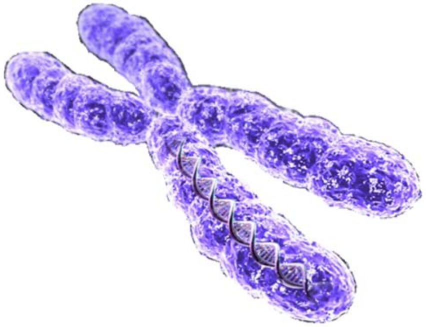
Advanced Introduction to Machine Learning

10715, Fall 2014

Structured Sparsity, with application in Computational Genomics



Eric Xing

Lecture 3, September 15, 2014

Reading:

Structured Sparsity

$$\beta^* = \arg\min L(\mathbf{X}, \mathbf{Y}; \beta) + \boldsymbol{\Omega}(\beta)$$

- Sparsity

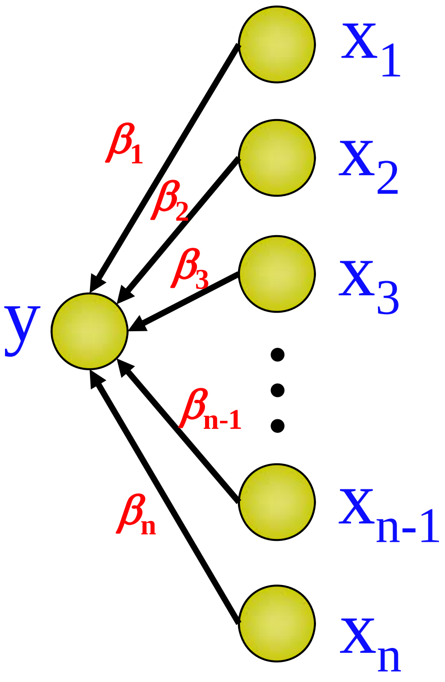
$$\Omega(\beta) = \sum_i |\beta_i|$$

- Group sparsity

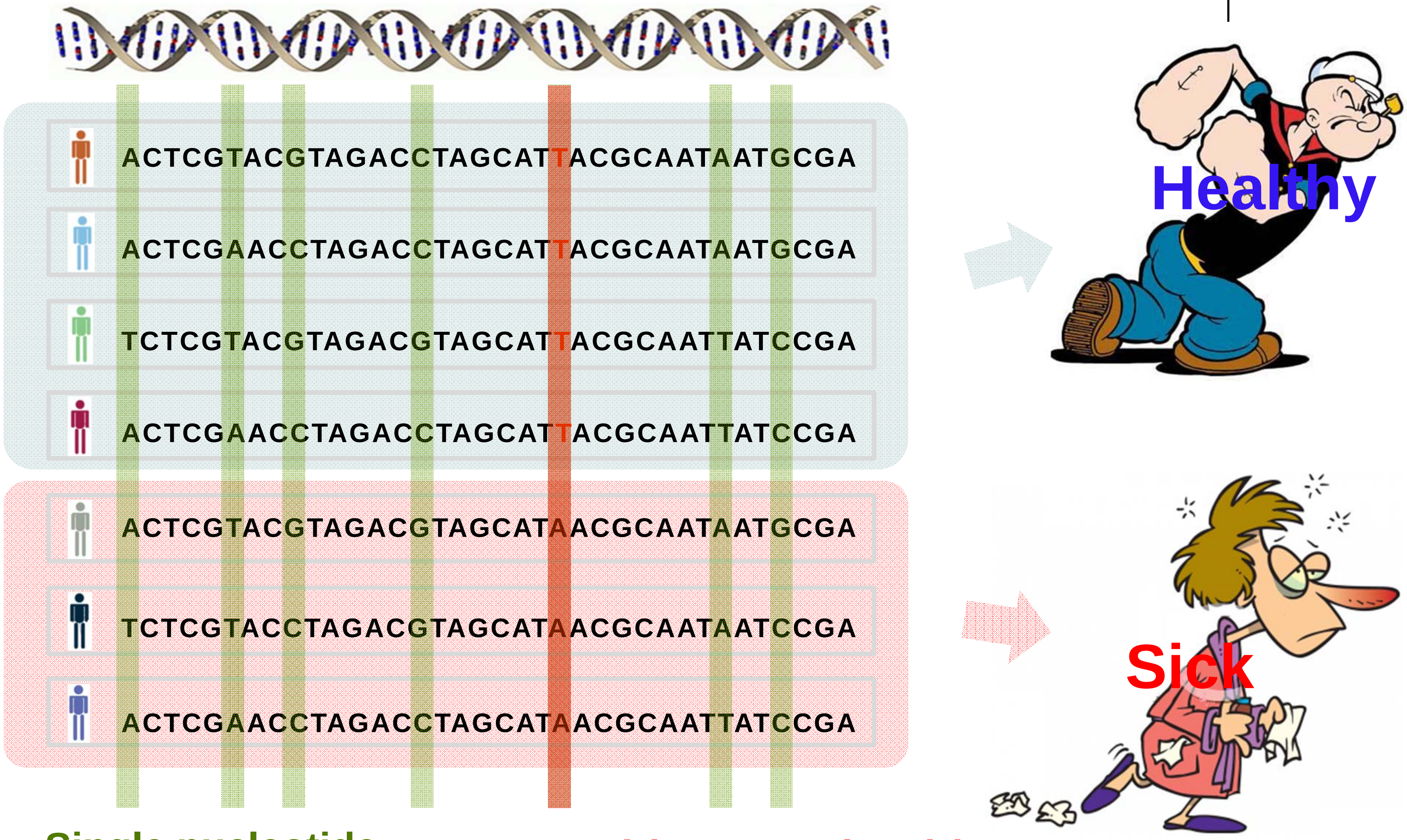
$$\Omega(\beta) = \sum_c |\beta_{G_c}|_2 = \sum_c \sqrt{\sum_{i \in G_c} \beta_i^2}$$

- Total variation sparsity

$$\Omega(\beta) = \sum_c |\beta_{G_c}|_{TV} = \sum_c \sum_{i \in G_c} |\beta_i - \beta_{i-1}|$$



