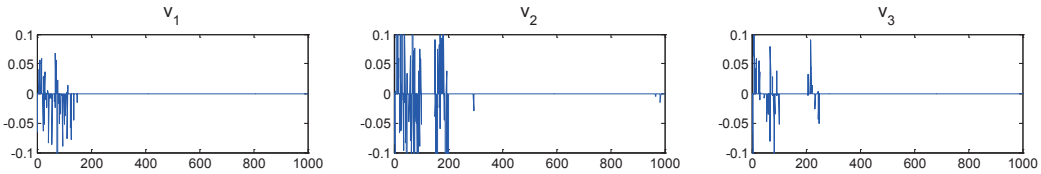
4 Figures



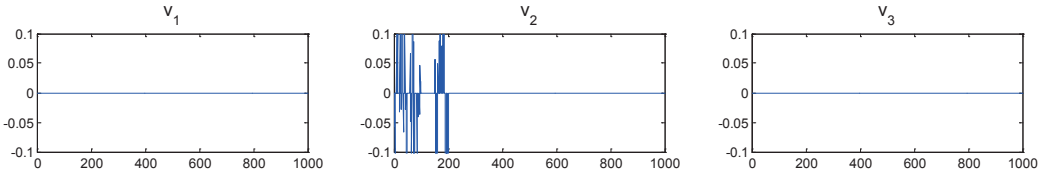
(a) Ground truth



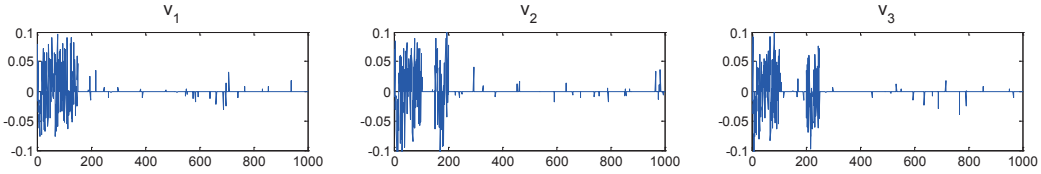
(b) $\lambda_v = 0.8$



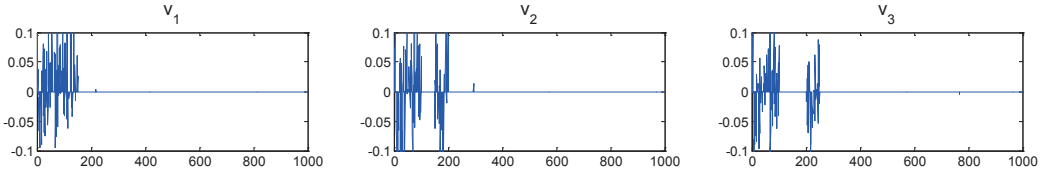
(c) $\lambda_v = 1.4$



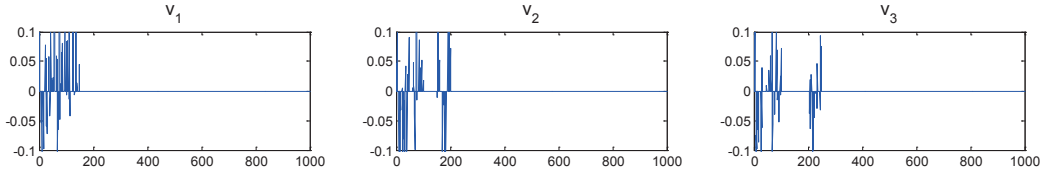
(d) $\lambda_v = 1.6$



(e) $\kappa_v = 160$



(f) $\kappa_v = 80$



(g) $\kappa_v = 40$

Figure S 1: The comparison of the solutions of JSCCA using fixed λ_v and fixed κ_v.

