



---

# Mining Imaging Genetics Data via Sparsity-Induced Machine Learning Models

Hua Wang, Ph.D.

Department of Computer Science  
Colorado School of Mines

# Machine Learning and Data Mining Are Everywhere of Our Lives



Game playing and problem solving



**Machine Learning  
and Data Mining**



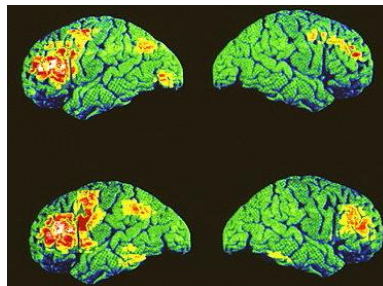
Intelligent virtual environments for treatment and rehabilitation



Industrial and engineering applications



Homeland security applications



Medicine, bioinformatics and systems biology



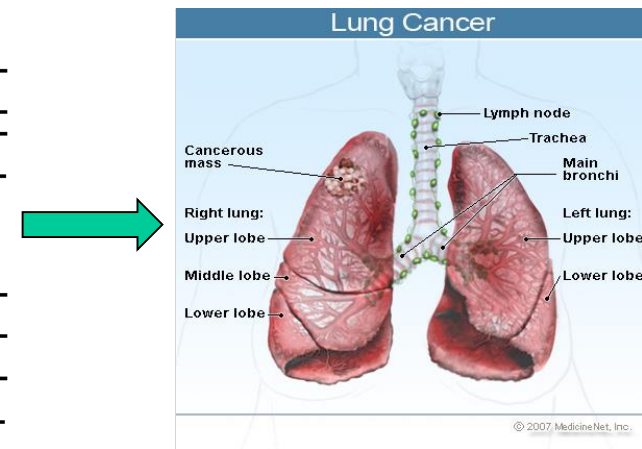
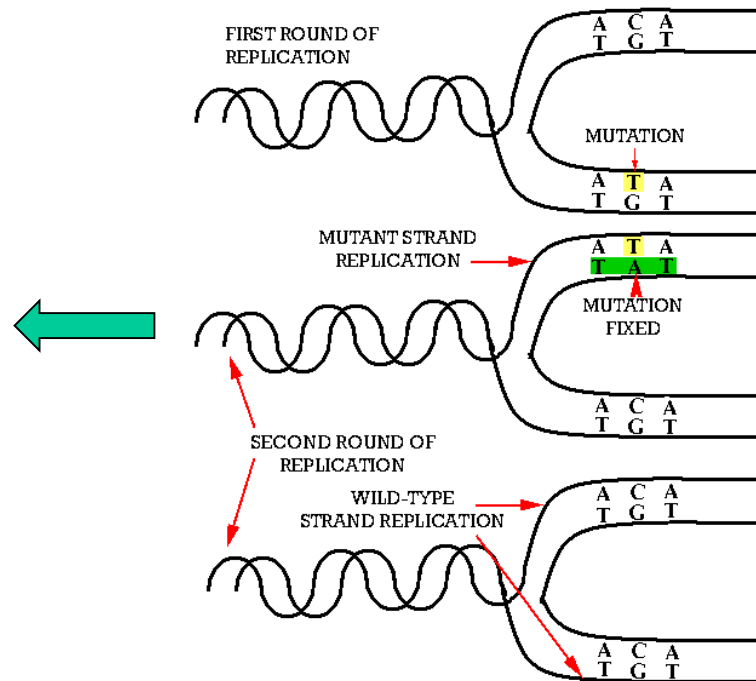
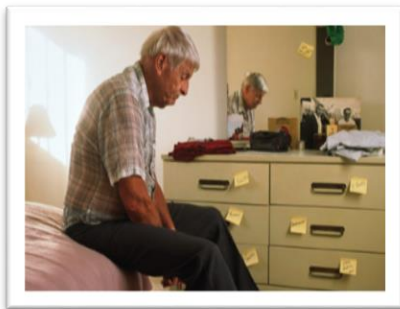
Economics, business and forecasting applications

... ..

# Many Major Human Diseases Have Been Connected to Gene Mutations



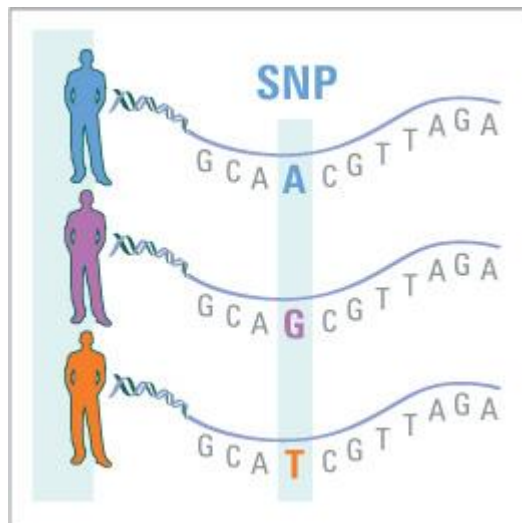
- Many major human diseases, such as cancer and neurodegenerative disorders, affects millions of people worldwide.
- Cause of these diseases: gene mutations (such as single-nucleotide polymorphism (SNP) or copy number variations).



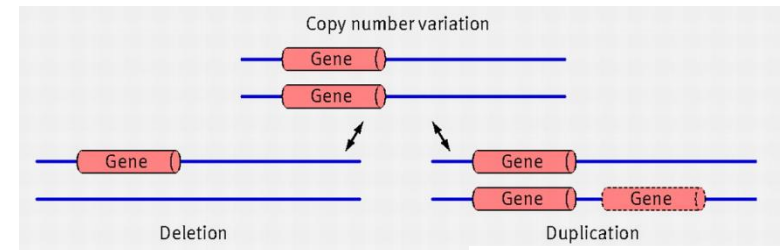
# Identification of Genetic/Genomic Biomarkers for Diagnosis And Therapy



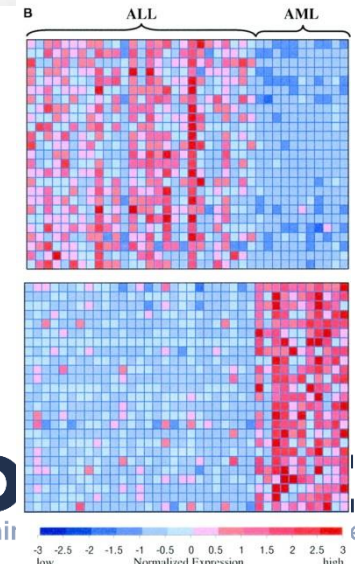
- Genetic variations characterized by SNP underlie differences in our susceptibility to, or protection from all kinds of diseases.
  - For example, a single base difference in the *Apolipoprotein E* is associated with a higher risk for Alzheimer's disease.



- CGH is a molecular-cytogenetic method for the analysis of copy number changes (gains/losses) in the DNA content of a given subject's DNA and often in tumor cells.



- Abnormal gene expression level could also indicate human diseases, such as cancers.

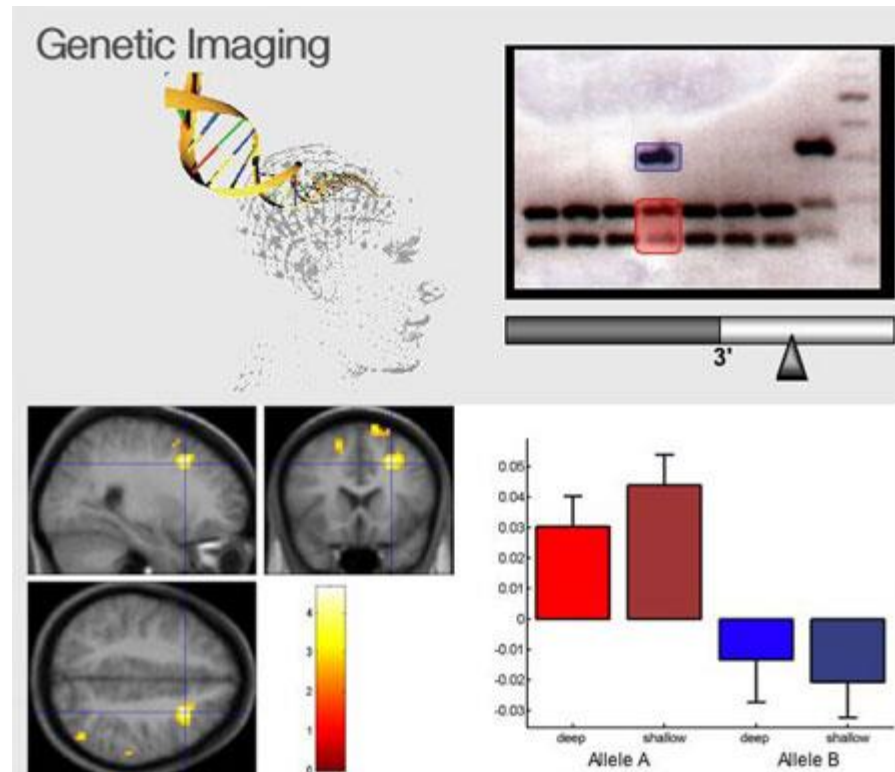




# Imaging Genetics via Machine Learning

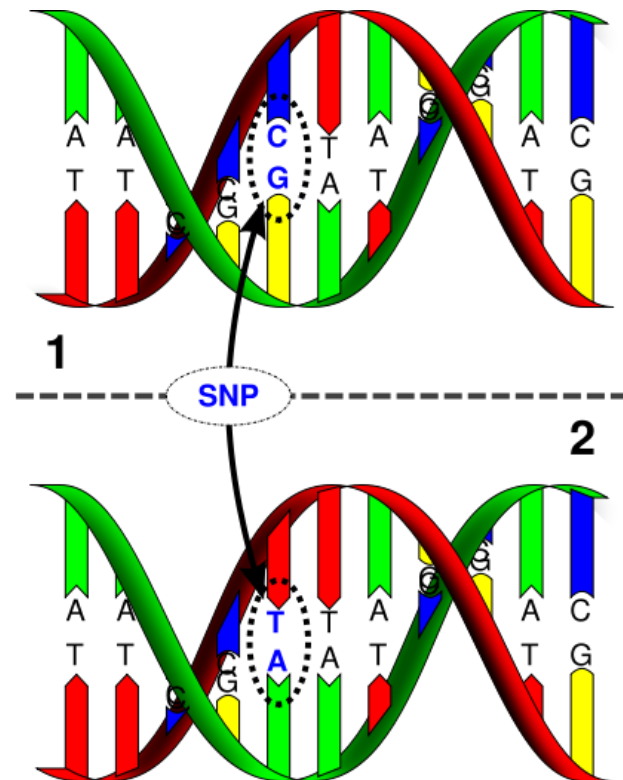
**Imaging genetics** refers to the use of anatomical or physiological imaging technologies as phenotypic assays to evaluate variations.

——wikipedia.org

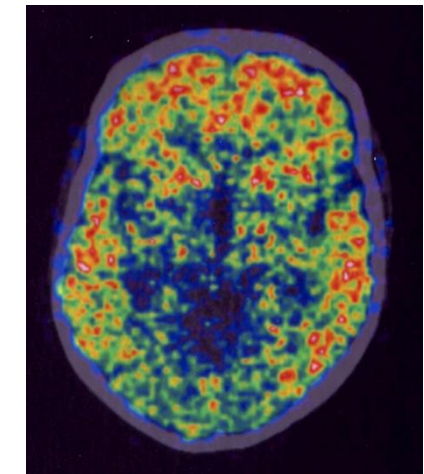
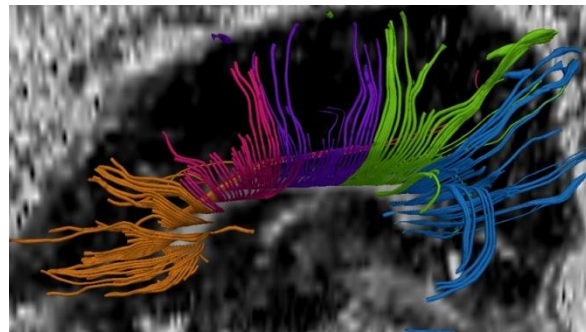
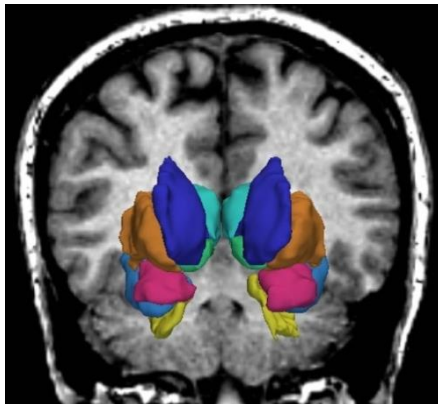
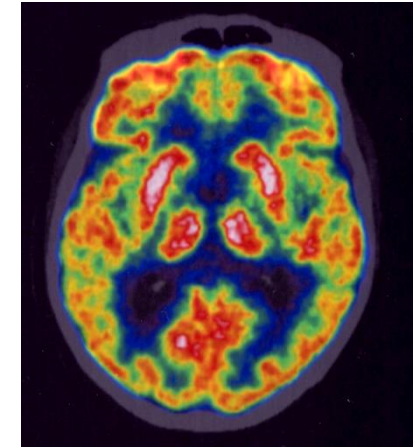
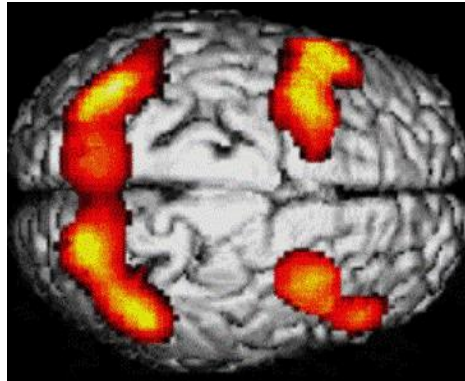
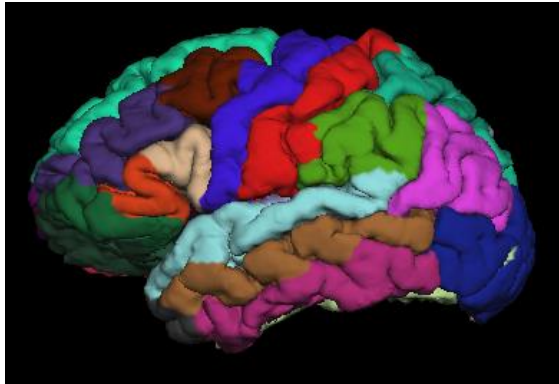


# Genotypes: Human Genome and SNP

- SNP (Single Nucleotide Polymorphism) - single nucleotide site where two or more different nucleotides occur in a large percentage of population.
- Total number of SNP: In the current dbSNP build, 132, the number of uniquely mapped refSNP (rs) numbers has grown to about 59 million+. (October 2011)