Genetic Basis of Diseases



ACTCGTACGTAGACCTAGCATTACGCAATAATGCGA

ACTCGAACCTAGACCTAGCATTACGCAATAATGCGA

TCTCGTACGTAGACGTAGCATTACGCAATTATCCGA

ACTCGAACCTAGACCTAGCATTACGCAATTATCCGA

ACTCGTACGTAGACGTAGCATAACGCAATAATGCGA

TCTCGTACCTAGACGTAGCATAACGCAATAATCCGA

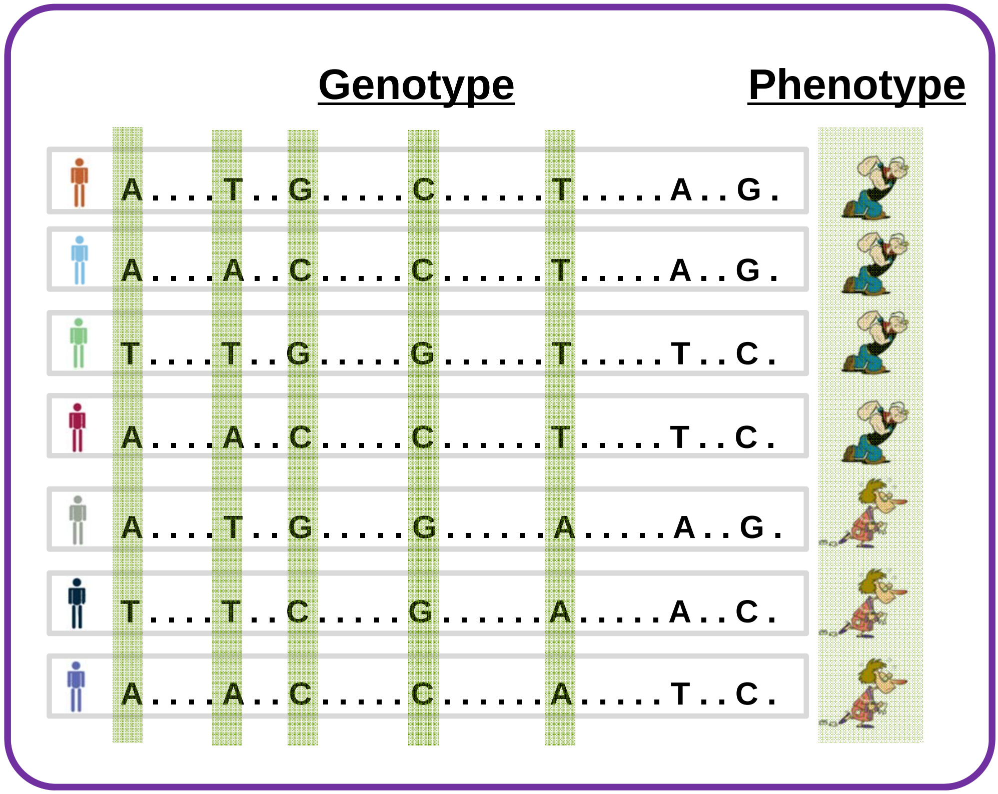
ACTCGAACCTAGACCTAGCATAACGCAATTATCCGA

Single nucleotide polymorphism (SNP)

Causal (or "associated") SNP

Genetic Association Mapping

Data



Genotype **Phenotype**

A . . . . T . . G . . . . . C . . . . . . T . . . . . A . . G .

A . . . . A . . C . . . . . C . . . . . . T . . . . . A . . G .

T . . . . T . . G . . . . . G . . . . . . T . . . . . T . . C .

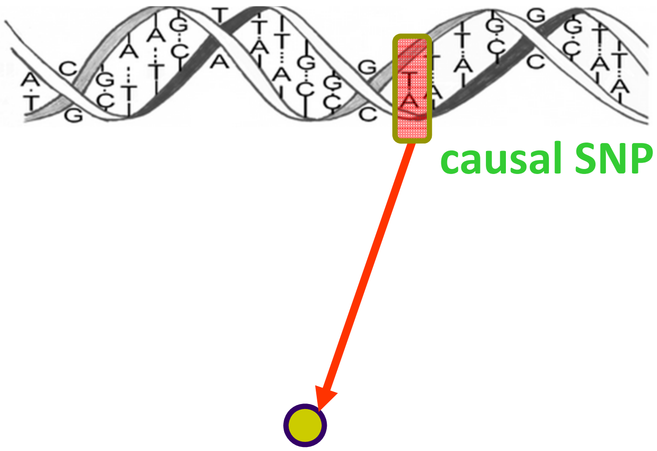
A . . . . A . . C . . . . . C . . . . . . T . . . . . T . . C .

A . . . . T . . G . . . . . G . . . . . . A . . . . . A . . G .

T . . . . T . . C . . . . . G . . . . . . A . . . . . A . . C .

A . . . . A . . C . . . . . C . . . . . . A . . . . . T . . C .

