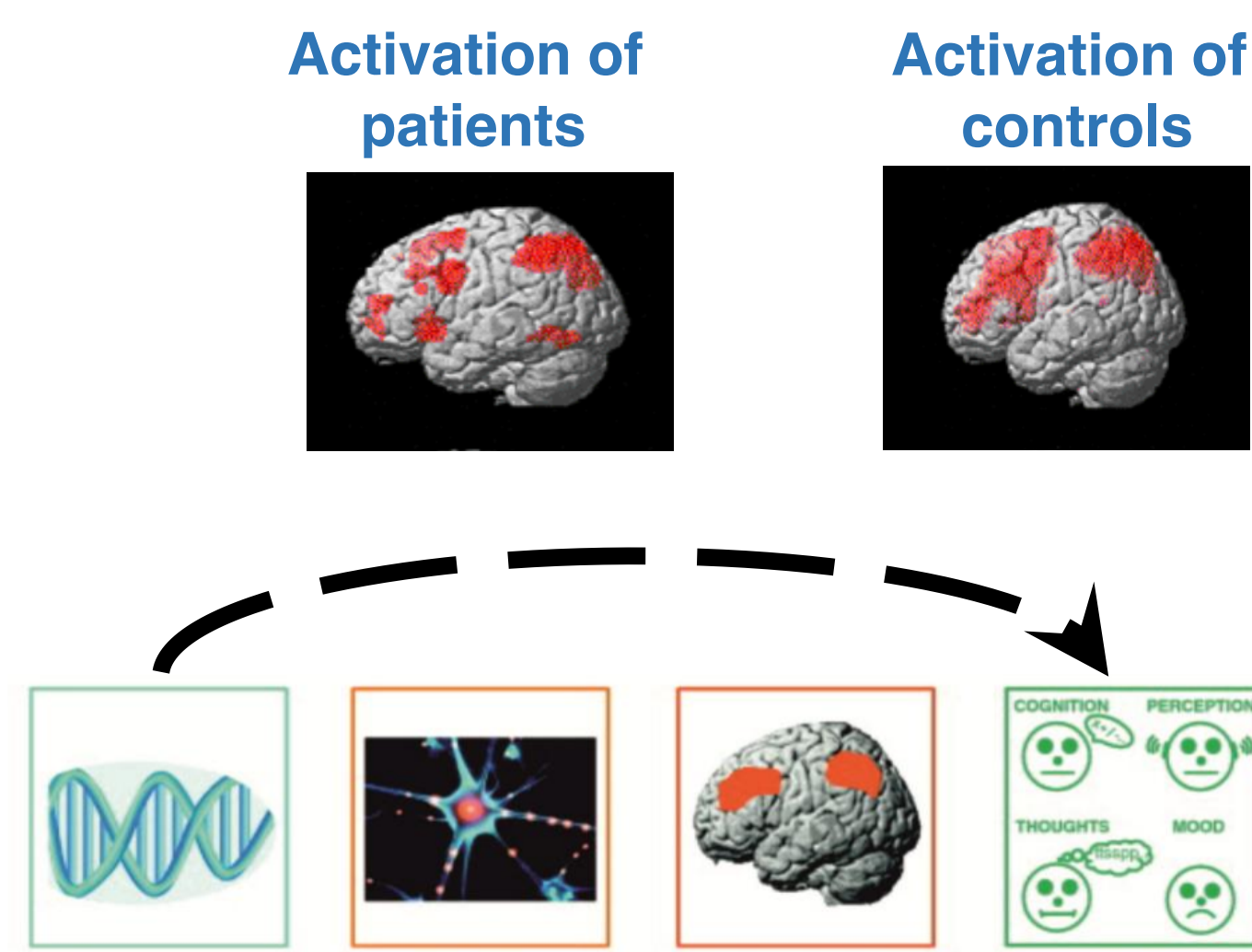


S. Ghosal¹, Q. Chen², A. L. Goldman², W. Ulrich², D. R. Weinberger^{2,3}, V. S. Mattay^{2,4}, and A. Venkataraman¹
¹Dept. of Electrical and Computer Engineering, Johns Hopkins University, ²Lieber Institute for Brain Development.
³Dept. of Psychiatry, Neurology and Neuroscience, Johns Hopkins University School of Medicine
⁴Dept. of Neurology and Radiology, Johns Hopkins University School of Medicine.