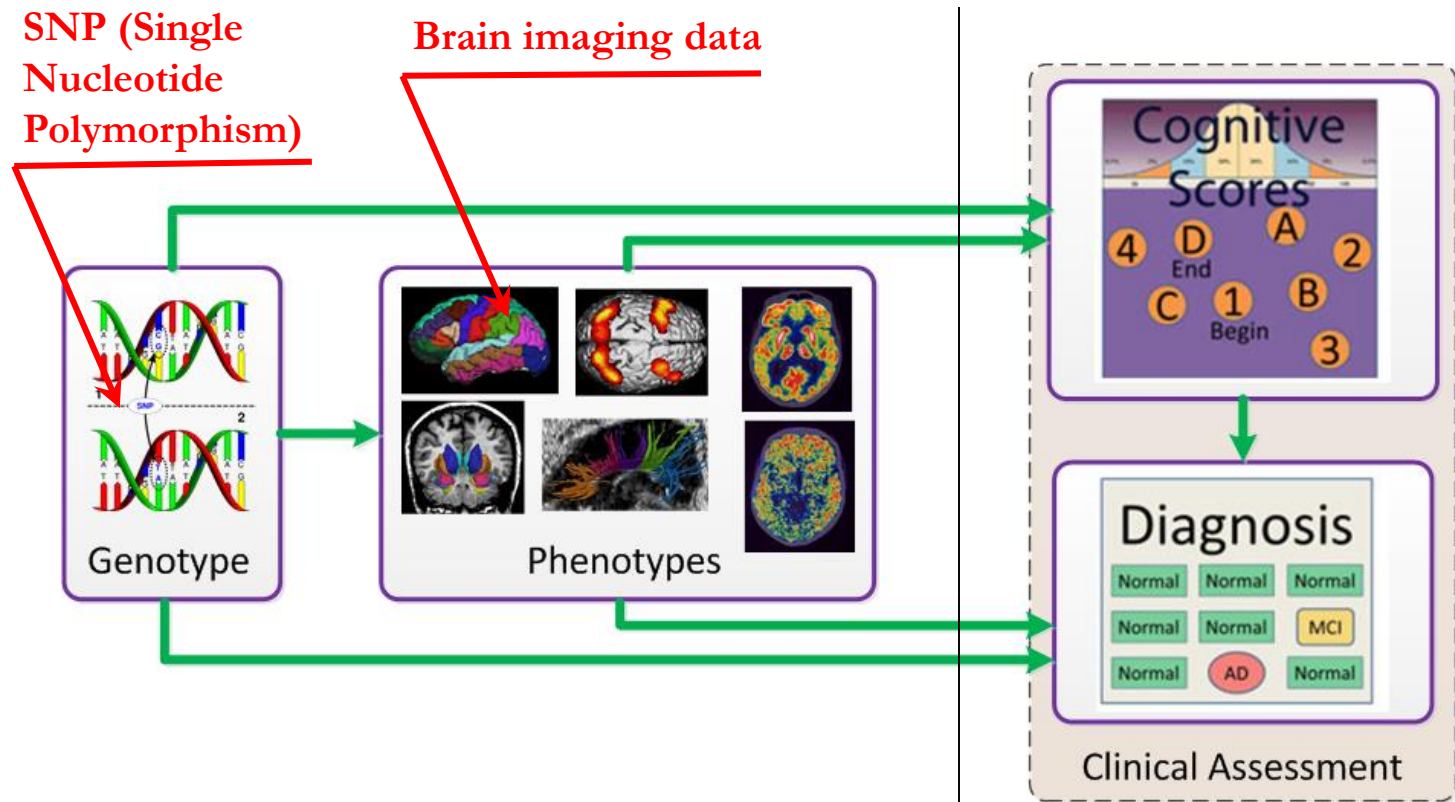


# Quantitative Phenotypes



Structural, functional, diffusion MRI  
FDG, PiB PET  
Fluid, cognitive biomarkers

# A Comprehensive Research Platform to Study Alzheimer's Disease - ADNI



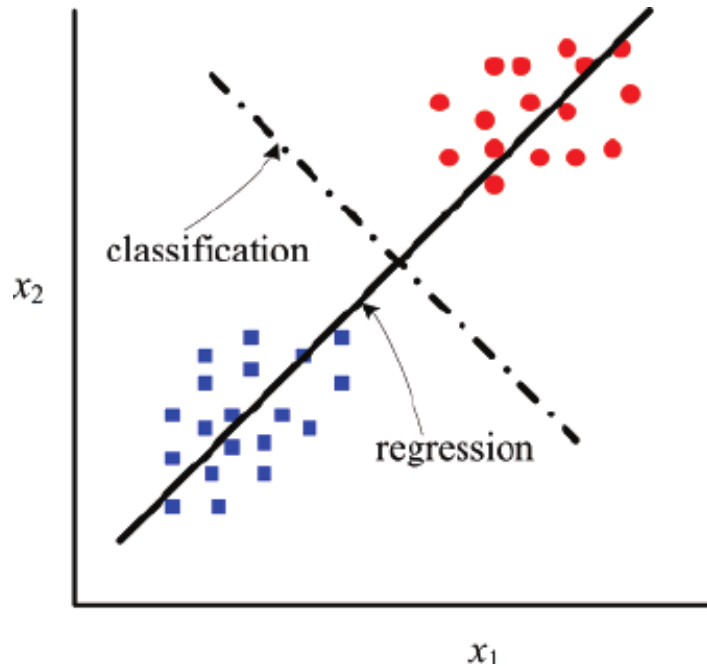
- NIH recently funded the ADNI (Alzheimer's Disease Neuroimaging Initiative) project, which has provided a comprehensive platform for imaging genetics studies.

Input

Output



# Association Studies in Statistical Learning



$$\min_{\mathbf{w}} f(\mathbf{w}; \mathbf{X}, \mathbf{y}) + r(\mathbf{w})$$

- **Classification:** associate a set of input samples into a set of predefined discrete targets.
- **Regression:** associate a set of input samples into a set of predefined continuous targets.

# Structured Sparse Learning for Biomarker Identification

---



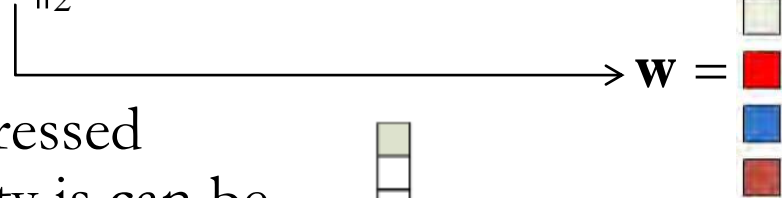
- Sparse learning has recently been successfully applied to solve a number of machine learning and data mining problems.
- Our structured sparse learning based biomarker selection approaches are able to
  - impose sparsity and
  - incorporate useful structural information contained in imaging and genetic data.

# Sparsity Is Achieved by $\ell_1$ -Norm Regularization



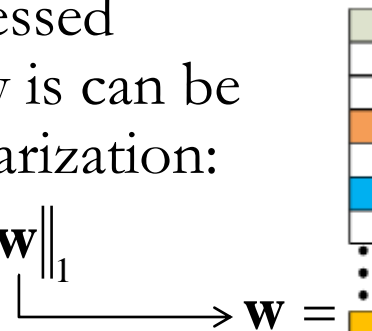
- In traditional statistical learning, in order to avoid over-fitting of the learning model,  $\ell_2$ -norm regularization is used, which leads to non-sparse result.

$$\min_{\mathbf{w}} f(\mathbf{w}; \mathbf{X}, \mathbf{y}) + a \|\mathbf{w}\|_2^2$$

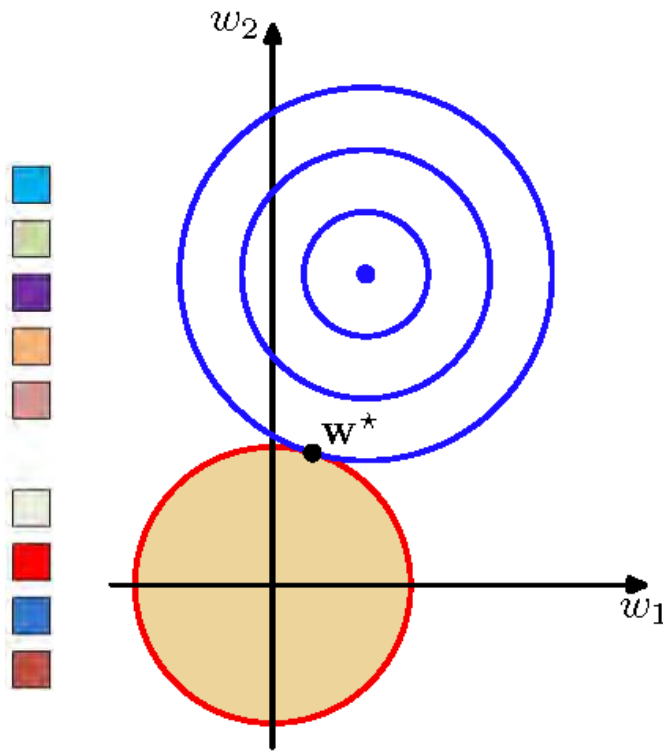


- Recent progress in compressed sensing shows that sparsity is can be achieved by  $\ell_1$ -norm regularization:

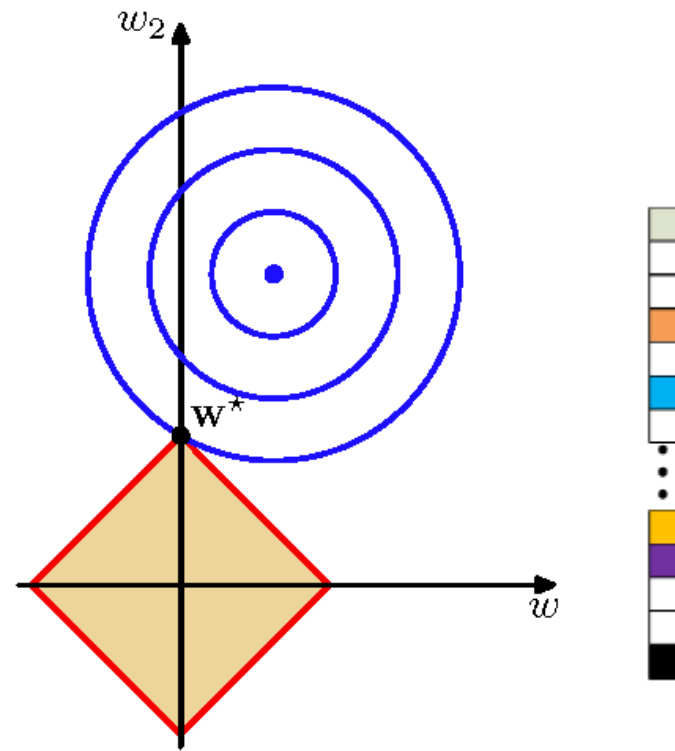
$$\min_{\mathbf{w}} f(\mathbf{w}; \mathbf{X}, \mathbf{y}) + a \|\mathbf{w}\|_1$$



# Sparsity Is Achieved by $\ell_1$ -Norm Regularization (cont.)



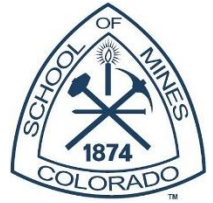
$$\min_{\mathbf{w}} f(\mathbf{w}; \mathbf{X}, \mathbf{y}) \quad s.t. \|\mathbf{w}\|_2^2 = c$$



$$\min_{\mathbf{w}} f(\mathbf{w}; \mathbf{X}, \mathbf{y}) \quad s.t. \|\mathbf{w}\|_1 = c$$



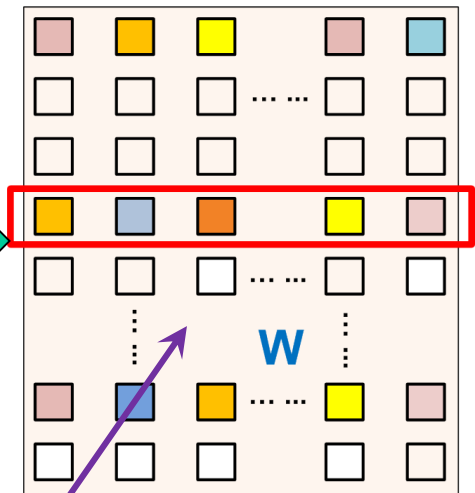
# Structured Sparsity Is Achieved by Mixed Norm Regularization



$\ell_2$ -norm regularization penalizes the involved coefficients as a whole

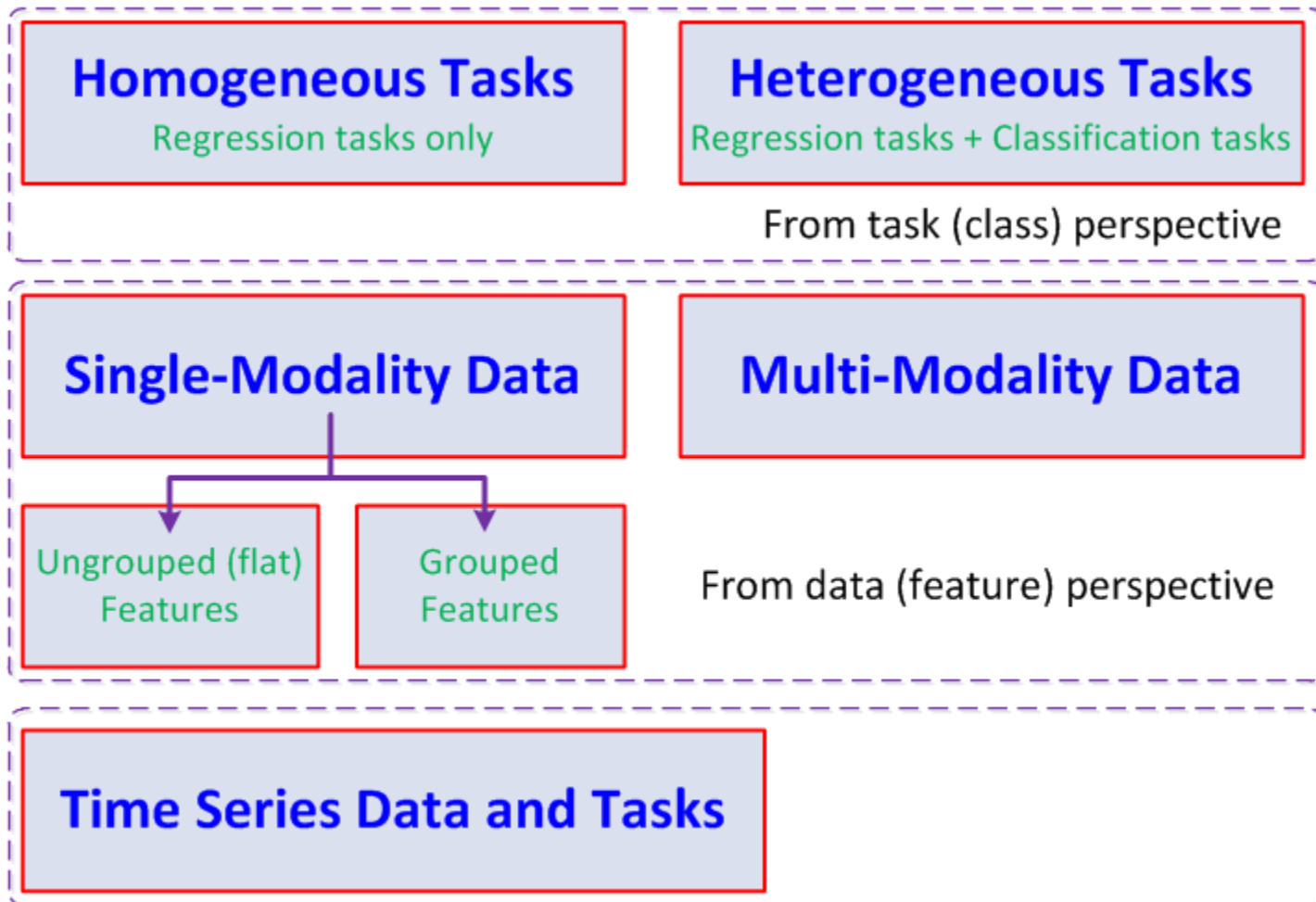
$\ell_1$ -norm regularization impose sparsity over the penalized coefficients