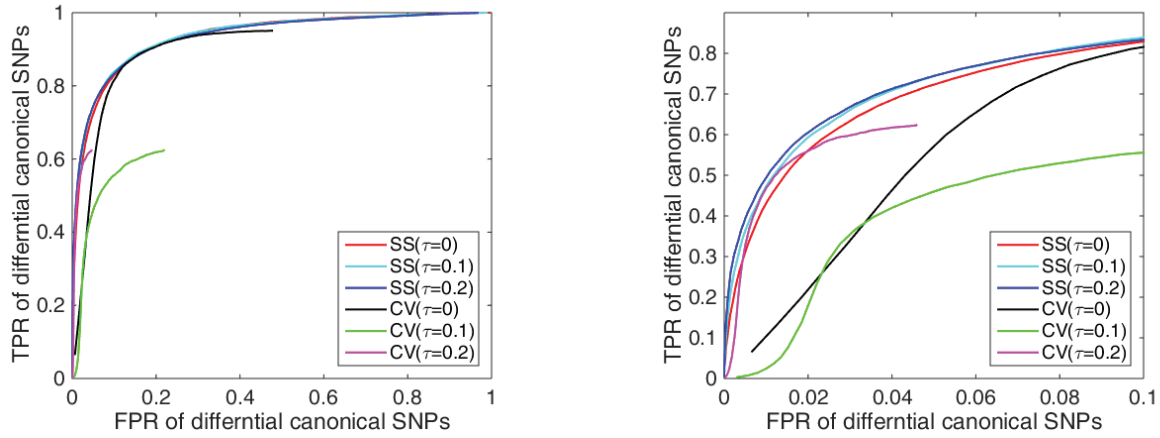


Figure S 2: A comparison between stability selection(SS) and cross-validaion(CV) in detecting differential correlated SNPs.

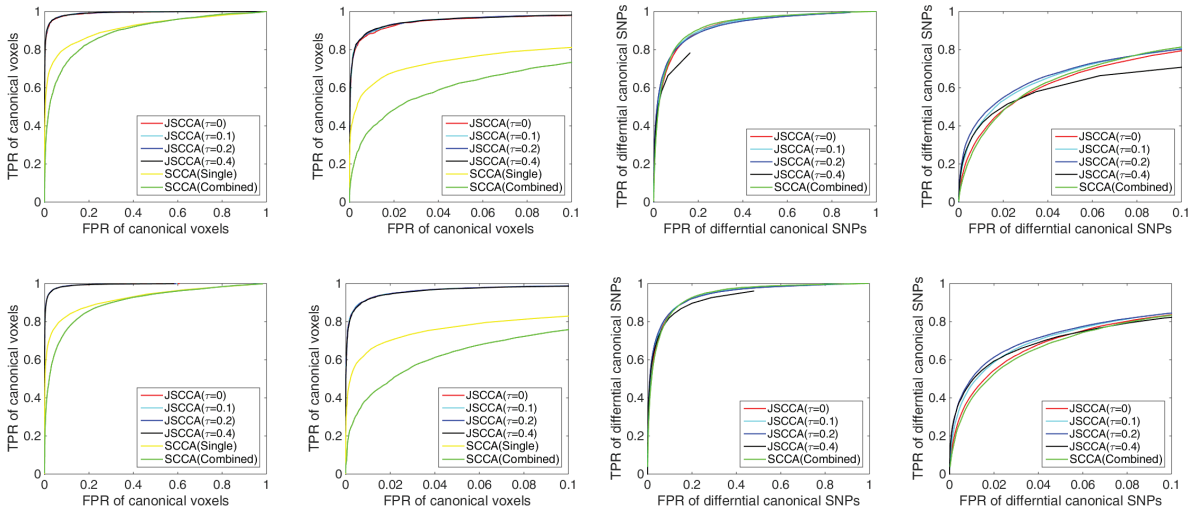


Figure S 3: The effects of l. Top: $l = 50$. Bottom: $l = 200$.