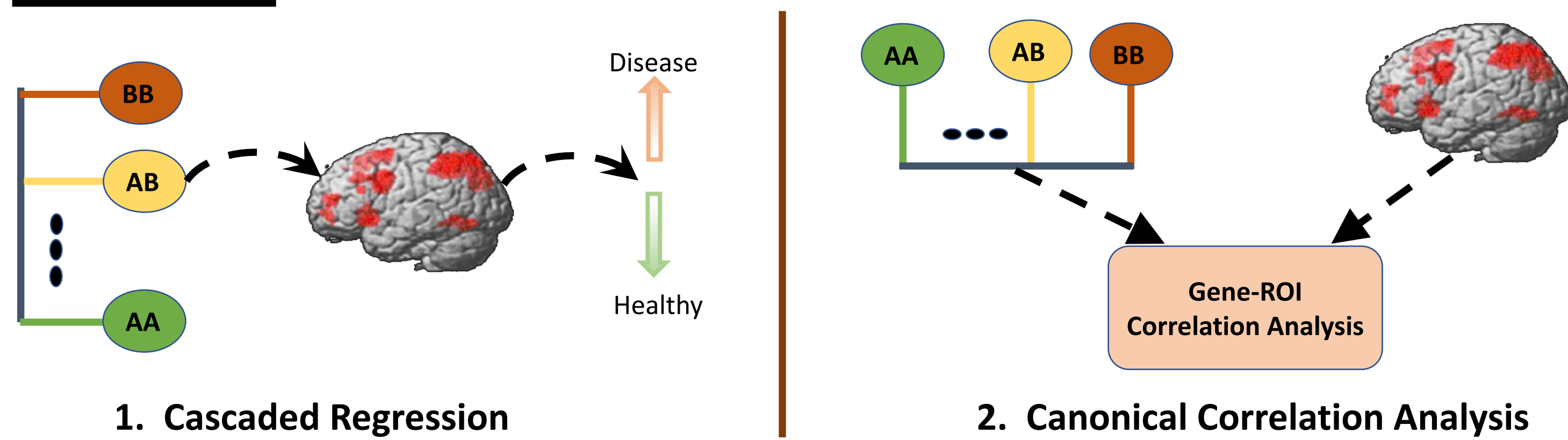


Motivation

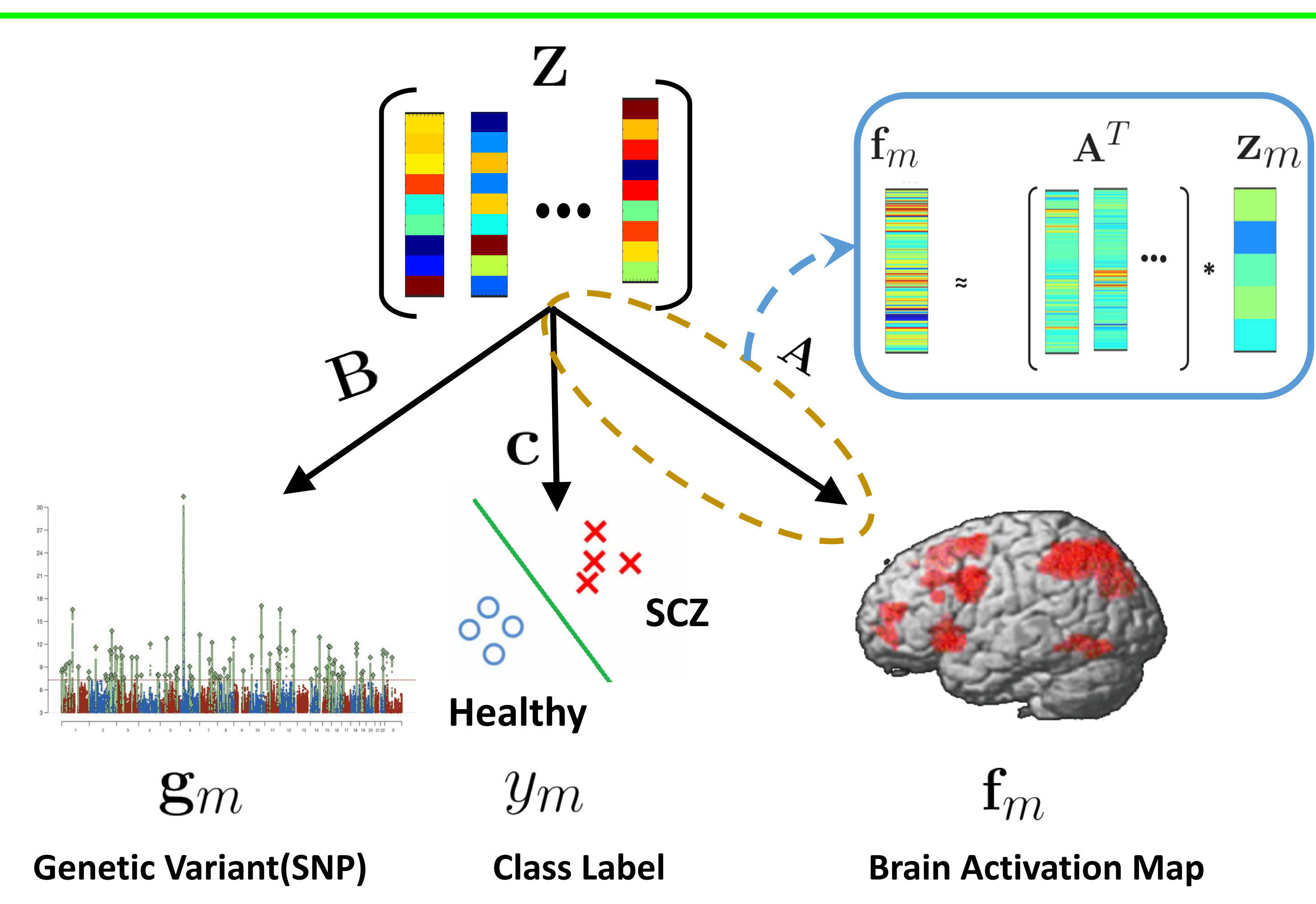
- Neuropsychiatric diseases are mainly characterized by atypical neural functioning.
- The Genetic variants play a crucial role in altering the brain activity patterns.