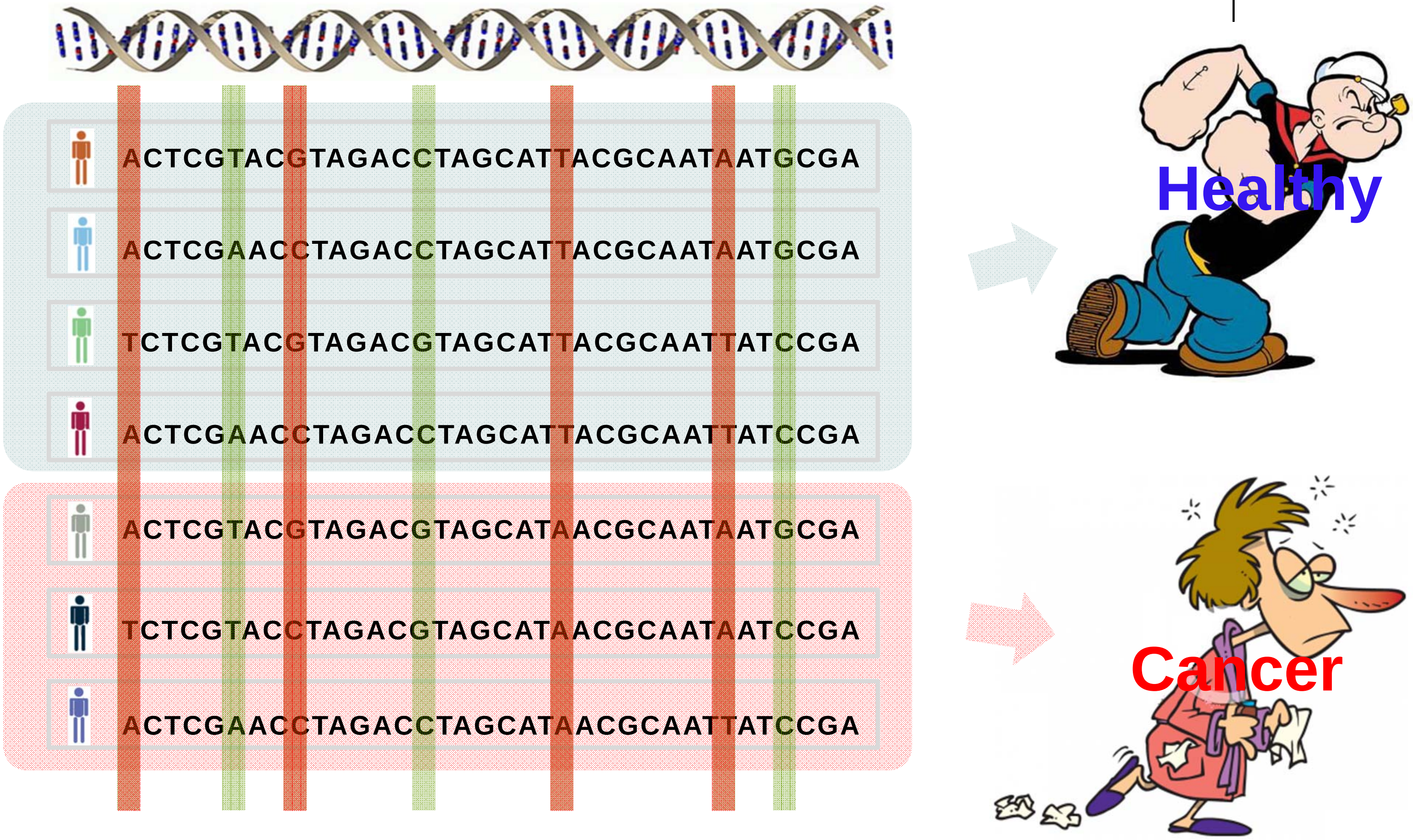
Standard Approach



a univariate **phenotype**:
e.g., disease/control

- **Cancer: Dunning et al. 2009.**
- **Diabetes: Dupuis et al. 2010.**
- **Atopic dermatitis: Esparza-Gordillo et al. 2009.**
- **Arthritis: Suzuki et al. 2008**