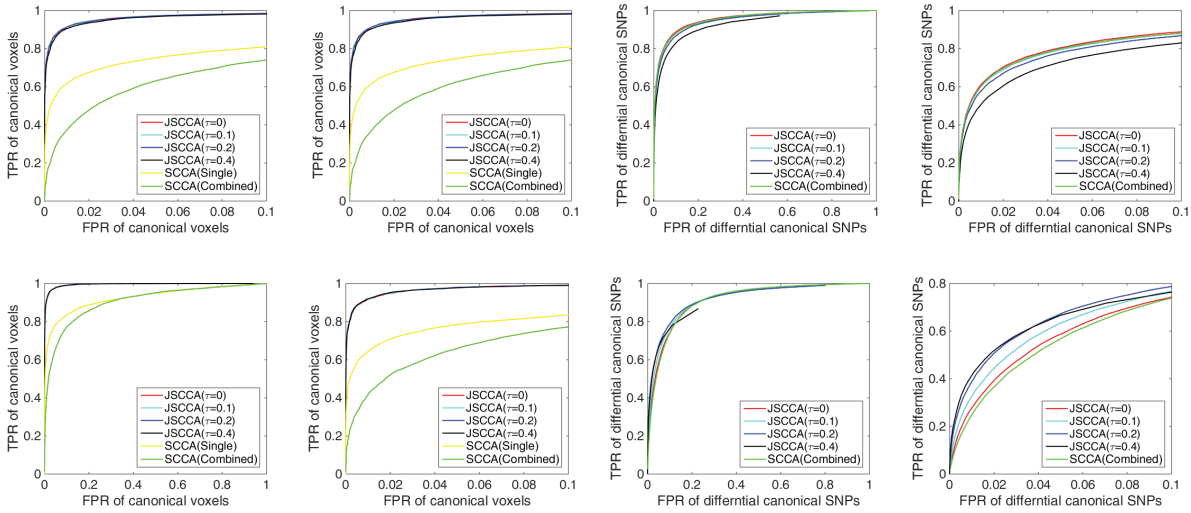


Figure S 4: The effects of m_s. Top: $m_s = 0$. Bottom: $m_s = 200$.

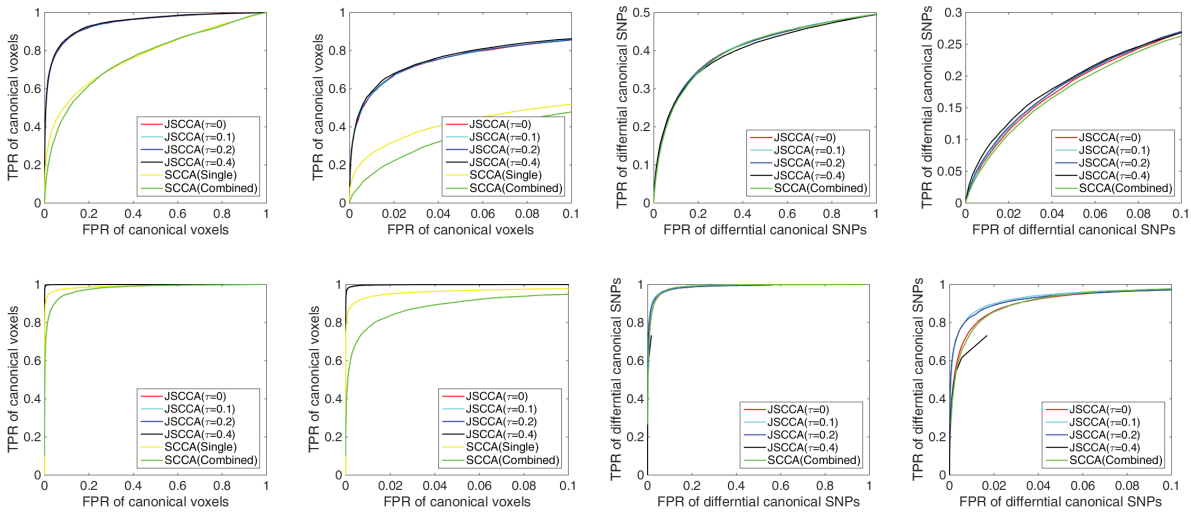


Figure S 5: The effects of sample size n. Top: $n = 50$. Bottom: $n = 200$.