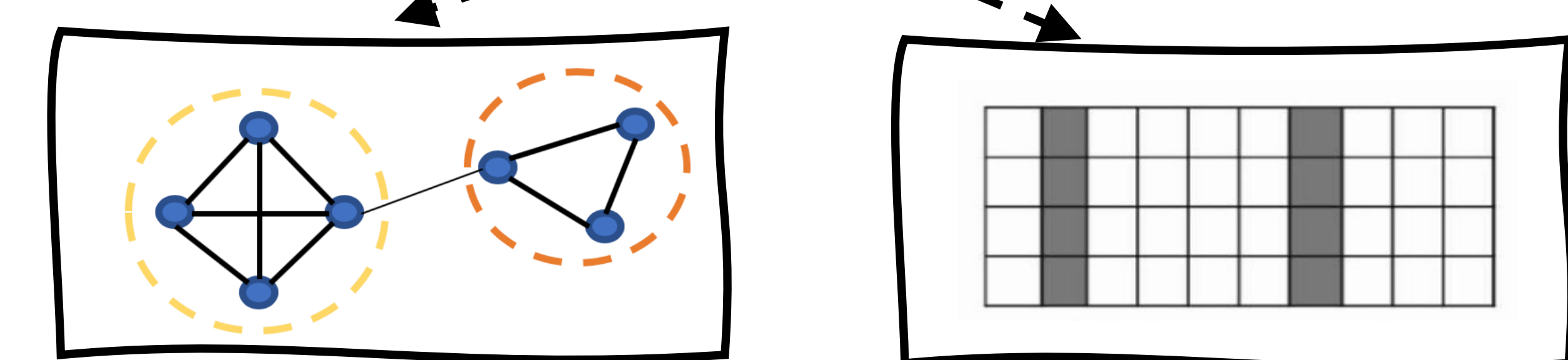
Prior Work



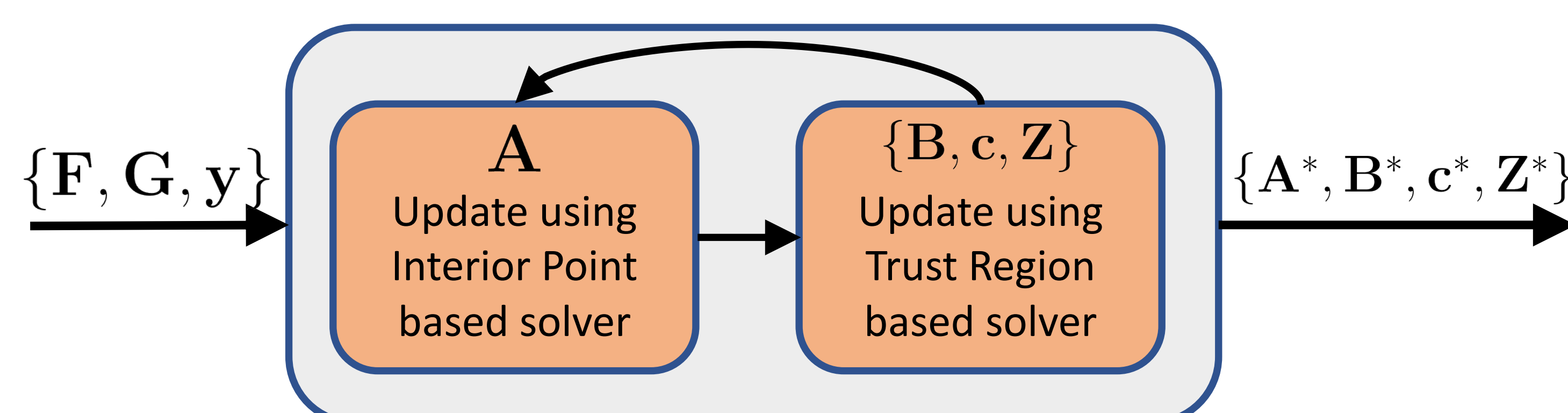
Methods



$$\mathcal{J} = \underbrace{\|F - A^T Z\|_F^2}_{\text{Imaging Loss}} + \underbrace{\|G - B^T Z\|_F^2}_{\text{Genetic Loss}} - \underbrace{\sum_{m=1}^M (y_m \log(\sigma(z_m^T c)) + (1 - y_m) \log(1 - \sigma(z_m^T c)))}_{\text{Logistic Loss}} + \underbrace{\frac{\lambda_1}{2} \text{Tr}(A^T L A)}_{\text{Graph Loss}} + \underbrace{\lambda_2 \|B\|_{2,1}}_{\text{Group Sparsity Loss}} + \underbrace{\frac{\lambda_3}{2} \|Z\|_F^2 + \frac{\lambda_4}{2} \|c\|_2^2}_{\text{Ridge Loss}} \quad \text{s.t. } A^T A = I$$