Genetic Basis of Complex Diseases



ACTCGTACGTAGACCTAGCATTACGCAATAATGCGA

ACTCGAACCTAGACCTAGCATTACGCAATAATGCGA

TCTCGTACGTAGACGTAGCATTACGCAATTATCCGA

ACTCGAACCTAGACCTAGCATTACGCAATTATCCGA

ACTCGTACGTAGACGTAGCATAACGCAATAATGCGA

TCTCGTACCTAGACGTAGCATAACGCAATAATCCGA

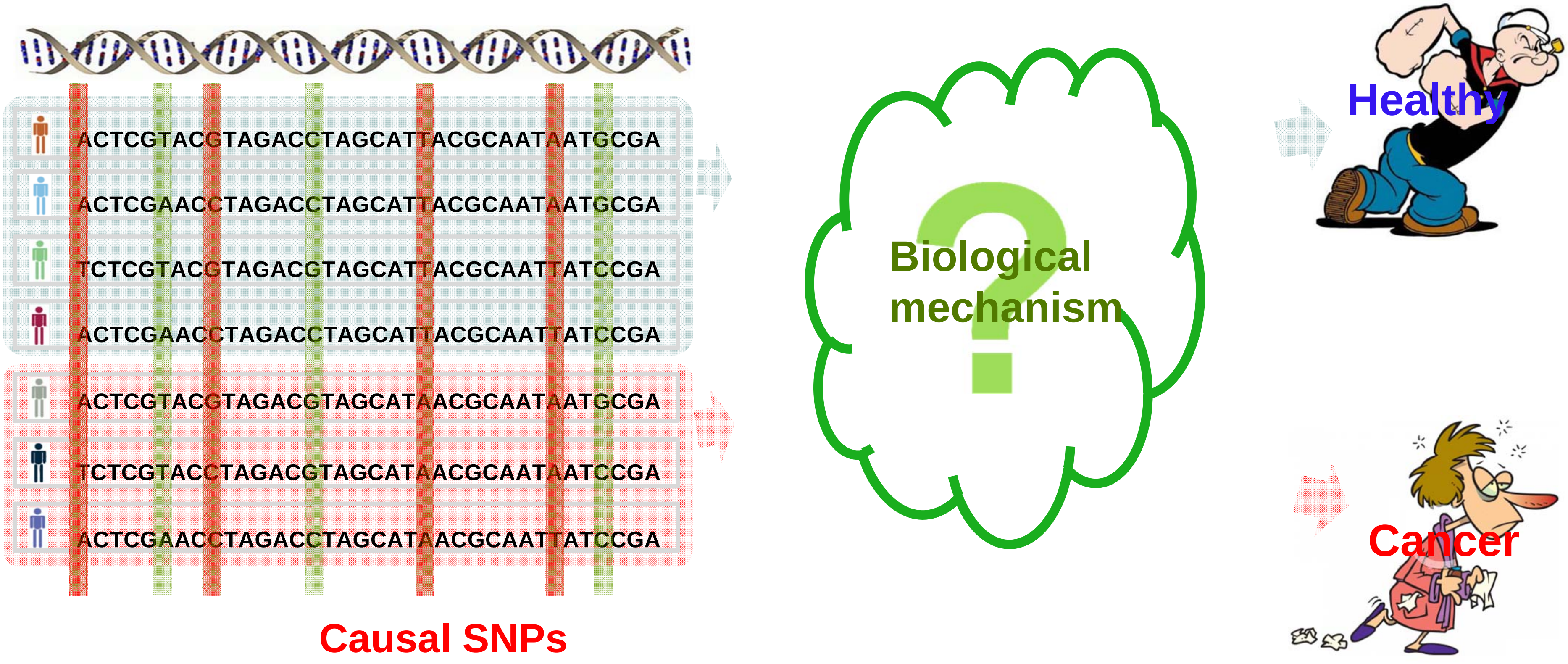
ACTCGAACCTAGACCTAGCATAACGCAATTATCCGA

Healthy

Cancer

Causal SNPs

Genetic Basis of Complex Diseases



ACTCGTACGTAGACCTAGCATTACGCAATAATGCGA

ACTCGAACCTAGACCTAGCATTACGCAATAATGCGA

TCTCGTACGTAGACGTAGCATTACGCAATTATCCGA

ACTCGAACCTAGACCTAGCATTACGCAATTATCCGA

ACTCGTACGTAGACGTAGCATAACGCAATAATGCGA

TCTCGTACCTAGACGTAGCATAACGCAATAATCCGA

ACTCGAACCTAGACCTAGCATAACGCAATTATCCGA

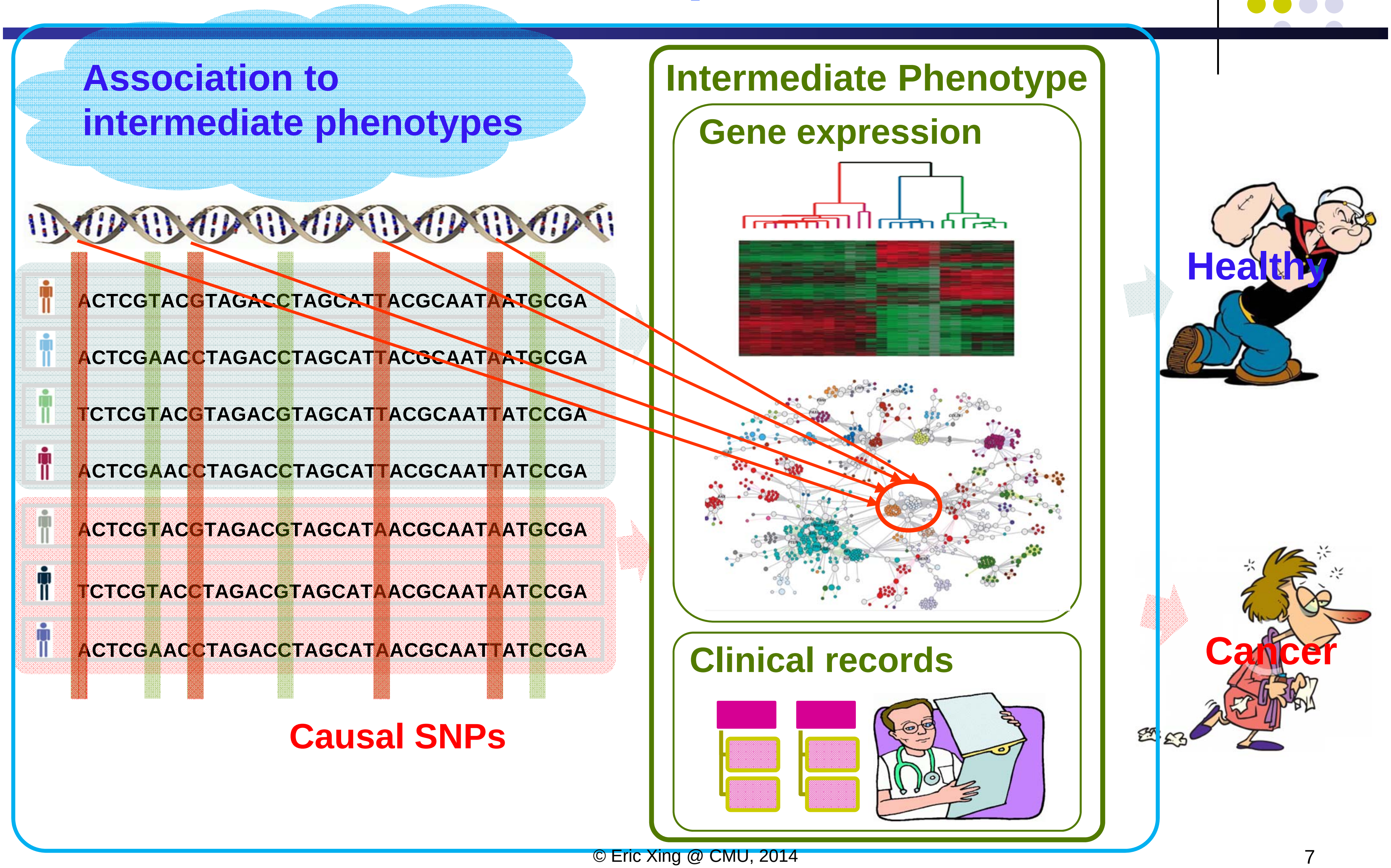
Causal SNPs

Biological mechanism

Healthy

Cancer

Genetic Basis of Complex Diseases