Mixed  $\ell_{21}$ -norm regularization impose structured sparsity

$$\|W\|_{2,1} = \sum_{i=1}^d \|\mathbf{w}^i\|_2$$


Either select an entire row, or discard an entire row

# A unified structured sparse learning framework for imaging genetics studies



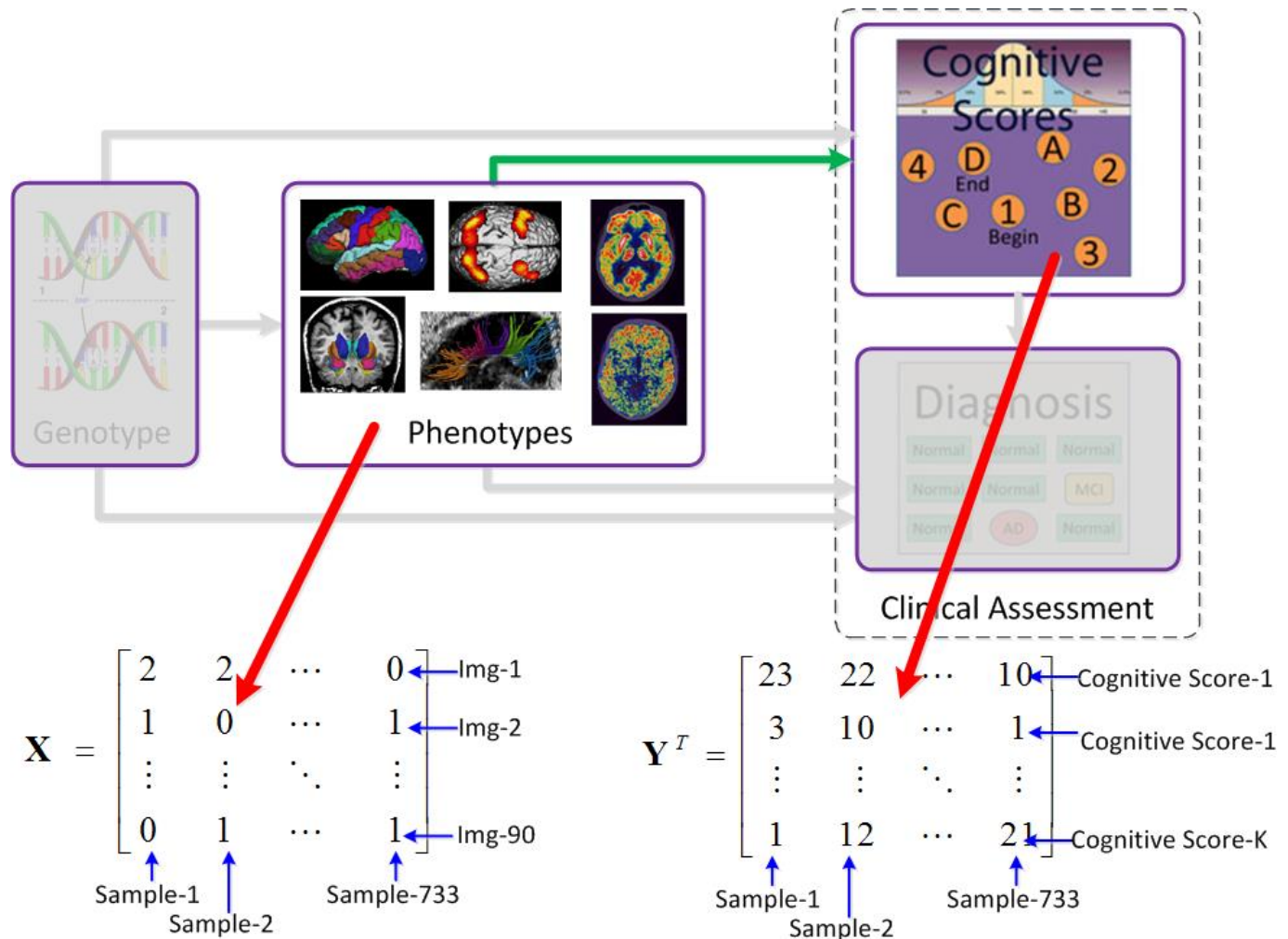


# Outline

---

- Background and motivation
- **Learning with homogeneous tasks**
- Learning with heterogeneous tasks
- Learning with group-structured single-modality data
- Learning with multi-modality data
- Learning with longitudinal data

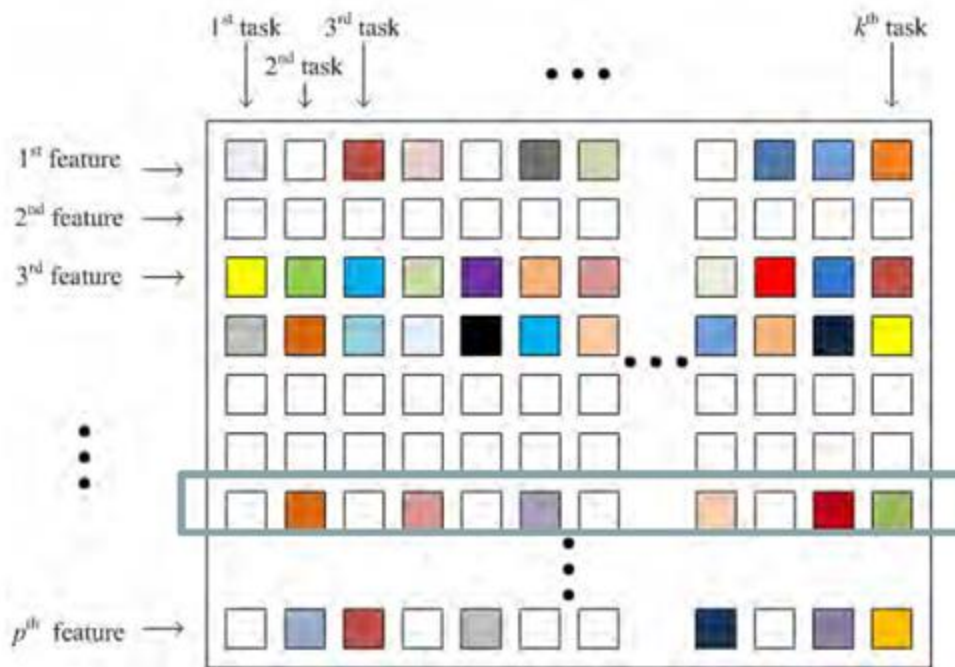
# Learning with Homogeneous Tasks





# A Convex Learning Objective

$$\min_W \left\| X^T W - Y \right\|_F^2 + \gamma_1 \|W\|_1 + \gamma_2 \|W\|_{2,1}$$



$$\lambda_1 \sum_{i=1}^p \|X^i\|_1 + \lambda_q \sum_{i=1}^p \|X^i\|_2$$

(Wang et al. ICCV'11)

# An Efficient Iterative Algorithm with Global Convergence



---

## Algorithm 1: Algorithm

---

**Input:**  $X, Y$

Initialize  $W^1 \in \mathbb{R}^{d \times c}, t = 1$  ;

**while** *not converge* **do**

1. Calculate the diagonal matrices

$D_i^{(t)} (1 \leq i \leq c)$  and  $\tilde{D}^{(t)}$ , where the  $k$ -th diagonal element of  $D_i^{(t)}$  is  $\frac{1}{2|w_{ki}^{(t)}|}$ , the  $k$ -th

diagonal element of  $\tilde{D}^{(t)}$  is  $\frac{1}{2\|(w^{(t)})^k\|_2}$  ;

2. For each  $i (1 \leq i \leq c)$ ,

$w_i^{(t+1)} = (XX^T + \gamma_1 D_i^{(t)} + \gamma_2 \tilde{D}^{(t)})^{-1} X y_i$  ;

3.  $t = t + 1$  ;

**Output:**  $W^{(t)} \in \mathbb{R}^{d \times c}$ .

(Wang *et al.* ICCV'11)

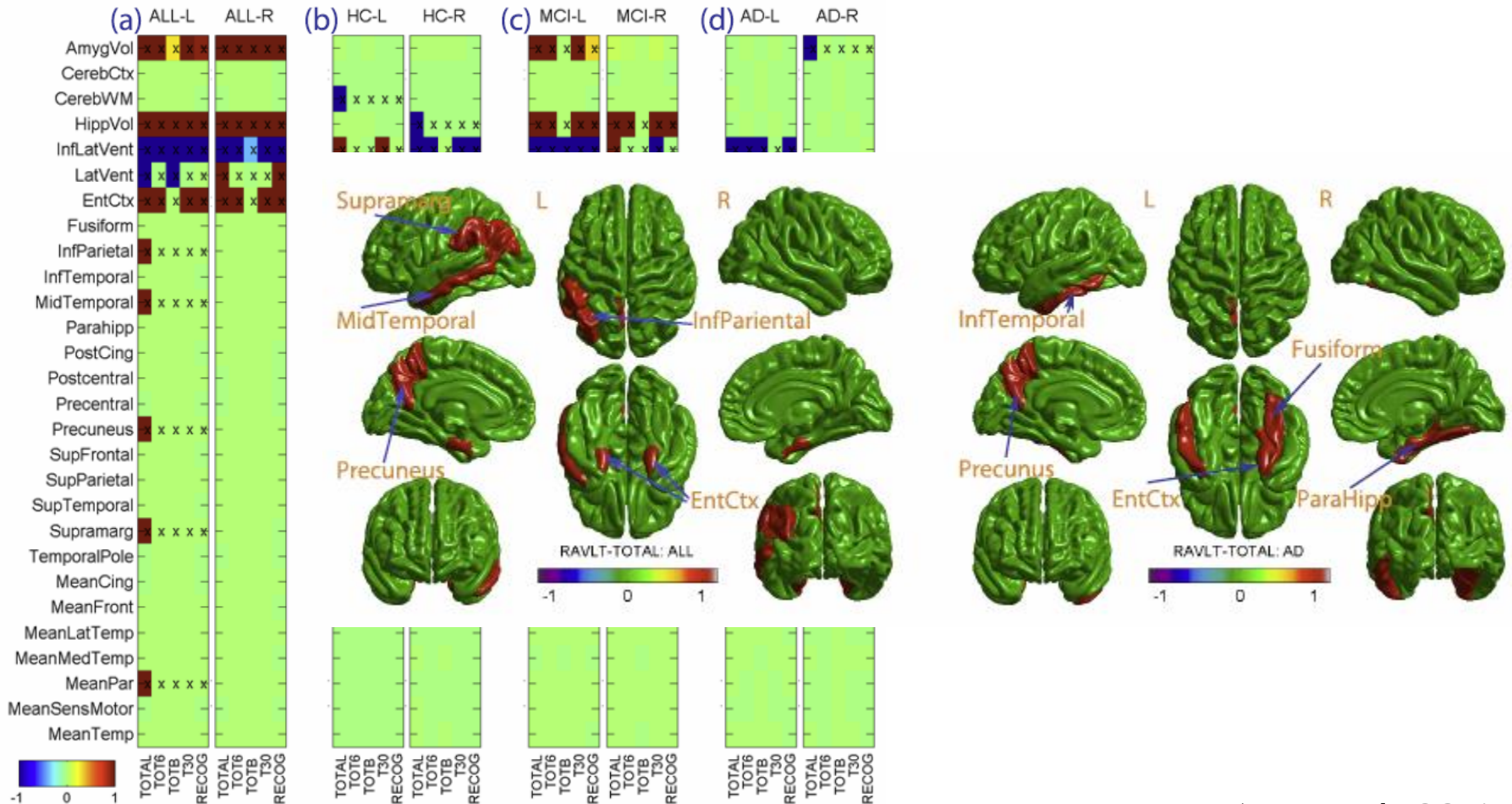
# Experimental Results

Table 2. Prediction performance measured by RMSE.

Test cases		TOTAL	TOT6	TOTB	T30	RECOG
FreeSurfer_HC	MVR	8.762	4.362	3.281	4.305	4.021
	SMART	6.645	2.940	2.235	2.806	3.621
FreeSurfer_MCI	MVR	6.998	2.765	2.399	2.480	3.427
	SMART	5.600	1.990	1.953	1.709	3.181
FreeSurfer_AD	MVR	5.897	1.768	2.058	1.382	3.390
	SMART	5.042	1.452	1.716	1.050	2.830
FreeSurfer_all	MVR	5.926	2.238	2.036	2.090	3.342
	SMART	5.736	2.139	1.961	1.966	3.196
VBM_HC	MVR	8.651	3.772	2.885	3.496	4.776
	SMART	6.705	2.844	2.139	2.656	3.584
VBM_MCI	MVR	11.495	4.256	4.621	4.032	5.598
	SMART	5.584	1.832	1.931	1.669	3.017
VBM_AD	MVR	7.223	2.162	2.622	1.479	4.163
	SMART	5.120	1.518	1.826	0.904	2.781
VBM_all	MVR	6.090	2.290	2.140	2.141	3.396
	SMART	5.718	2.103	1.993	1.921	3.182
VBM+FreeSurfer_HC	MVR	12.265	5.416	4.349	5.089	6.703
	SMART	6.664	2.829	2.230	2.683	3.577
VBM+FreeSurfer_MCI	MVR	68.222	26.146	23.489	30.033	34.306
	SMART	5.533	1.901	1.869	1.606	3.114
VBM+FreeSurfer_AD	MVR	14.552	4.307	5.141	4.297	8.430
	SMART	4.805	1.218	1.731	0.858	2.865
VBM+FreeSurfer_all	MVR	6.505	2.596	2.258	2.540	3.582
	SMART	5.809	2.208	2.000	2.051	3.214

(Wang *et al.* ICCV'11)

# Experimental Results (Cont.)



(Wang *et al.* ICCV'11)