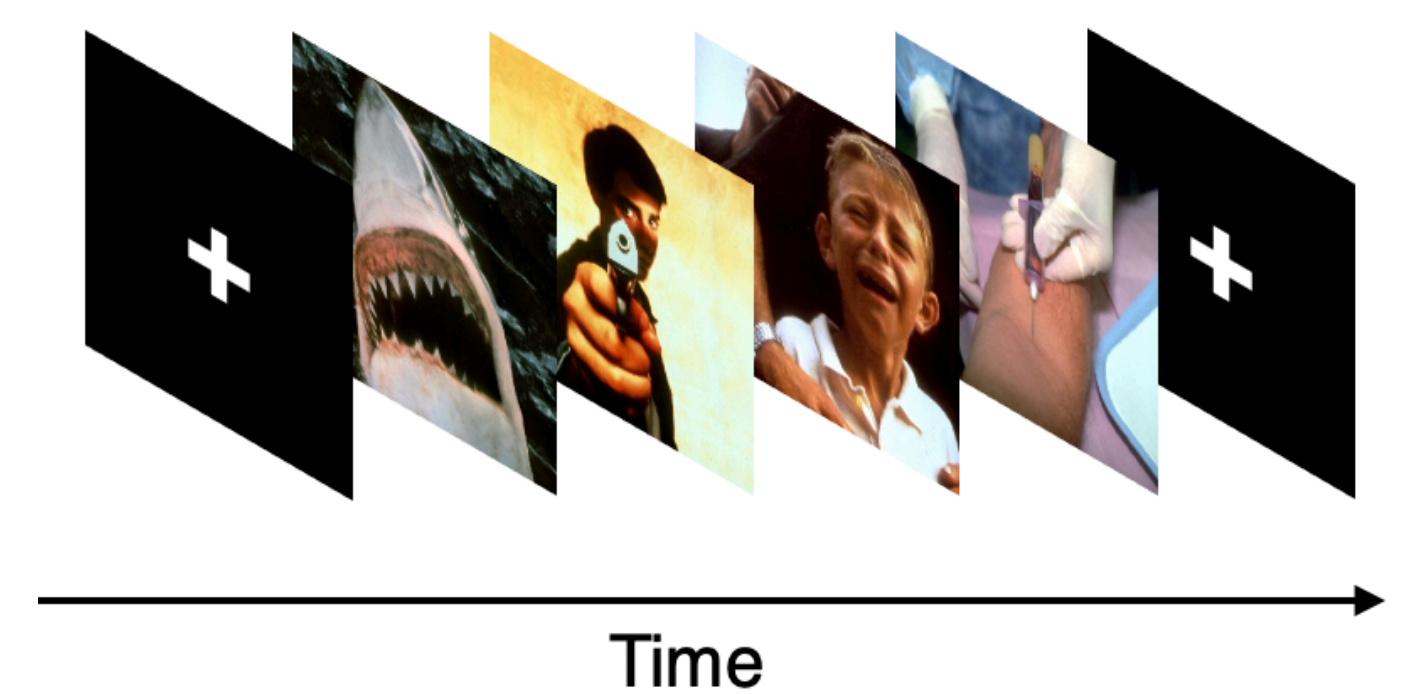