Association to intermediate phenotypes

ACTCGTACGTAGACCTAGCATTACGCAATAATGCGA

ACTCGAACCTAGACCTAGCATTACGCAATAATGCGA

TCTCGTACGTAGACGTAGCATTACGCAATTATCCGA

ACTCGAACCTAGACCTAGCATTACGCAATTATCCGA

ACTCGTACGTAGACGTAGCATAACGCAATAATGCGA

TCTCGTACCTAGACGTAGCATAACGCAATAATCCGA

ACTCGAACCTAGACCTAGCATAACGCAATTATCCGA

Causal SNPs

Intermediate Phenotype

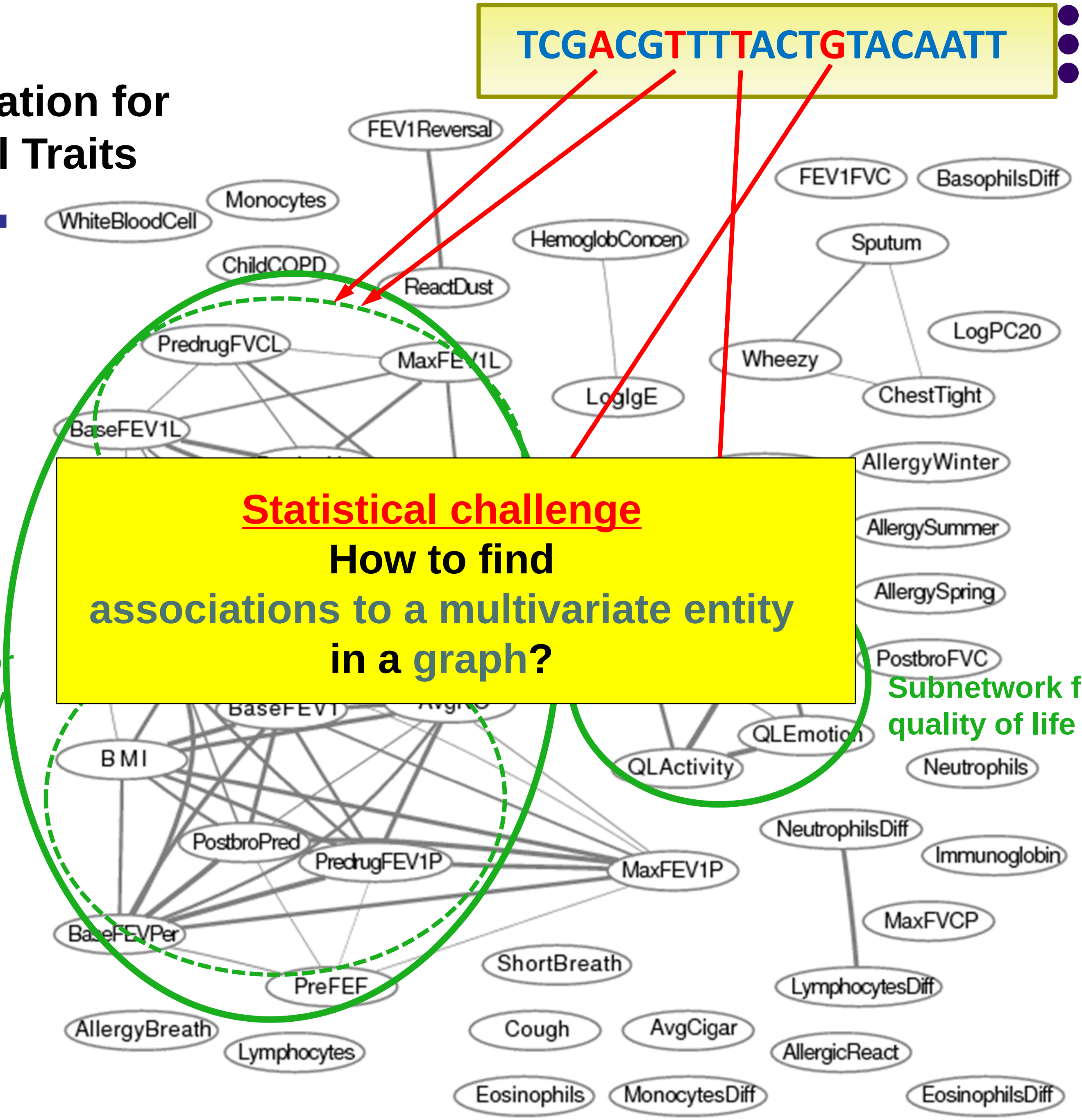
Gene expression

Clinical records

Healthy

Cancer

Genetic Association for Asthma Clinical Traits



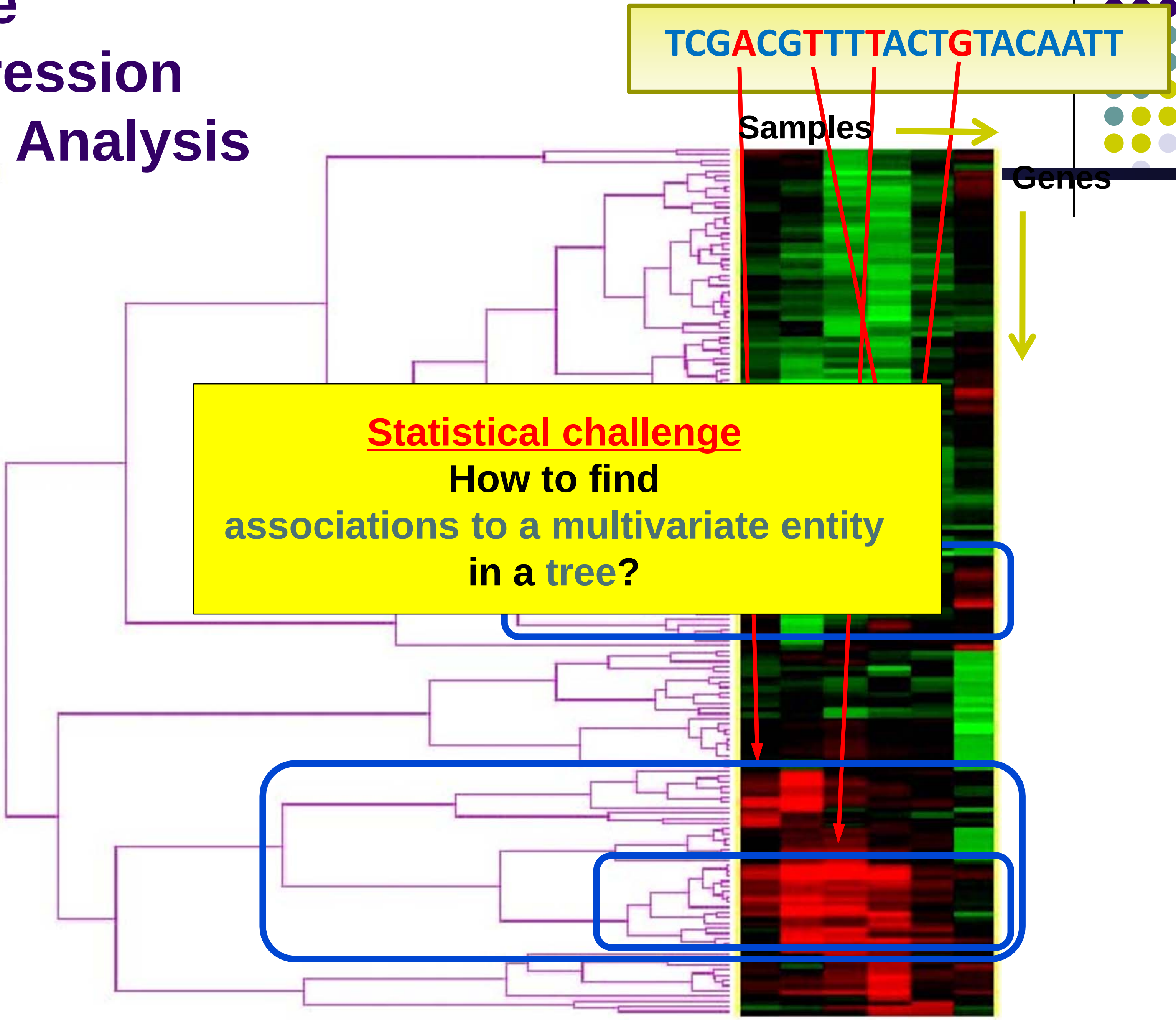
Statistical challenge

How to find associations to a multivariate entity in a graph?

Subnetworks for lung physiology

Subnetwork for quality of life

Gene Expression Trait Analysis



Statistical challenge

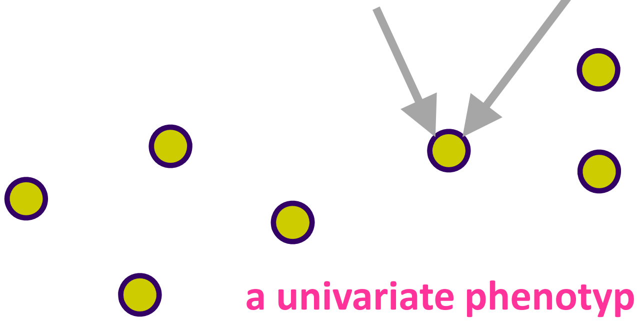
How to find associations to a multivariate entity in a tree?