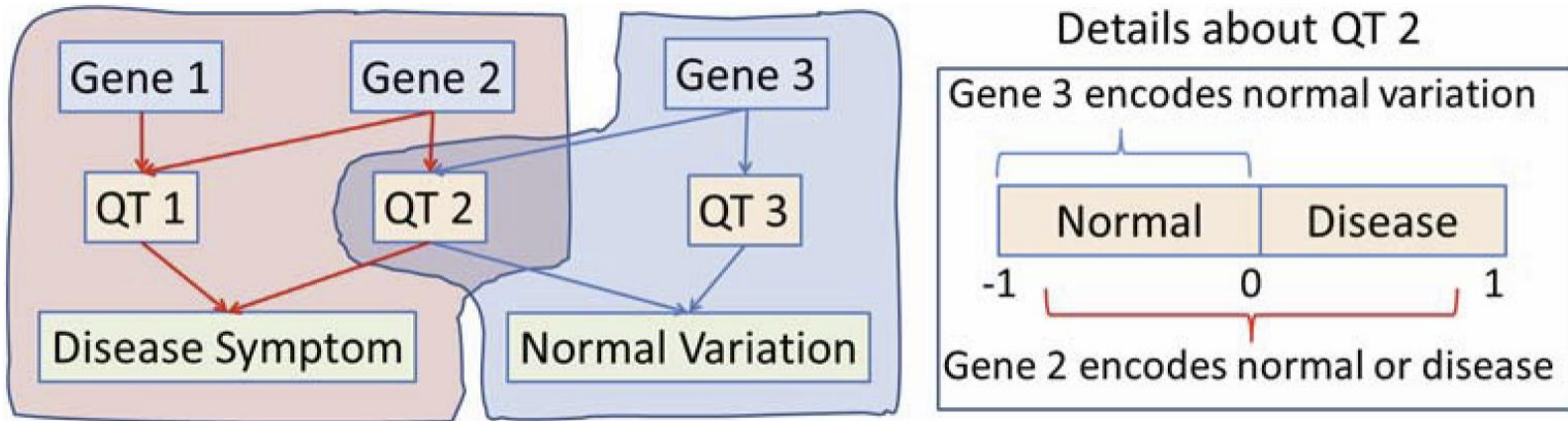


# Outline

---

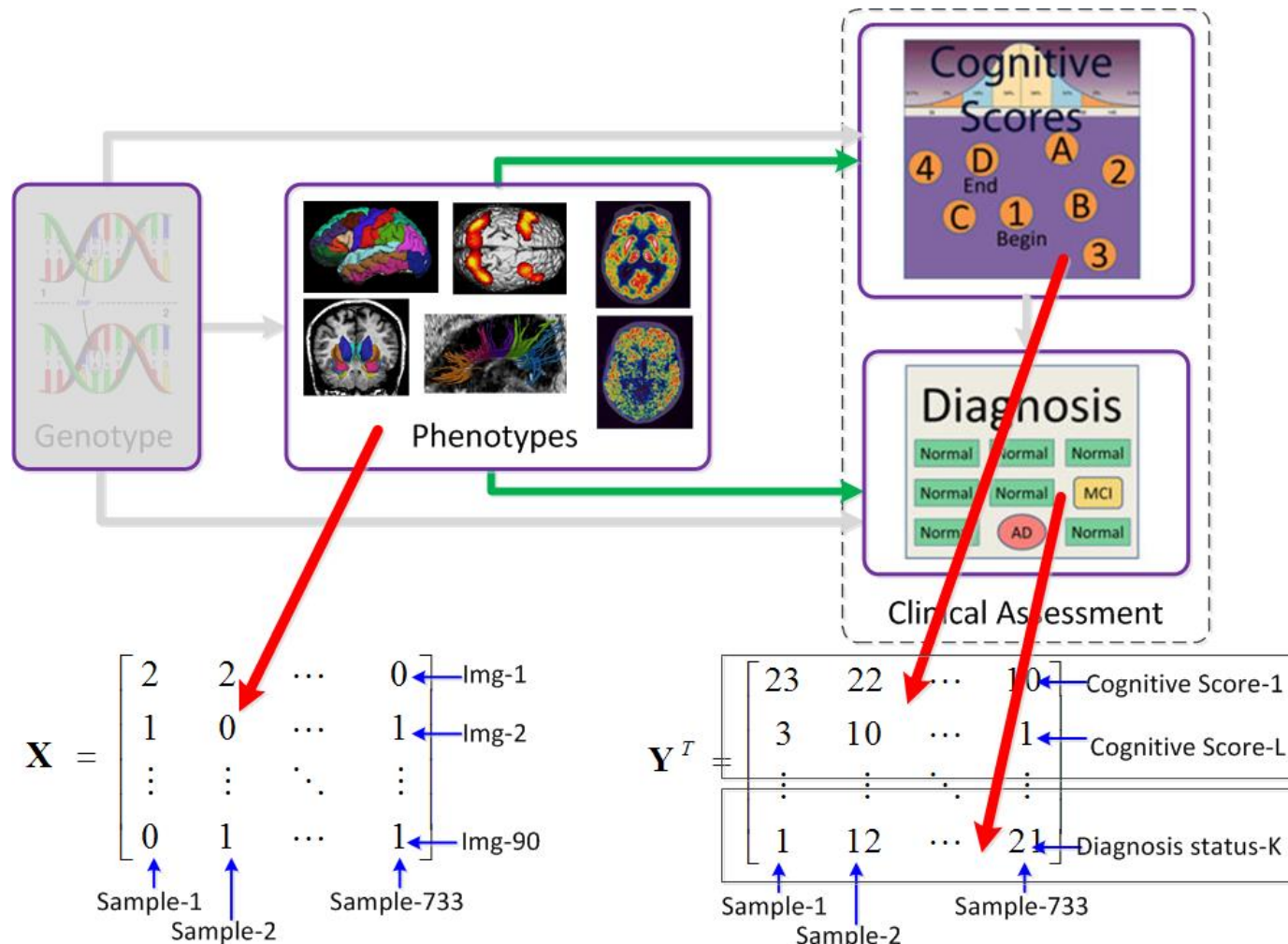
- Background and motivation
- Learning with homogeneous tasks
- **Learning with heterogeneous tasks**
- Learning with group-structured single-modality data
- Learning with multi-modality data
- Learning with longitudinal data

# Learning with Heterogeneous Tasks



- A simplified schematic example of two pathways from gene to QTs to phenotypic endpoints:
  - the red one is disease relevant
  - while the blue one yields only normal variation.
- Traditional two-stage imaging genetic strategy identifies QT 1 and QT 2 first and then Genes 1, 2, 3.
- Our new method will identify only disease relevant genes (i.e., Gene 1 and Gene 2); and Gene 3 won't be identified because it cannot be used to classify disease status.

# Learning with Heterogeneous Tasks (Cont.)



# Our Objective for Joint Classification and Regression



- Our joint learning objective

$$\min J(\mathbf{V}) = l_1(\mathbf{W}) + l_2(\mathbf{P}) + \gamma \|\mathbf{V}\|_{2,1},$$

where  $\mathbf{V} = [\mathbf{W} \ \mathbf{P}] \in \mathbf{R}^{d \times (c_1 + c_2)}$ .

- Logistic loss for disease status classification:

$$l_1(\mathbf{W}) = -\log \prod_{i=1}^n p(\mathbf{y}^i | \mathbf{x}_i, \mathbf{W}) = \sum_{i=1}^n \sum_{k=1}^{c_1} \left( y_{ik} \log \sum_{l=1}^{c_1} e^{\mathbf{w}_l^T \mathbf{x}_i} - y_{ik} \mathbf{w}_k^T \mathbf{x}_i \right).$$

- Least square loss for memory degradation score regression:

$$l_2(\mathbf{P}) = \|\mathbf{X}^T \mathbf{P} - \mathbf{Z}\|_F^2,$$

(Wang *et al.* MICCAI'11)



# An Efficient Iterative Algorithm with Global Convergence



---

## Algorithm 1. An efficient algorithm to solve Eq. (3)

---

**Input:**  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n] \in \mathbb{R}^{d \times n}$ ,  $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_n]^T \in \mathbb{R}^{n \times c_1}$ , and  $\mathbf{Z} = [\mathbf{z}_1, \dots, \mathbf{z}_n]^T \in \mathbb{R}^{n \times c_2}$ .

1. Initialize  $\mathbf{W} \in \mathbb{R}^{d \times c_1}$ ,  $\mathbf{P} \in \mathbb{R}^{d \times c_2}$ , and let  $\mathbf{V} = [\mathbf{W} \ \mathbf{P}] \in \mathbb{R}^{d \times (c_1 + c_2)}$ ;

**while not converge do**

2. Calculate the diagonal matrix  $\mathbf{D}$ , of which the  $k$ -th element is  $\frac{1}{2\|\mathbf{v}^k\|_2}$ ;

3. Update  $\mathbf{w}$  by  $\mathbf{w} - \mathbf{B}^{-1}\mathbf{a}$ , where  $(d \times (p-1) + u)$ -th element of  $\mathbf{a} \in \mathbb{R}^{dc_1 \times 1}$  is

$\frac{\partial[l_1(\mathbf{W}) + \gamma \text{tr}(\mathbf{W}^T \mathbf{D} \mathbf{W})]}{\partial \mathbf{W}_{up}}$  for  $1 \leq u \leq d, 1 \leq p \leq c_1$ , the  $(d \times (p-1) + u, d \times (q-1) + v)$ -th

element of  $\mathbf{B} \in \mathbb{R}^{dc_1 \times dc_1}$  is  $\frac{\partial[l_1(\mathbf{W}) + \gamma \text{tr}(\mathbf{W}^T \mathbf{D} \mathbf{W})]}{\partial \mathbf{W}_{up} \partial \mathbf{W}_{vq}}$  for  $1 \leq u, v \leq d$  and  $1 \leq p, q \leq c_1$ .

Construct the updated  $\mathbf{W} \in \mathbb{R}^{d \times c_1}$  by the updated vector  $\mathbf{w} \in \mathbb{R}^{dc_1}$ , where the  $(u, p)$ -th element of  $\mathbf{W}$  is the  $(d \times (p-1) + u)$ -th element of  $\mathbf{w}$ ;

4. Update  $\mathbf{P}$  by  $\mathbf{P} = (\mathbf{X}\mathbf{X}^T + \gamma\mathbf{D})^{-1} \mathbf{X}\mathbf{Z}$ ;

5. Update  $\mathbf{V}$  by  $\mathbf{V} = [\mathbf{W} \ \mathbf{P}]$ ;

**end**

**Output:**  $\mathbf{W} \in \mathbb{R}^{d \times c_1}$  and  $\mathbf{P} \in \mathbb{R}^{d \times c_2}$ .

---

(Wang *et al.* MICCAI'11)



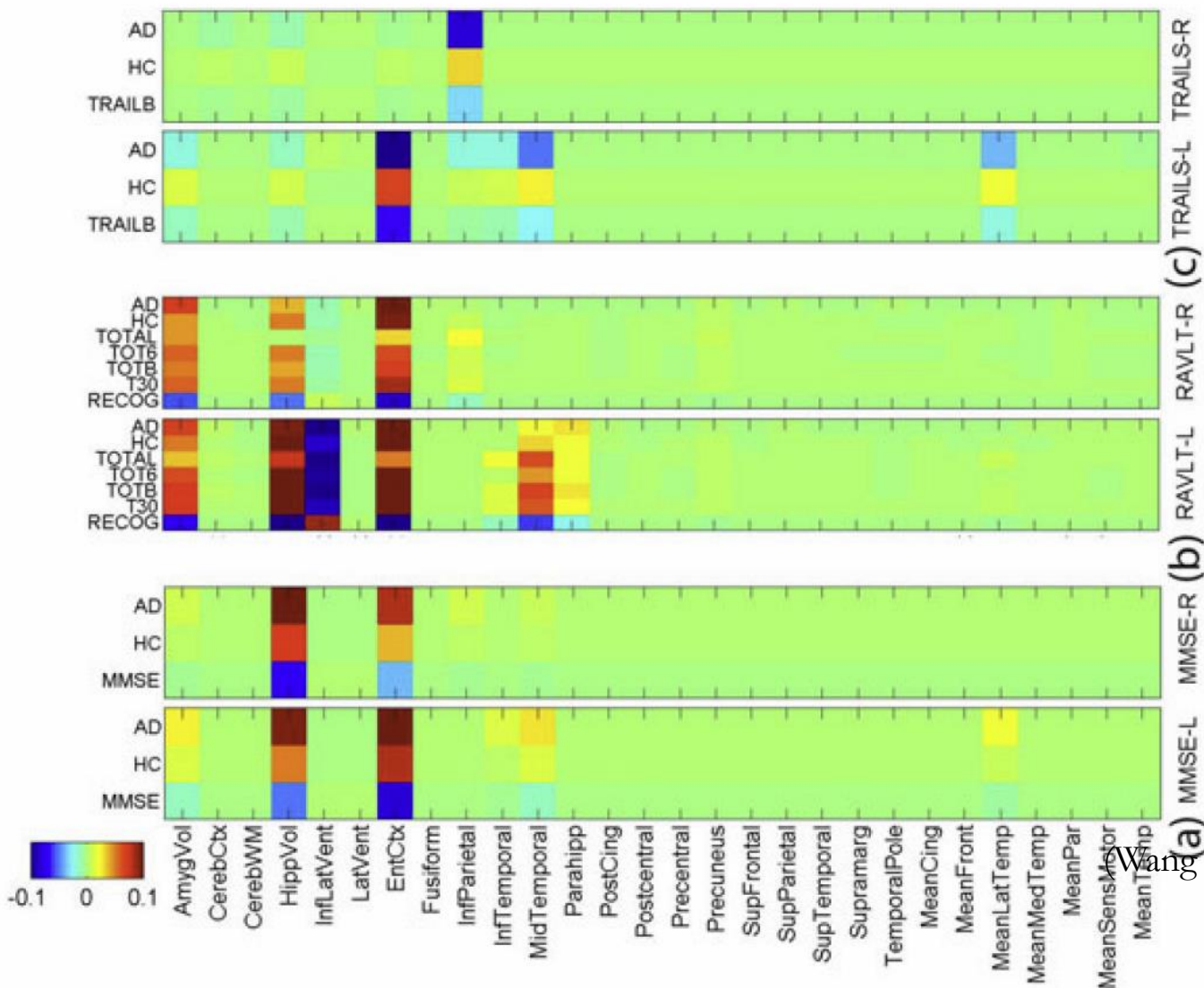
# Experimental Results

**Table 1.** Comparison of classification and regression performance

				Our method		Classification accuracy		RMSE (mean $\pm$ std)	
Memory score	# subjects	# AD	# HC	Classification accuracy	Regression RMSE	Logistic regression	SVM	Multivariate regression	Ridge regression
MMSE	378	175	203	0.881	0.034 $\pm$ 0.002	0.832	0.783 (linear kernel)	0.041 $\pm$ 0.003	0.039 $\pm$ 0.004
RAVLT	371	172	199	0.884	0.019 $\pm$ 0.001		0.839 (Polynomial kernel)	0.028 $\pm$ 0.002	0.024 $\pm$ 0.003
TRAILS	369	166	203	0.864	0.043 $\pm$ 0.002		0.796 (Gausssian kernel)	0.049 $\pm$ 0.003	0.046 $\pm$ 0.003

(Wang *et al.* MICCAP'11)

# Experimental Results (Cont.)



Wang et al. MICCAI'11)