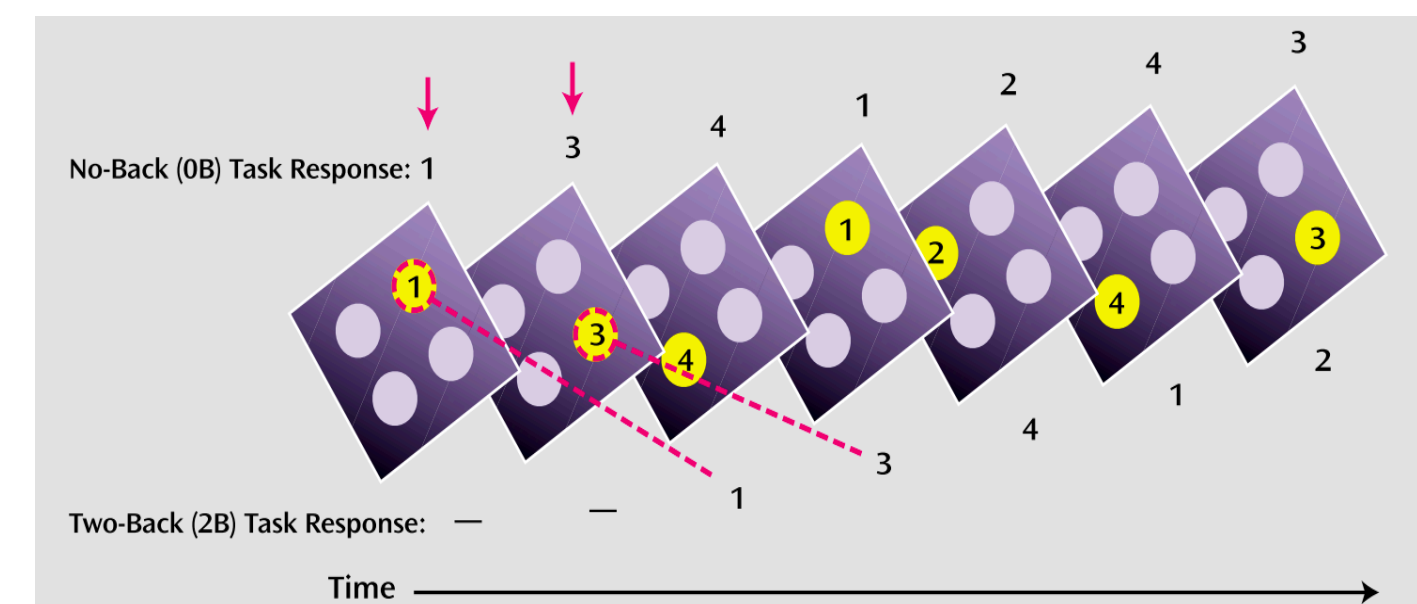
Optimization