Structured Association

Traditional Approach

causal SNP

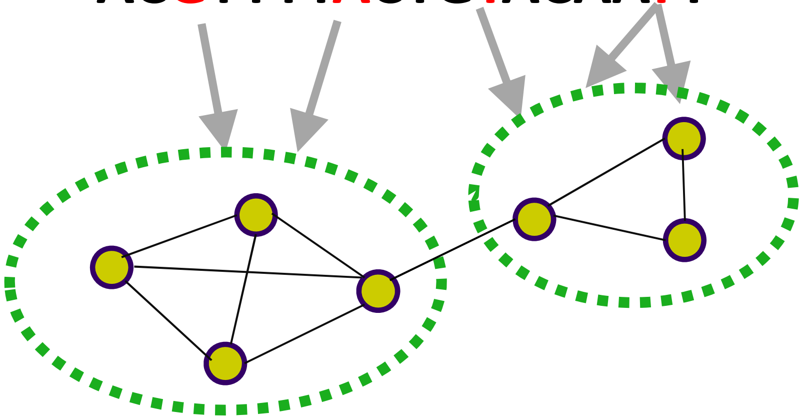
ACGTTTTACTGTACAATT



a univariate phenotype:
gene expression level

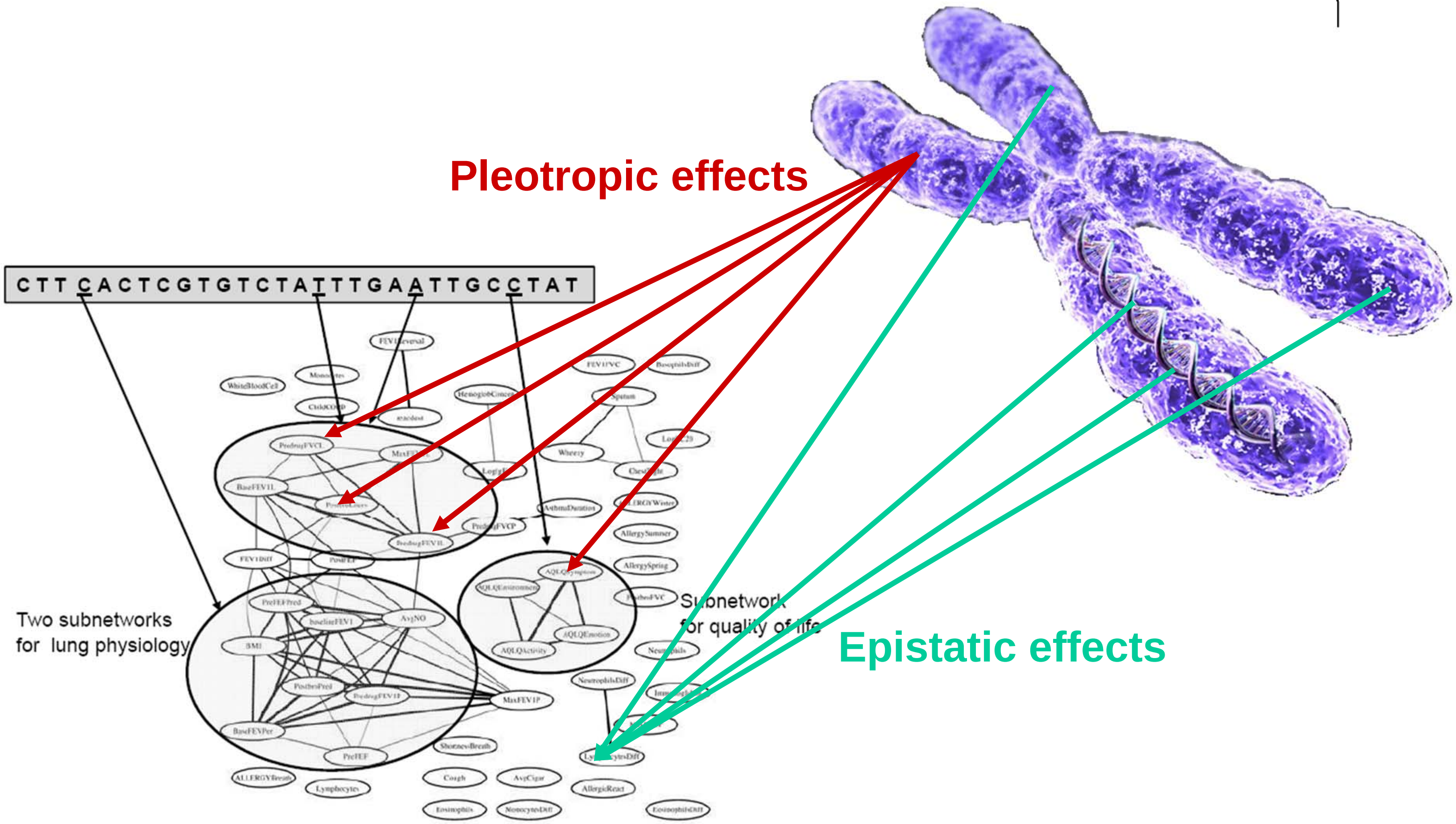
Association with Phenome

ACGTTTTACTGTACAATT



Multivariate complex syndrome (e.g., asthma)
age at onset, history of eczema
genome-wide expression profile

Sparse Associations



Structured Sparse Association : a New Paradigm

Standard Approach

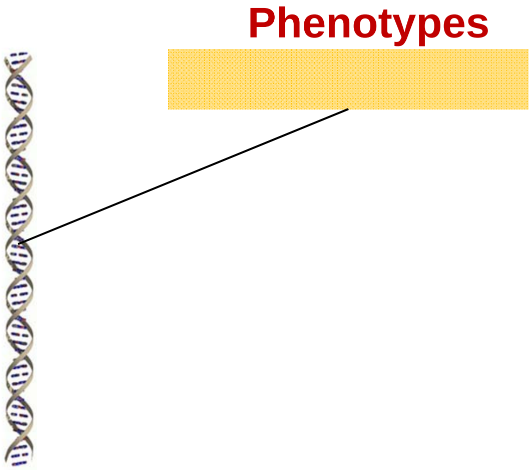
Consider
one phenotype at a
time

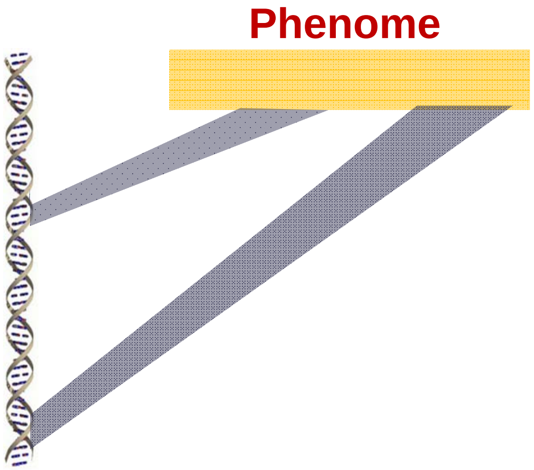
vs

.