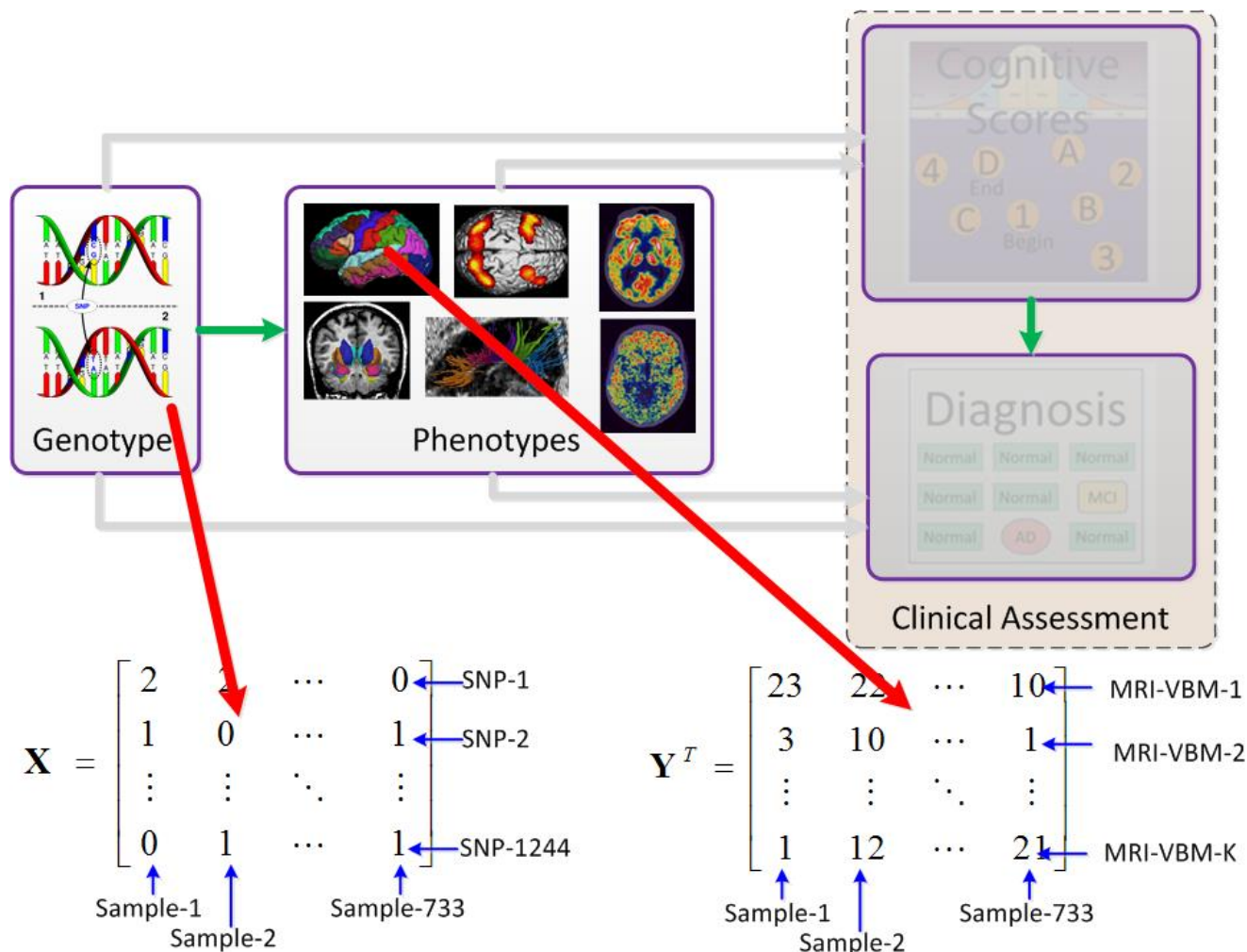


# Outline

---

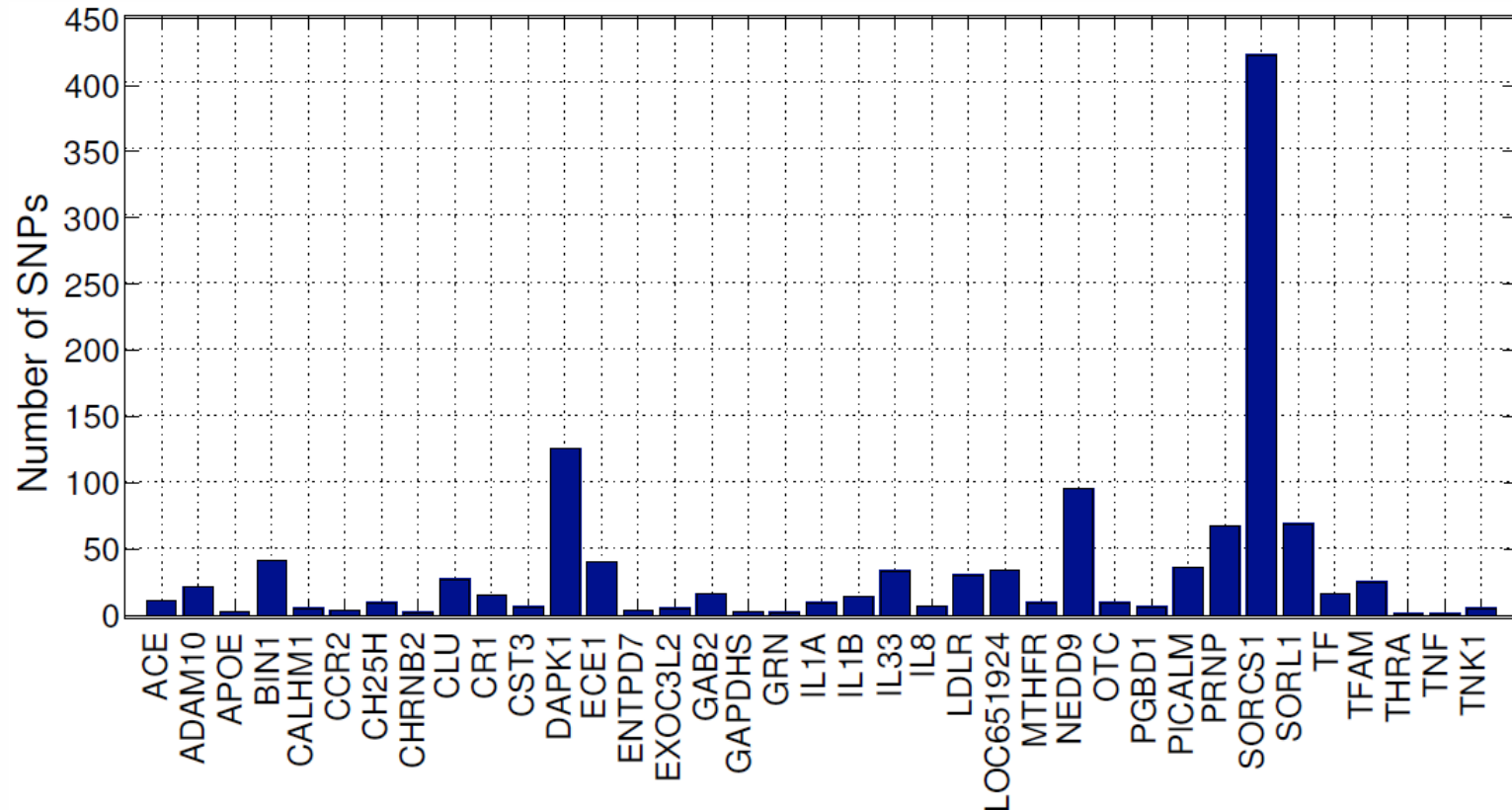
- Background and motivation
- Learning with homogeneous tasks
- Learning with heterogeneous tasks
- **Learning with group-structured single-modality data**
- Learning with multi-modality data
- Learning with longitudinal data

# From Genotype to Phenotype



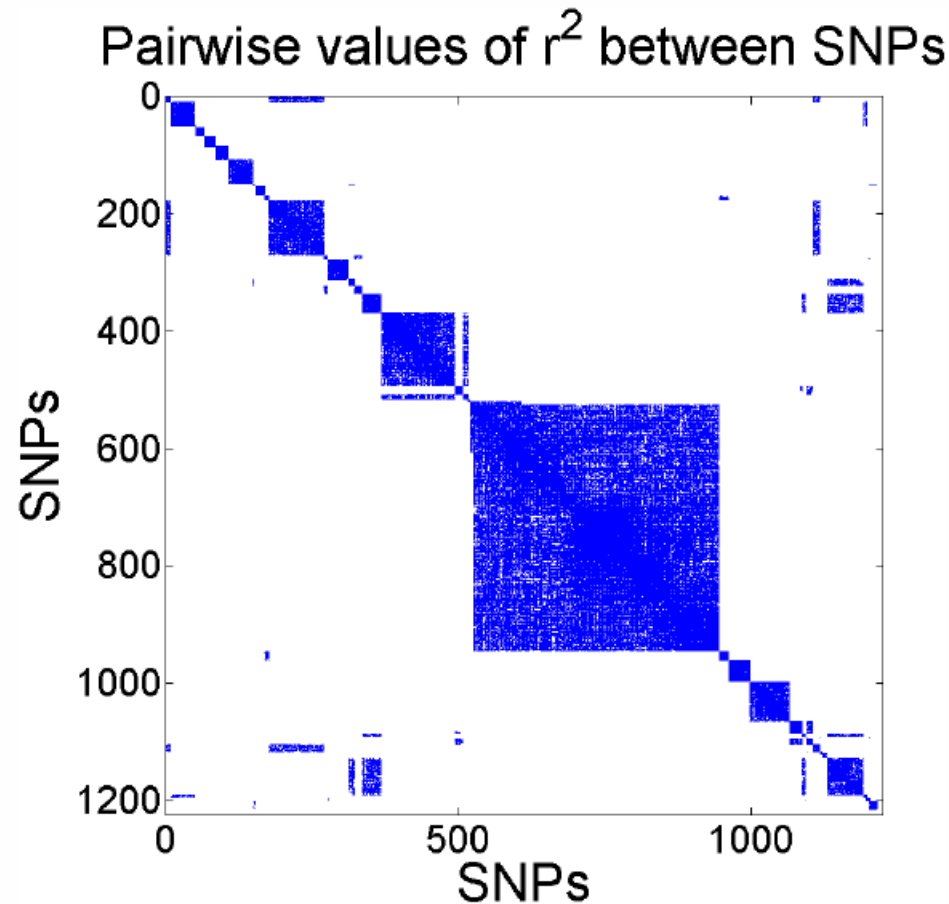


# SNPs are Genetically Linked with Group Structures — Genes



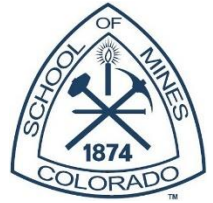
(Wang *et al.* Bioinformatics'12)

# SNPs are Genetically Linked with Group Structures — Linkage Disequilibrium (LD)

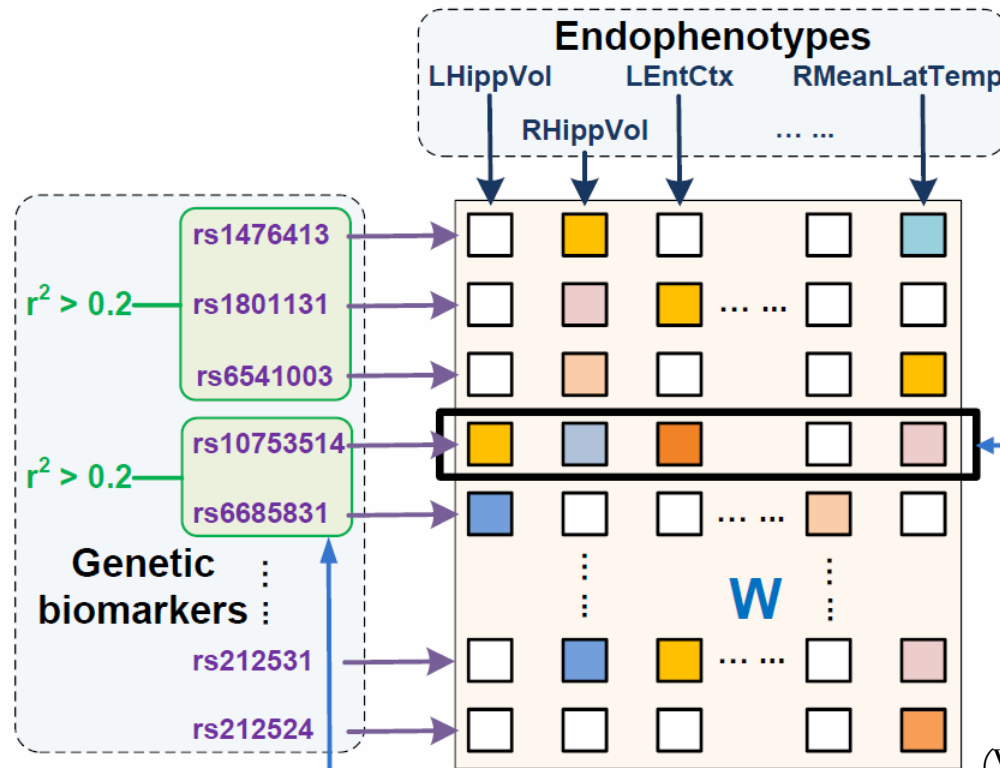


(Wang *et al.* Bioinformatics'12)

# A Novel Group-2,1 Norm to Capture Grouping Structures



$$\min_{\mathbf{W}} \sum_{i=1}^n ||\mathbf{W}^T \mathbf{X} - \mathbf{Y}||_F^2 + \gamma_1 ||\mathbf{W}||_{G_{2,1}} + \gamma_2 ||\mathbf{W}||_{2,1}$$



$$||\mathbf{W}||_{G_{2,1}} = \sum_{k=1}^K \sqrt{\sum_{i \in \pi_k} ||\mathbf{w}^i||_2^2}$$

$$||\mathbf{W}||_{2,1} = \sum_{i=1}^d ||\mathbf{w}^i||_2$$

(Wang *et al.* Bioinformatics'12)

# An Efficient Iterative Algorithm with Global Convergence



**Input:**  $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbb{R}^{d \times n}$ ,

$\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n] \in \mathbb{R}^{c \times n}$

Initialize  $\mathbf{W}_1 \in \mathbb{R}^{d \times c}$ ,  $t = 1$ ;

**while** *not converge* **do**

1. Calculate the block diagonal matrix  $\mathbf{D}_t$ , where the  $k$ -th diagonal is  $\frac{1}{2\|\mathbf{w}_t^k\|_F} \mathbf{I}_k$ ; Calculate the diagonal matrix  $\tilde{\mathbf{D}}_t$ , where the  $i$ -th diagonal element is  $\frac{1}{2\|\mathbf{w}_t^i\|_2}$ ;

2.  $\mathbf{W}_{t+1} = (\mathbf{X}\mathbf{X}^T + \gamma_1 \mathbf{D}_t + \gamma_2 \tilde{\mathbf{D}}_t)^{-1} \mathbf{X}\mathbf{Y}^T$ ;