Model Evaluation

Datasets: (1) N-back Working memory t-fMRI of 53 controls and 53 schizophrenic patients.

(2) Simple declarative memory t-fMRI (SDMT) of 46 controls and 47 schizophrenic patients.

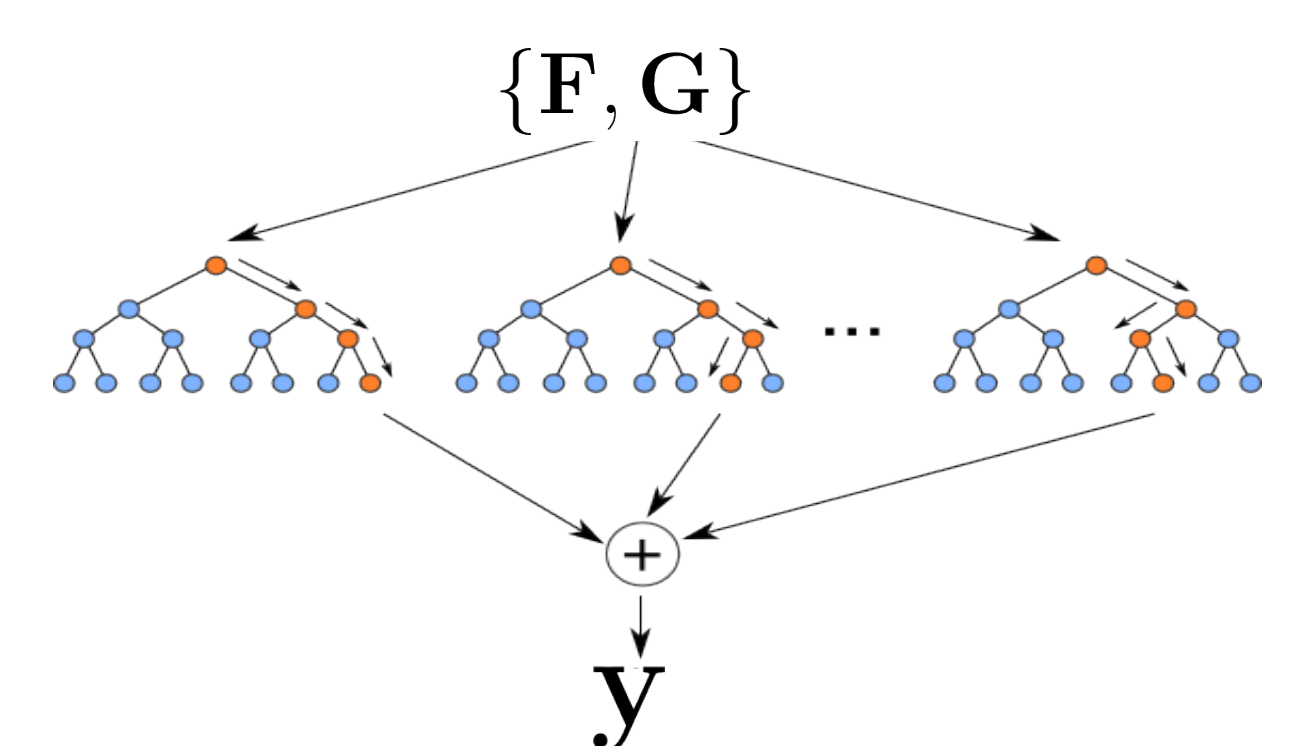


Baselines

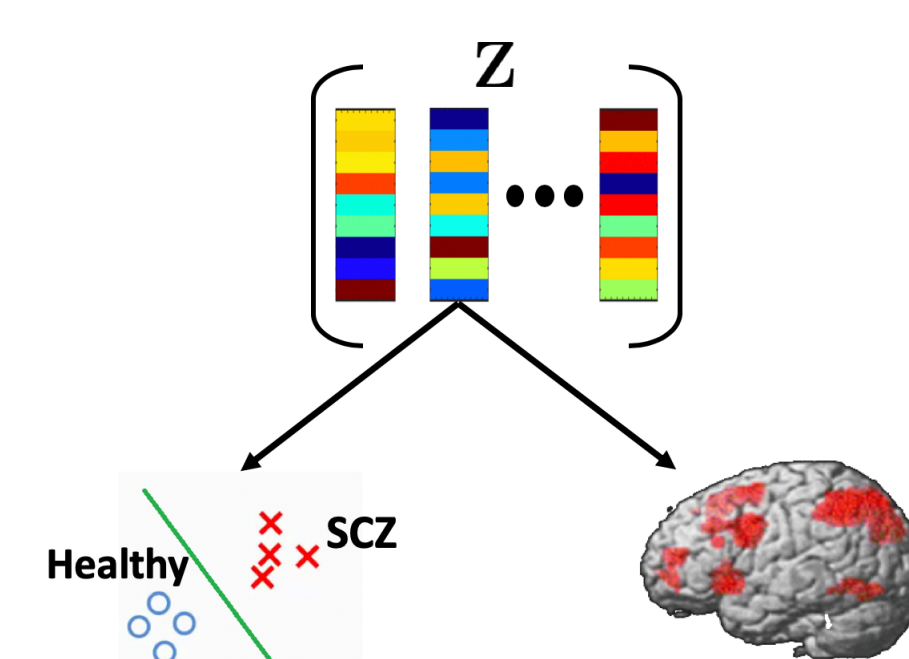
CCA + Random Forest:

The input to the RF classifier is the aligned imaging & genetics data after performing CCA.