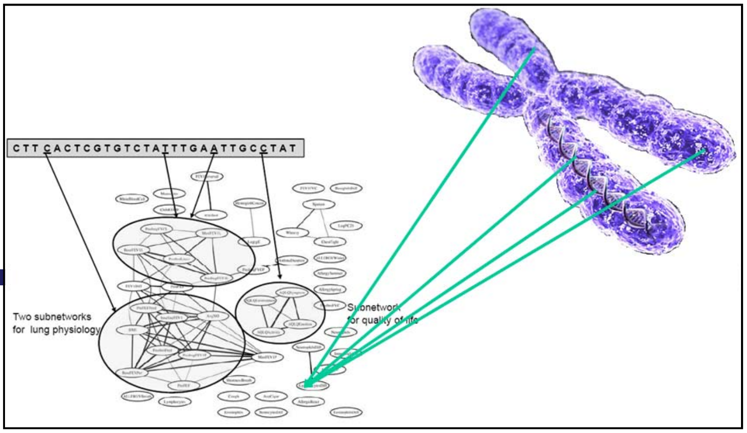
New Approach

Consider
multiple correlated
phenotypes (phenome)
jointly





Sparse Learning


CTT CACTCGTGTCTATTTGAATTGCCTAT
Two subnetworks for lung physiology
Subnetwork for quality of life

- Linear Model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \mathbf{y} \in \mathbb{R}^{N \times 1}, \quad \mathbf{X} \in \mathbb{R}^{N \times J}, \quad \boldsymbol{\epsilon} \sim N(0, \sigma^2 I_{N \times N})$$

$$\boldsymbol{\beta} = (\beta_1, \ldots, \beta_j, \ldots, \beta_J)^T \in \mathbb{R}^J$$

- Lasso (Sparse Linear Regression)

[R.Tibshirani 96]

$$\arg\min_{\boldsymbol{\beta} \in \mathbb{R}^J} f(\boldsymbol{\beta}) \equiv \frac{1}{2}\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \Omega(\boldsymbol{\beta}) \quad \Omega(\boldsymbol{\beta}) = \lambda\|\boldsymbol{\beta}\|_1$$

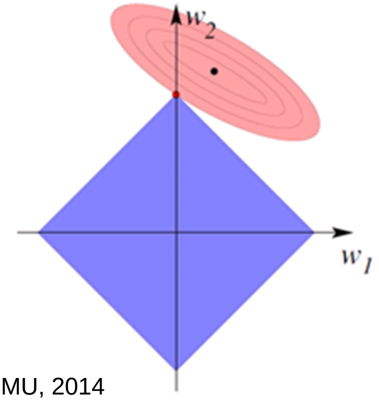
$$\|\boldsymbol{\beta}\|_1 = \sum_{j=1}^{J} |\beta_j|$$

- Why sparse solution?

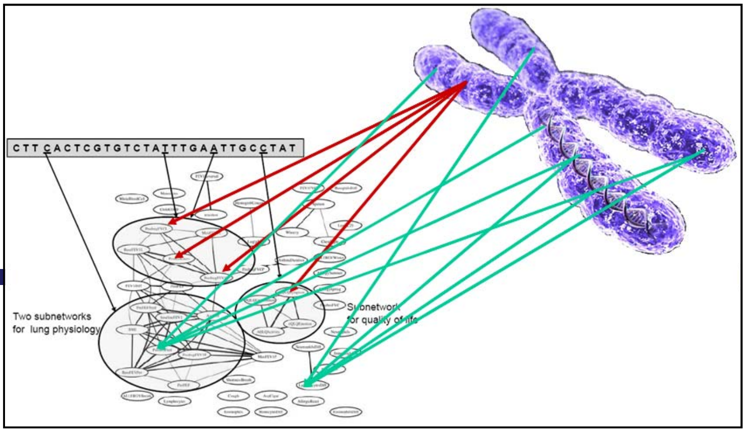
penalizing $\lambda\|\boldsymbol{\beta}\|_1$

$\Updownarrow$

constraining $\|\boldsymbol{\beta}\|_1 \leq \gamma$



Multi-Task Extension


CTT CACTCGTGTCTATTTGAATTGCCTAT
Two subnetworks for lung physiology
Subnetwork for quality of life

- Multi-Task Linear Model:

Input: $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_J) \in \mathbb{R}^{N \times J}$

Output: $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_K) \in \mathbb{R}^{N \times K}$

$$\boxed{\mathbf{y}_k = \mathbf{X}\boldsymbol{\beta}_k + \epsilon_k, \quad \forall k = 1, \ldots, K}$$

Coefficients for *k-th* task: $\boldsymbol{\beta}_k = (\beta_{1k}, \ldots, \beta_{Jk})^T \in \mathbb{R}^J$

Coefficient Matrix: $\mathbf{B} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K) \in \mathbb{R}^{J \times K}$

$$\mathbf{B} = \begin{pmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1K} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{J1} & \beta_{J2} & \cdots & \beta_{JK} \end{pmatrix}$$

Coefficients for a variable (2nd)

Coefficients for a task (2nd)

Outline

- Background: Sparse multivariate regression for disease association studies

- Structured association – a new paradigm
  - Association to a **graph**-structured phenome
    - Graph-guided fused lasso (Kim & Xing, PLoS Genetics, 2009)

  - Association to a **tree**-structured phenome
    - Tree-guided group lasso (Kim & Xing, ICML 2010)

Multivariate Regression for Single-Trait Association Analysis

Trait		Genotype		Association Strength
2.1	=	T G A A C C A T G A A G T A	x	?
y	=	X	x	β

Multivariate Regression for Single-Trait Association Analysis

Trait

2.1

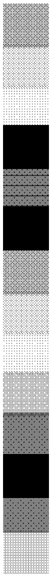
=

Genotype

TGAACCATGAAGTA

x

Association Strength

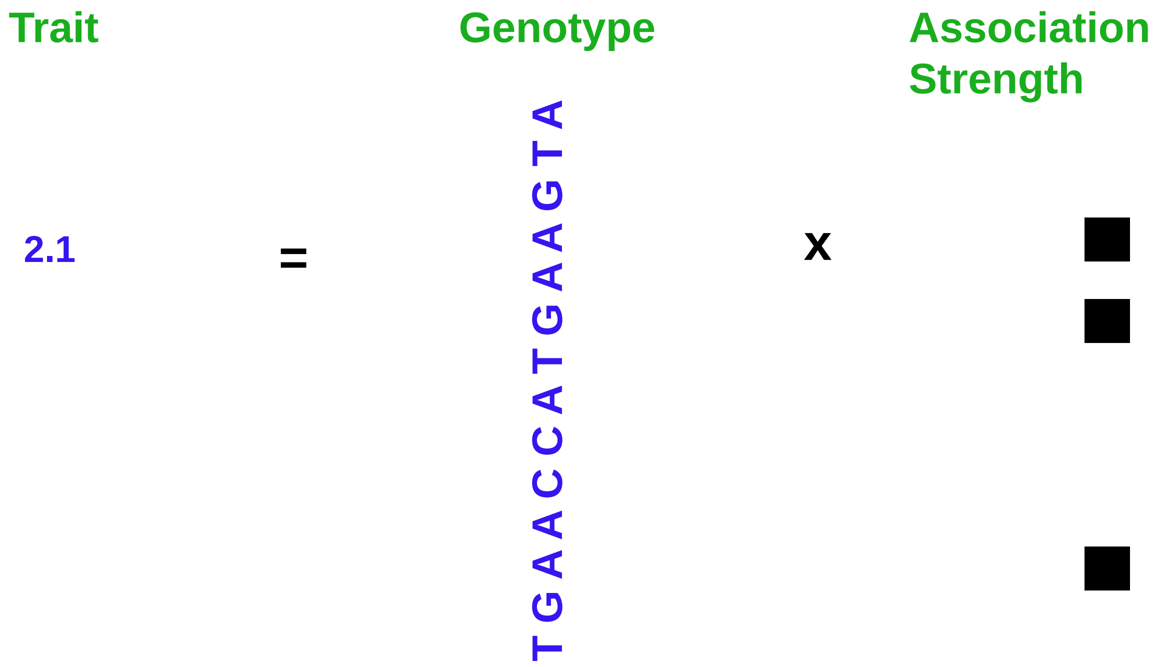


$$\beta^* = \arg\min_{\beta} (\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta)$$

Many non-zero associations: Which SNPs are truly significant?