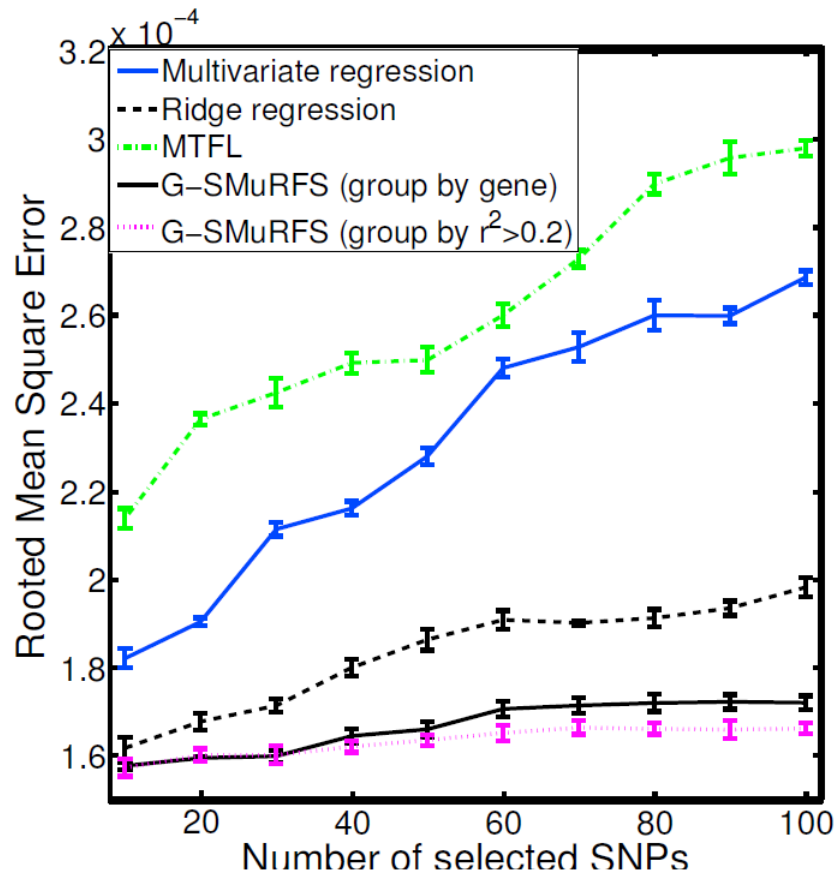
3.  $t = t + 1$ ;

**end**

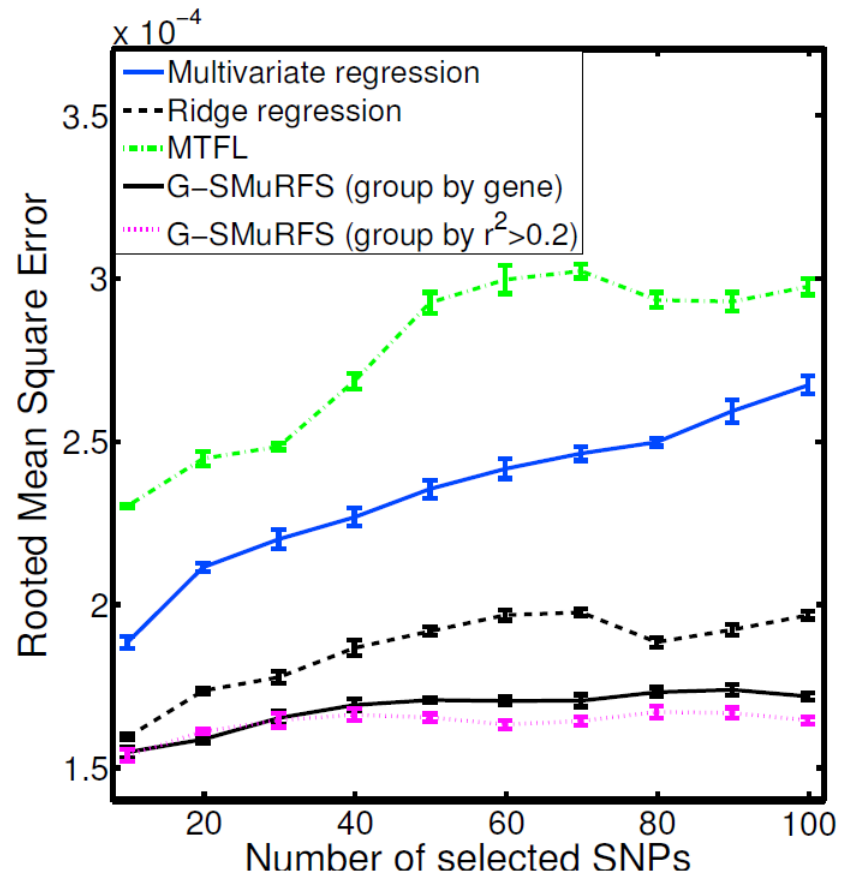
**Output:**  $\mathbf{W}_t \in \mathbb{R}^{d \times c}$ .

(Wang *et al.* Bioinformatics'12)

# Experimental Results



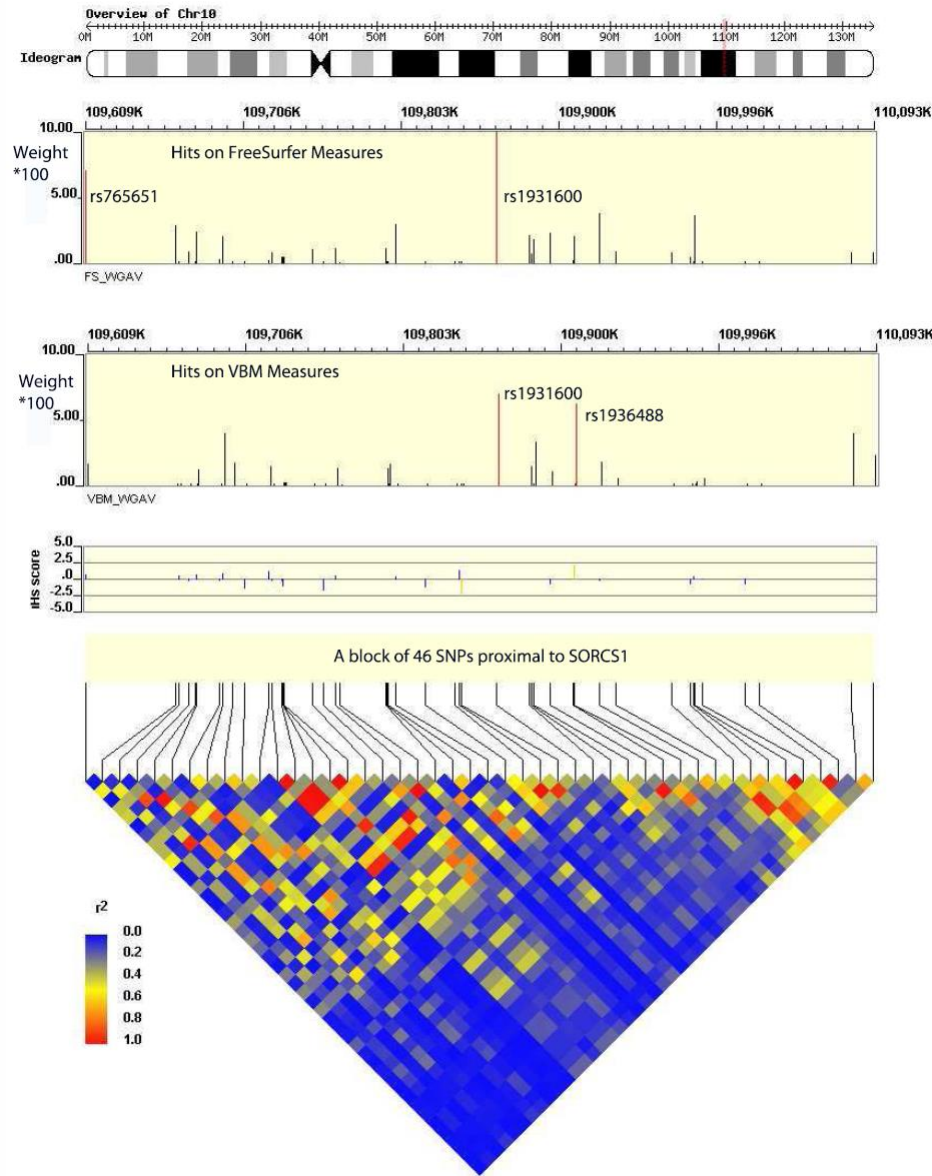
(a) FreeSurfer imaging genotypes.



(b) VBM imaging genotypes.

(Wang *et al.* Bioinformatics'12)

# Experimental Results (Cont.)



(Wang *et al.* Bioinformatics'12)



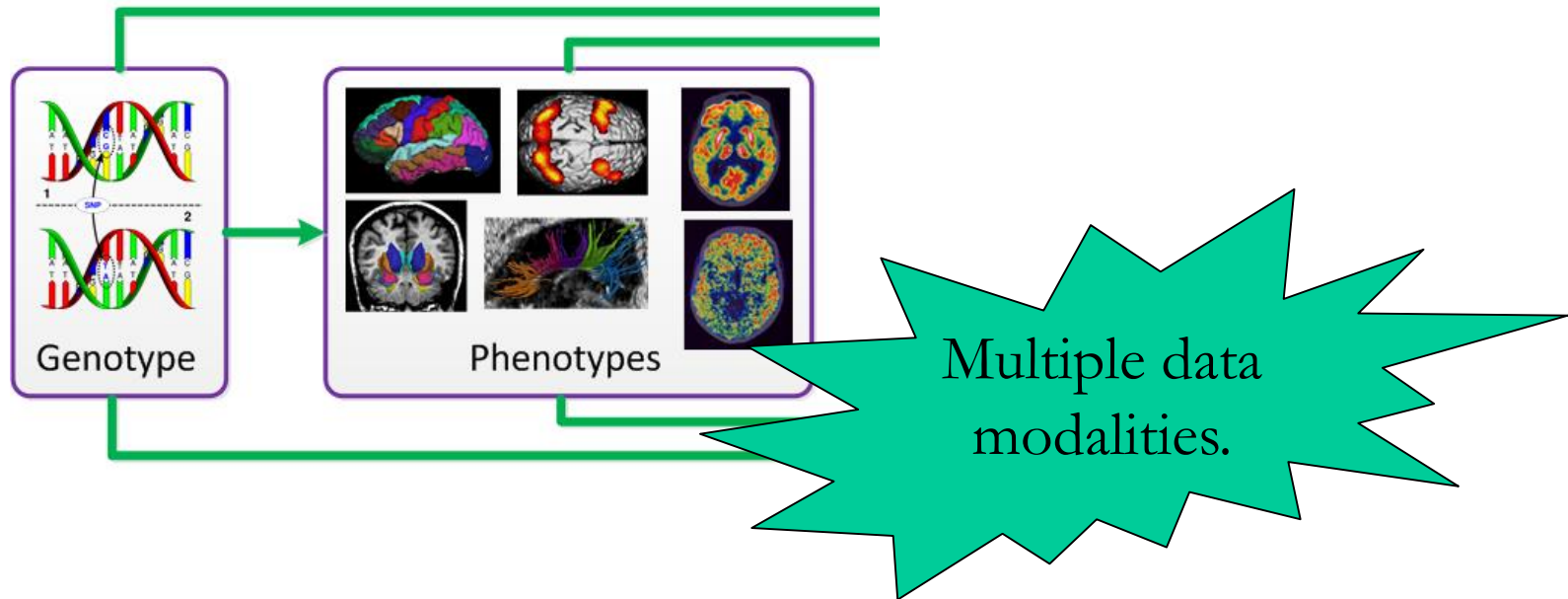


# Outline

---

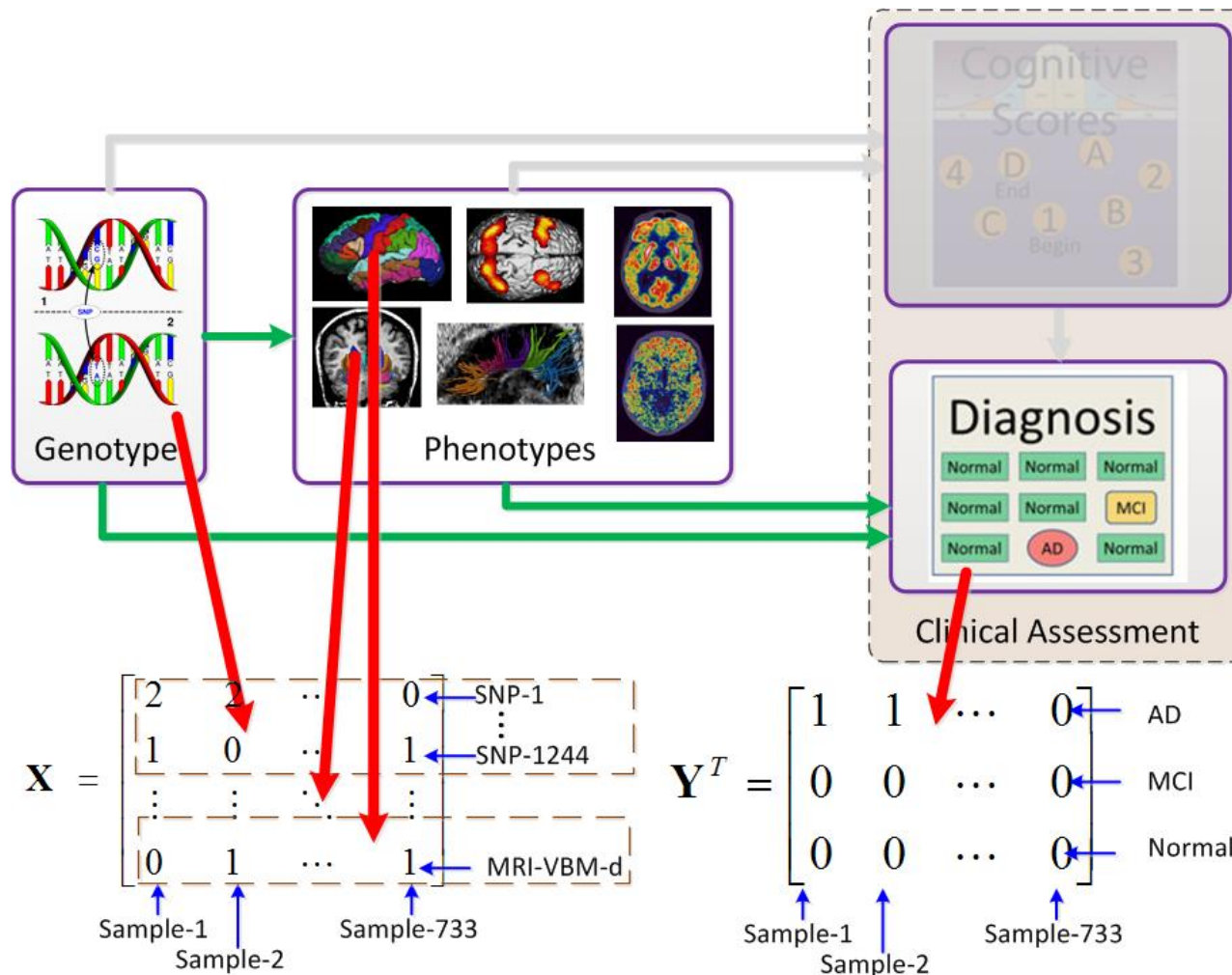
- Background and motivation
- Learning with homogeneous tasks
- Learning with heterogeneous tasks
- Learning with group-structured single-modality data
- **Learning with multi-modality data**
- Learning with longitudinal data