Random Forest Classification:



Our method using Imaging variant only:



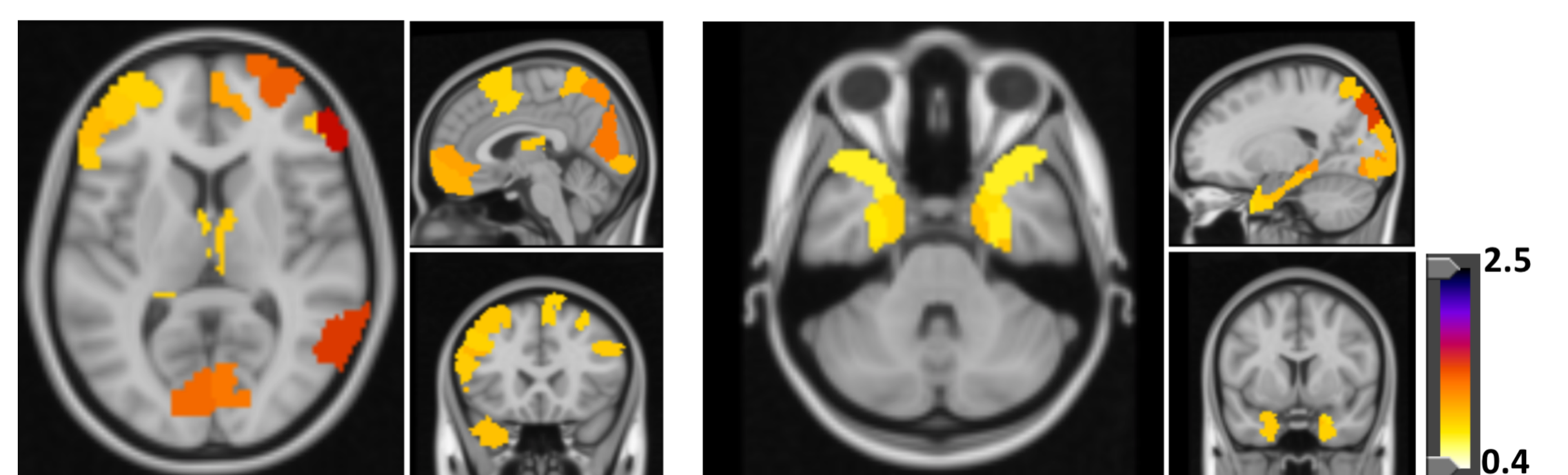
Evaluation Strategy

- 10 fold cross validation
- Interpretability
- Gene Ontology Analysis

Results

Methods	Nback Task			SDMT Task		
	Sens	Spec	Acc	Sens	Spec	Acc
RF	0.58	0.56	0.57	0.68	0.56	0.62
CCA+RF	0.41	0.47	0.44	0.55	0.69	0.62
Our Method (Imaging Only)	0.66	0.52	0.60	0.63	0.63	0.63
Our Method (Imaging + Genetics)	0.71	0.68	0.70	0.60	0.76	0.68

Table 1. Performance of each of the methods during 10 fold cross validation.



(a) Nback

(b) SDMT

Figure 1. The representative set of regions captured by the matrix, A.

Datasets	Biological Processes	FDR
Nback	Central nervous system development	0.03
	→ Nervous system development	0.0002
	→ System development	0.001
	Generation of neurons	0.03
	→ Neurogenesis	0.02
	→ Cell differentiation	0.003
SDMT	Forebrain neuron differentiation	0.04
	→ Nervous system development	0.002
	→ Generation of neurons	0.004
	→ Central nervous system neuron differentiation	0.04
	Central nervous system neuron development	0.02
	Regulation of neurogenesis	0.03

Table 2. The enriched biological processes along with their level of significance.

Conclusion

- Our matrix decomposition framework identifies the imaging and genetic biomarkers as guided by the clinical diagnosis
- Robustly and efficiently integrate diverse datatypes.