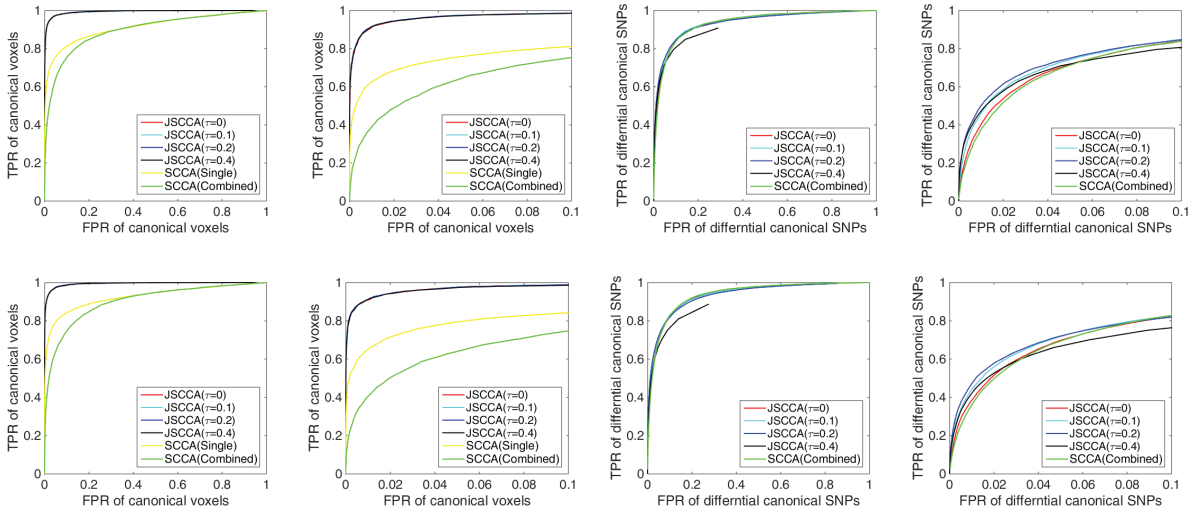


Figure S 6: The effects of m_c. Top: $m_c = 20$. Bottom: $m_c = 100$.

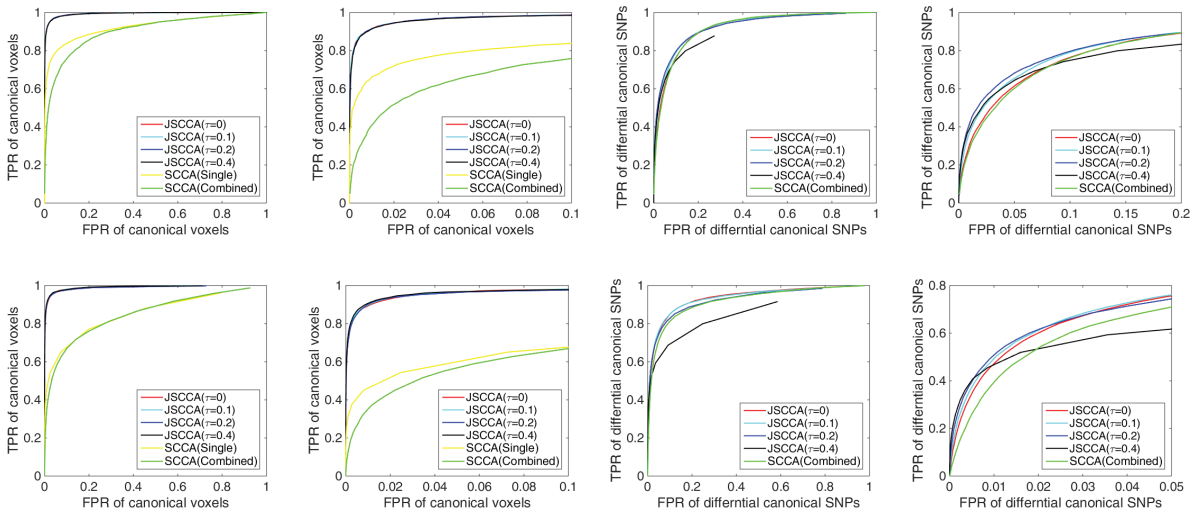


Figure S 7: The effects of dimension. Top: $p = q = 500$. Bottom: $p = q = 5000$.