# Integration of Genotypes and Phenotypes to Select Multi-Modal Biomarkers



Feature Selection + Data Fusion

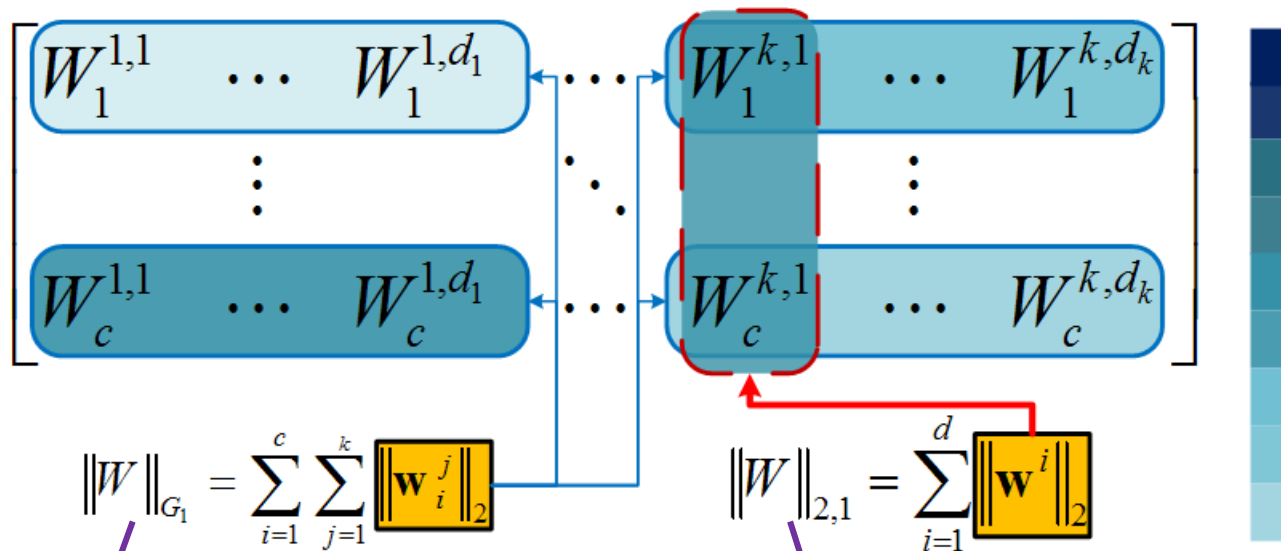
# Integration of Genotypes and Phenotypes to Select Multi-Modal Biomarkers



# New Objective Function for Multi-Modal Biomarker Integration



$$\min_{\mathbf{W}} \|\mathbf{X}^T \mathbf{W} - \mathbf{Y}\|_F^2 + \gamma_1 \|\mathbf{W}\|_{G_1} + \gamma_2 \|\mathbf{W}\|_{2,1}$$



Learning proper weight for each modality

Learning proper weight for each individual feature

(Wang *et al.* ISMB'12)

# New Objective Function for Multi-Modal Biomarker Integration (cont.)



- The formulated objective is highly non-smooth due to the  $\ell_{21}$ -norm regularization term and our new group  $\ell_1$ -norm regularization term.

**Input:**  $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbb{R}^{d \times n}$ ,  $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_c] \in \mathbb{R}^{n \times c}$ .  
**Output:**  $\mathbf{W}_t \in \mathbb{R}^{d \times c}$ .  
1. Let  $t = 1$ . Initialize  $\mathbf{W}_t \in \mathbb{R}^{d \times c}$ .  
**repeat**  
2. Calculate the block diagonal matrices  $\mathbf{D}_t^i$  ( $1 \leq i \leq c$ ), where the  $j$ -th diagonal block of  $\mathbf{D}_t^i$  is  $\frac{1}{2\|(\mathbf{w}_t)_i^j\|_2} \mathbf{I}_j$ . Calculate the diagonal matrix  $\tilde{\mathbf{D}}_t$ , where the  $i$ -th diagonal element is  $\frac{1}{2\|\mathbf{w}_t^i\|_2}$ .  
3. For each  $\mathbf{w}_i$  ( $1 \leq i \leq c$ ),  $(\mathbf{w}_{t+1})_i = (\mathbf{X}\mathbf{X}^T + \gamma_1 \mathbf{D}_t^i + \gamma_2 \tilde{\mathbf{D}}_t)^{-1} \mathbf{X}\mathbf{y}_i$ .  
4.  $t = t + 1$ .  
**until** Converges

Simple, efficient,  
easy to implement

**Theorem:** The objective value of

$$\min_{\mathbf{W}} \|\mathbf{X}^T \mathbf{W} - \mathbf{Y}\|_F^2 + \gamma_1 \|\mathbf{W}\|_{G_1} + \gamma_2 \|\mathbf{W}\|_{2,1}$$

decreases in each iteration of the algorithm.

(Wang *et al.* ISMB'12)

# Improved Early Alzheimer's Disease Detection

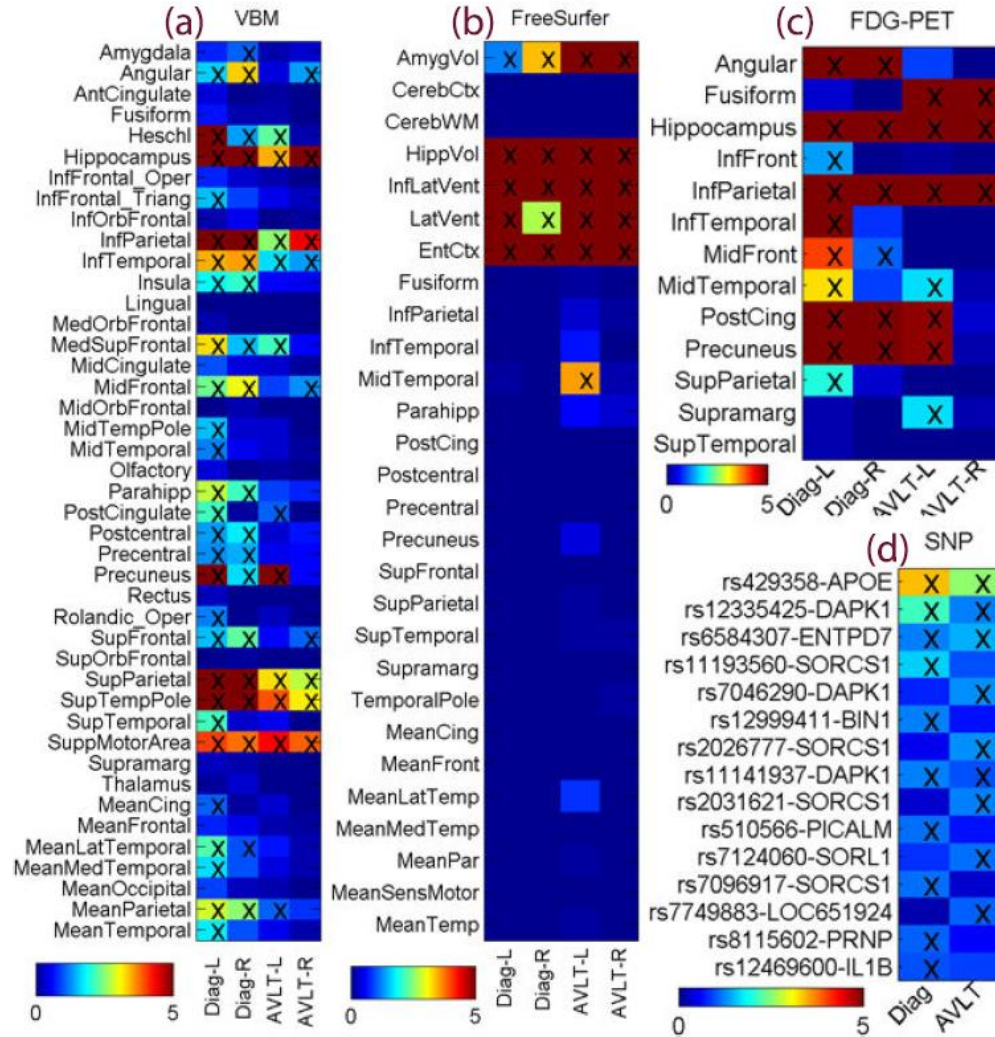


Methods	Accuracy (mean+std)
SVM (SNP)	0.561 $\pm$ 0.026
SVM (FreeSurfer)	0.573 $\pm$ 0.012
SVM (VBM)	0.541 $\pm$ 0.032
SVM (PET)	0.535 $\pm$ 0.026
SVM (all)	0.575 $\pm$ 0.019
SVM $\ell_\infty$ MKL method	0.624 $\pm$ 0.031
SVM $\ell_1$ MKL method	0.593 $\pm$ 0.042
SVM $\ell_2$ MKL method	0.561 $\pm$ 0.037
LSSVM $\ell_\infty$ MKL method	0.614 $\pm$ 0.031
LSSVM $\ell_1$ MKL method	0.585 $\pm$ 0.018
LSSVM $\ell_2$ MKL method	0.577 $\pm$ 0.033
Our method	<b>0.674 <math>\pm</math> 0.021</b>

(Wang *et al.* ISMB'12)



# Biomarker Selection



(Wang *et al.* ISMB'12)

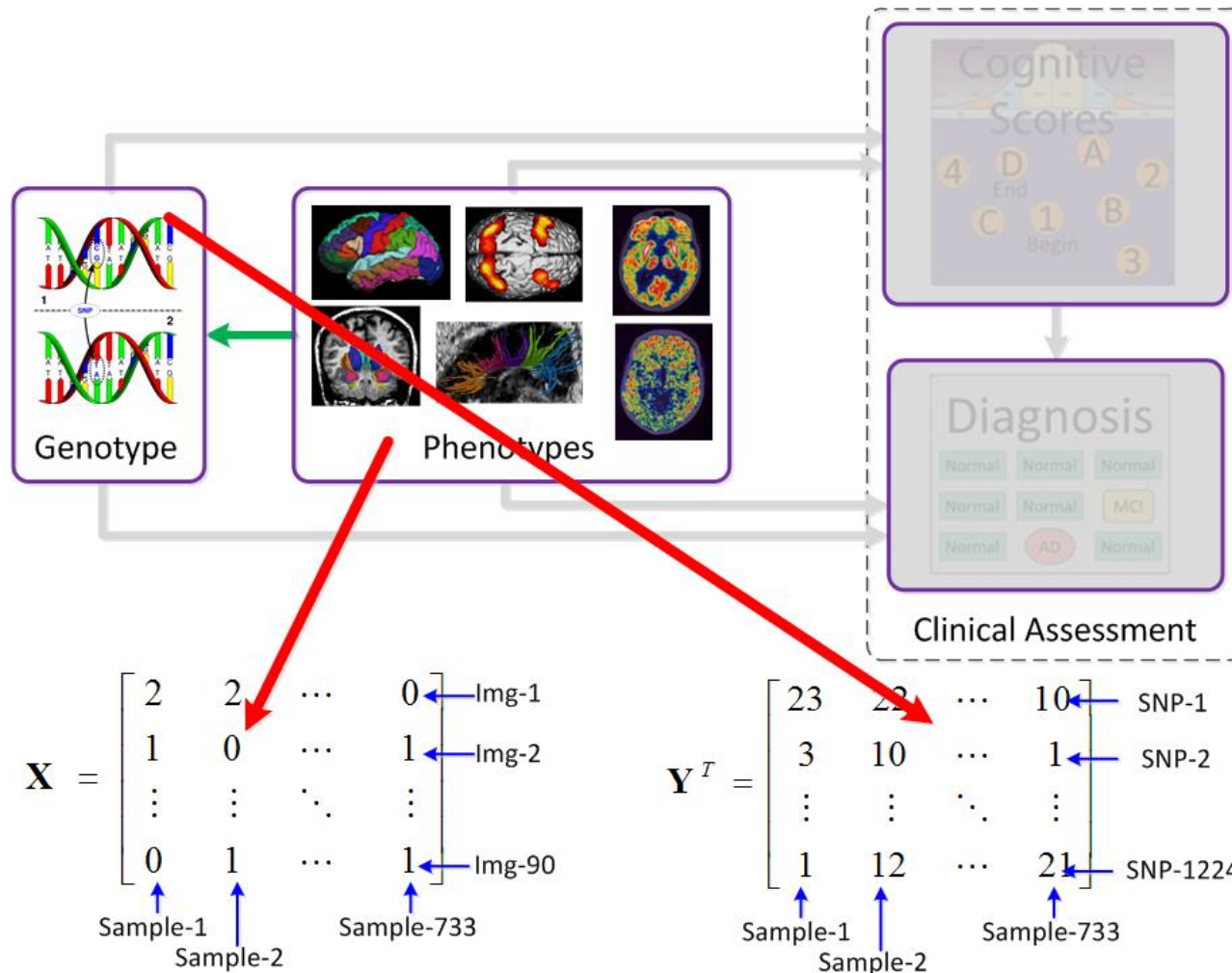


# Outline

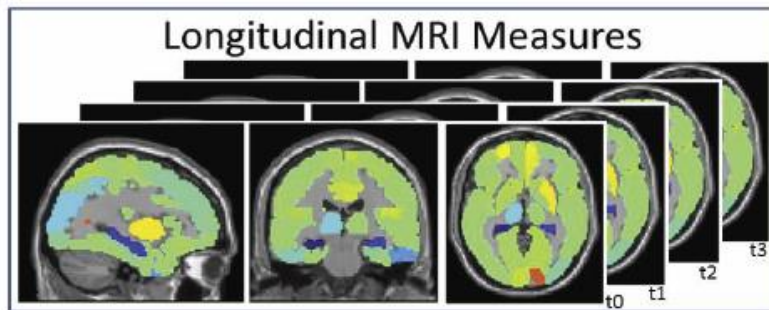
---


- Background and motivation
- Learning with homogeneous tasks
- Learning with heterogeneous tasks
- Learning with group-structured single-modality data
- Learning with multi-modality data
- **Learning with longitudinal data**

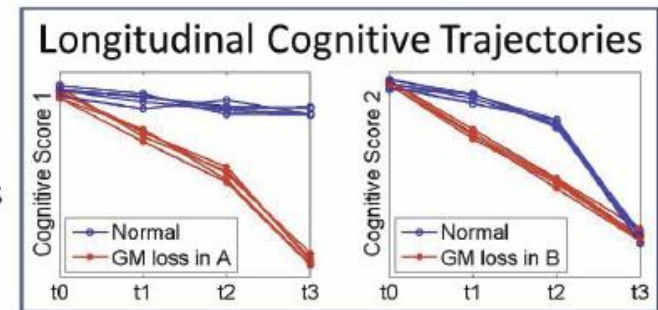
# Association from Phenotypes to Genotypes



# Longitudinal Multi-Task Regression Model



Longitudinal multi-task regression  
  
 to select MRI markers  
 predicting cognitive  
 trajectories



# Learning A Tensor of Regression Model

- We need to learn  $T$  regression coefficient matrices, which forms up a tensor:

$$\min_{\mathcal{B}} J_2 = \mathcal{L}(\mathcal{B}) + \gamma_1 \sum_{k=1}^d \sqrt{\sum_{t=1}^T \|\mathbf{b}_t^k\|_2^2} + \gamma_2 \|\mathcal{B}\|_*$$

where longitudinal loss is defined as:

$$\mathcal{L}(\mathcal{B}) = \|\mathcal{B} \otimes_1 \mathcal{X}^T - Y\|_F^2 = \sum_{t=1}^T \|X_t^T B_t - Y\|_F^2.$$

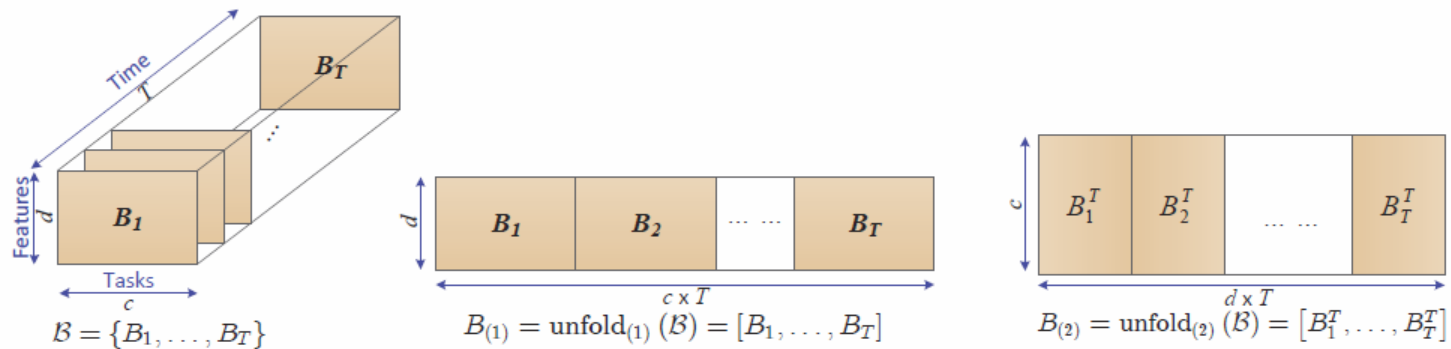


Figure 2: **Left:** visualization of the coefficient tensor  $\mathcal{B}$  learned for the association study on longitudinal data. **Middle:** the matrix unfolded from  $\mathcal{B}$  along the first mode (feature dimension). **Right:** the matrix unfolded from  $\mathcal{B}$  along the second mode (task dimension).

# An Efficient Iterative Algorithm with Global Convergence



---

**Algorithm 1:** A new algorithm to minimize  $J_2$  in Eq. (4).

---

**Data:**  $\mathcal{X} \in \mathbb{R}^{d \times n \times T}$ ,  $Y \in \mathbb{R}^{n \times c}$ .

1. Initialize  $\mathcal{B}^{(0)} \in \mathbb{R}^{d \times c \times T}$  using the regression results at each individual time point.

**repeat**

2. Calculate the diagonal matrix  $D$ , where the  $k$ -th diagonal element is computed as  $\frac{1}{2\sqrt{\sum_{t=1}^T \|\mathbf{b}_t^k\|_2^2}}$ .

3. Calculate  $\bar{D} = \frac{1}{2} (BB^T)^{-\frac{1}{2}}$ .

4. Update  $B_t$  by  $B_t = (X_t X_t^T + \gamma_1 D + \gamma_2 \bar{D})^{-1} X_t Y$ .

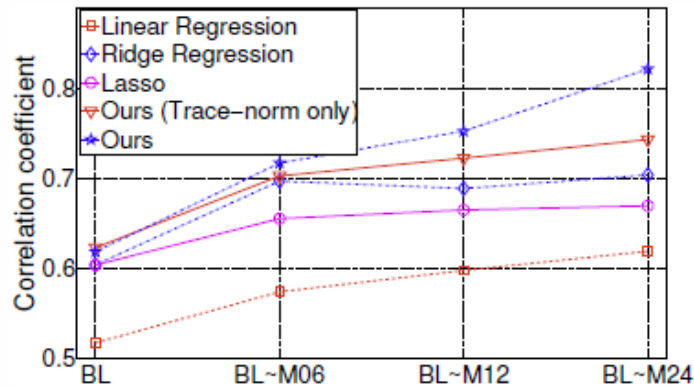
**until** *Converges*

**Result:**  $\mathcal{B} = \{B_1, B_2, \dots, B_T\} \in \mathbb{R}^{d \times c \times T}$ .

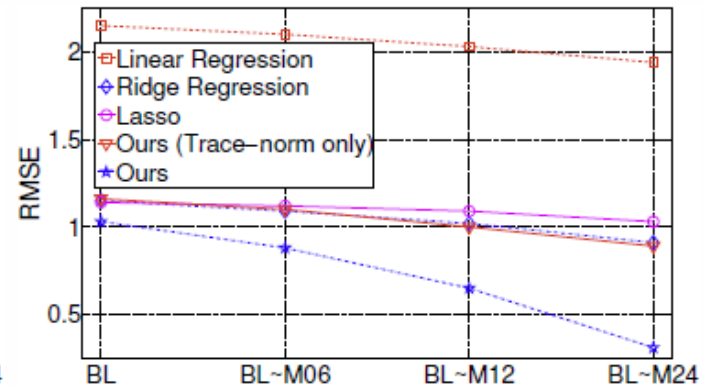
---



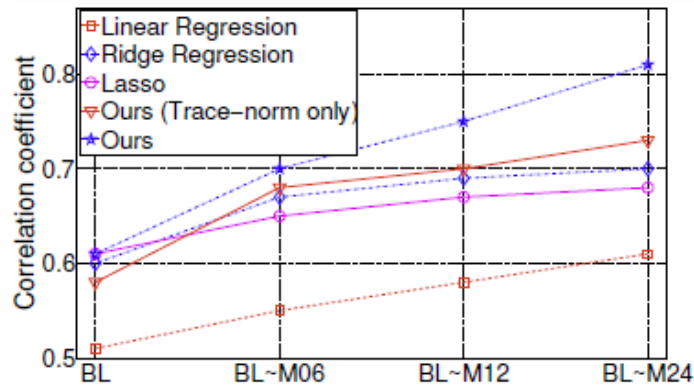
# Experimental Results



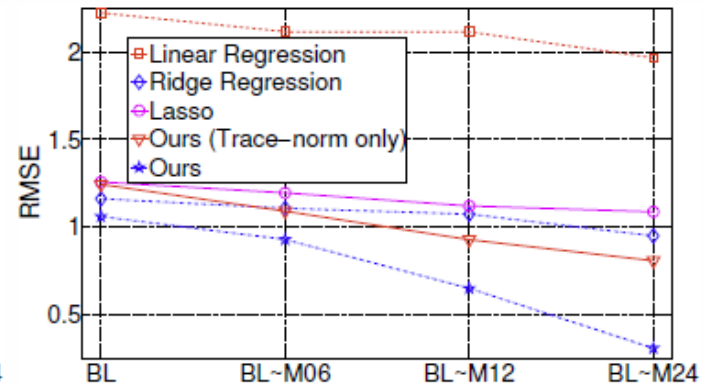
(a) VBM.



(b) VBM.

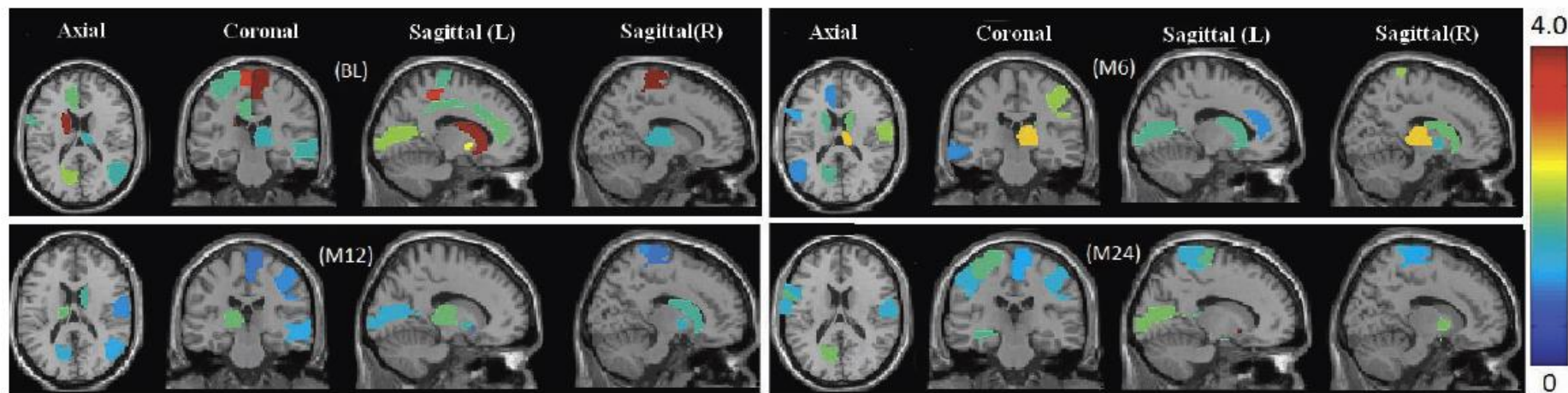
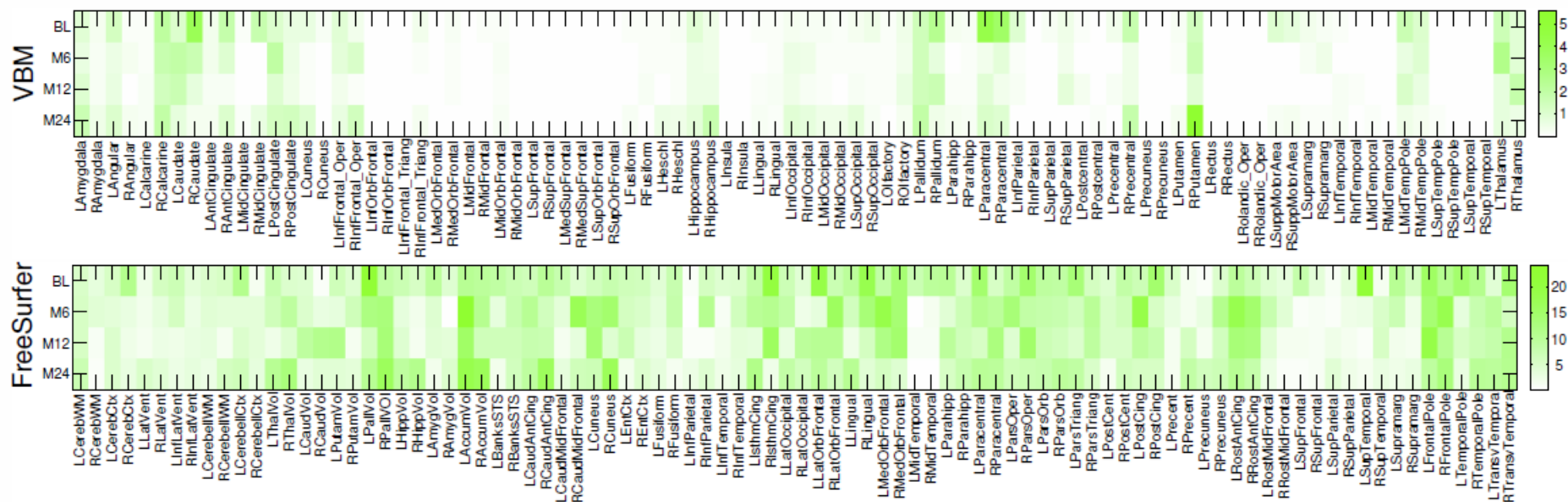


(c) FreeSurfer.



(d) FreeSurfer.

# Experimental Results (Cont.)



# Experimental Results (Cont.)

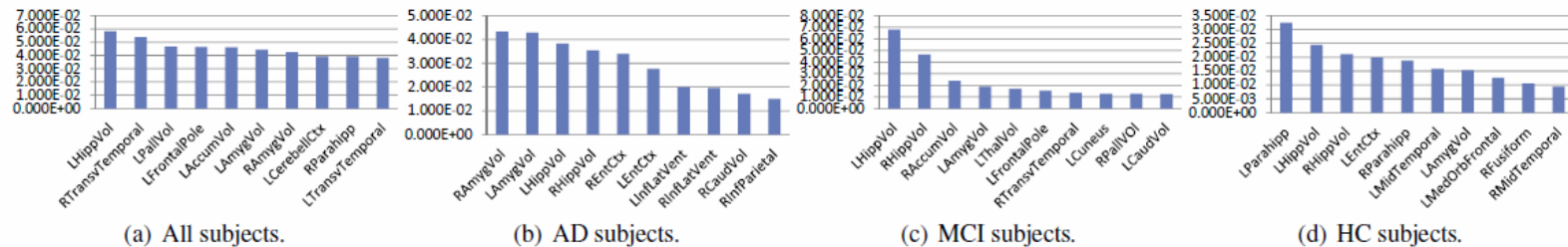


Fig. 4. Top 10 FreeSurfer markers identified for rs423958-APOE.

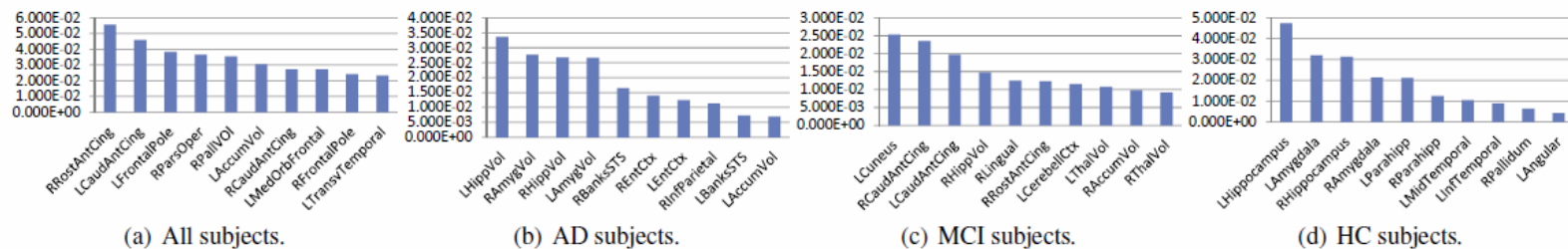


Fig. 5. Top 10 FreeSurfer markers identified for rs11136000-CLU.

# Conclusions

- Our research results have shown the effectiveness of using multi-modal data in early AD detection.
- Our algorithms have proven to be theoretically elegant and computationally efficient.