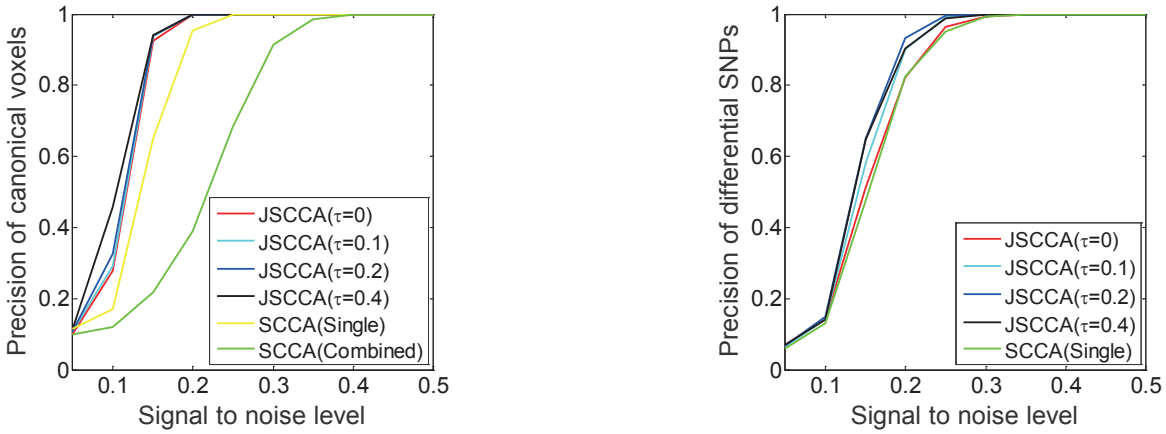


Figure S 8: Comparison of the precision with top ranked 50 voxels and SNPs.

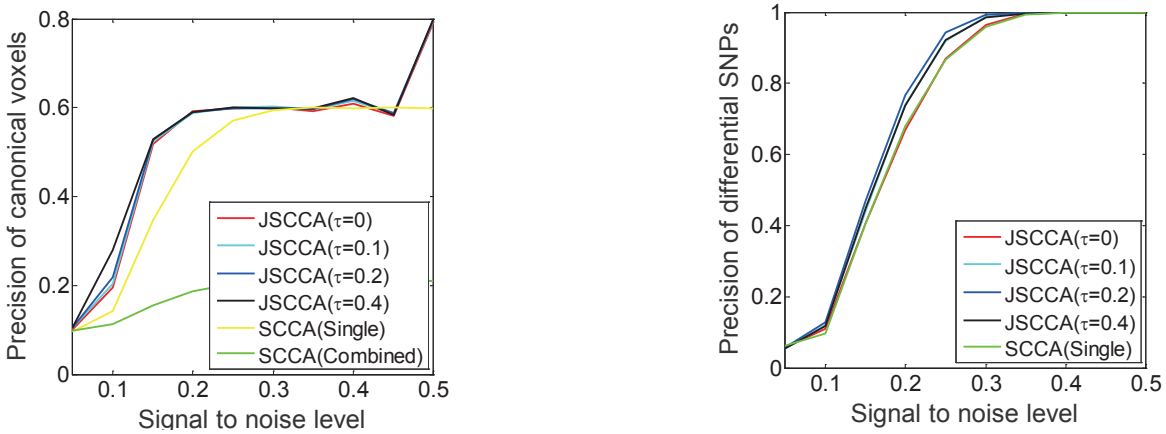


Figure S 9: Comparison of the precision with top ranked 200 voxels and SNPs.