Lasso for Reducing False Positives

(Tibshirani, 1996)

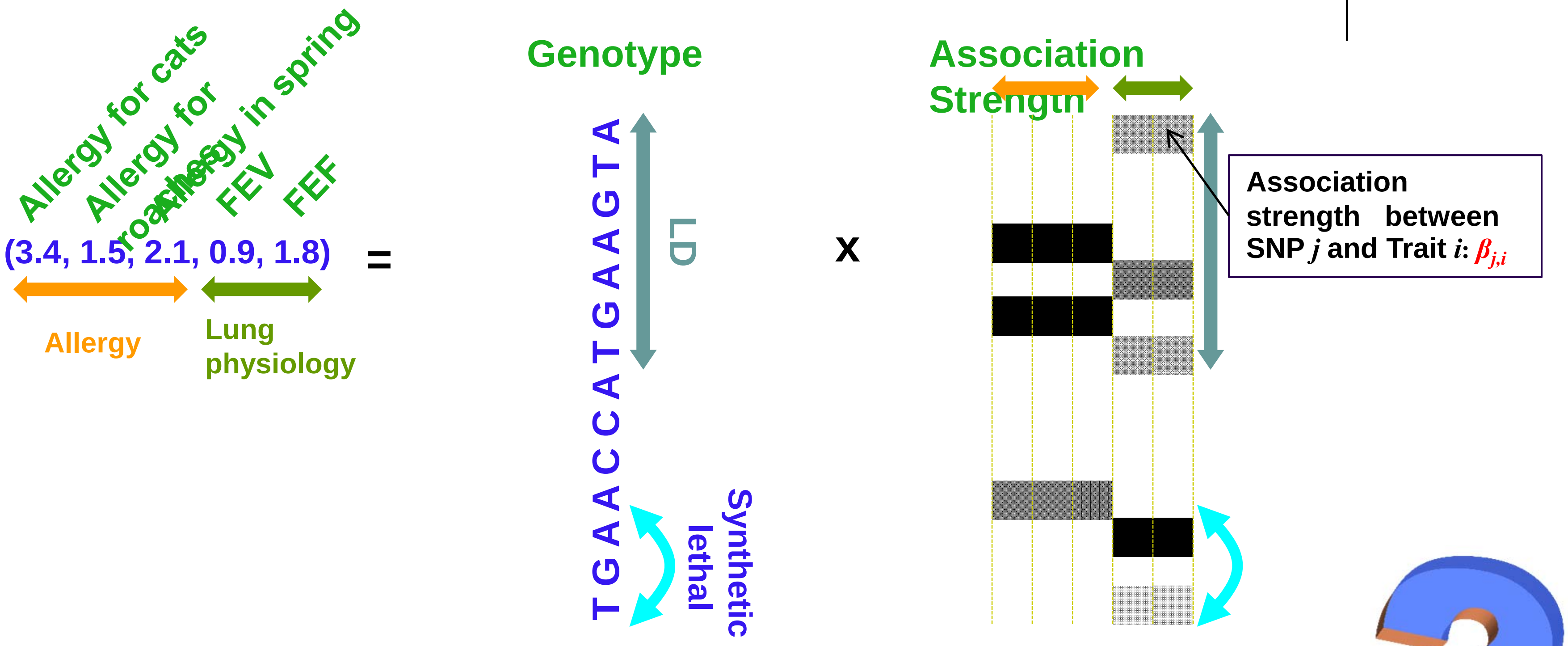


Lasso Penalty for sparsity

$$\beta^* = \arg\min_{\beta} (\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta) + \lambda \sum_{j=1}^{J} |\boldsymbol{\beta}_j|$$

Many zero associations (sparse results), but what if there are multiple related traits?

Multivariate Regression for Multiple-Trait Association Analysis



Allergy for cats

Allergy for roaches

Allergy in spring

FEV

FEF

(3.4, 1.5, 2.1, 0.9, 1.8) =

Allergy

Lung physiology

Genotype

T G A A C C A T G A A G T A

LD

Synthetic lethal

x

Association Strength

Association strength between SNP *j* and Trait *i*: $\beta_{j,i}$

$$\beta^* = \arg\min_{\beta} \sum_i (\mathbf{y}_i - \mathbf{X}_i\beta_i)^T (\mathbf{y}_i - \mathbf{X}_i\beta_i) + \lambda \sum_{i,j} |\beta_{j,i}|$$

How to combine information across multiple traits to increase the power?

Multivariate Regression for Multiple-Trait Association Analysis

Trait

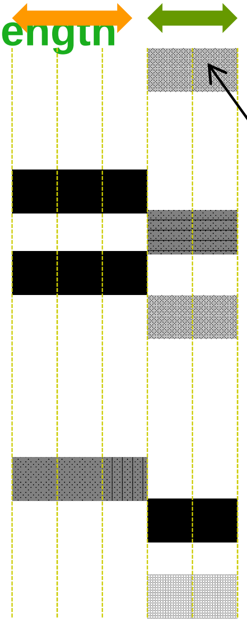
Genotype

Association Strength

(3.4, 1.5, 2.1, 0.9, 1.8) = TGAACCATGAAGTA x

Allergy

Lung physiology



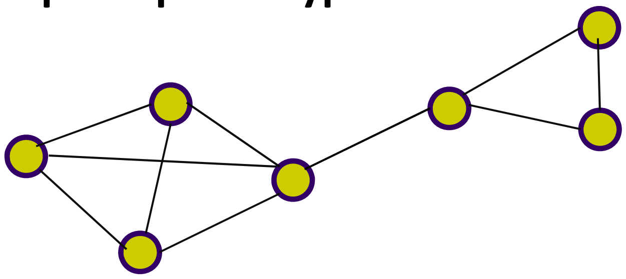
Association strength between SNP *j* and Trait *i*: $\beta_{j,i}$

$$\beta^* = \arg\min_{\beta} \sum_i (\mathbf{y}_i - \mathbf{X}_i\beta_i)^T(\mathbf{y}_i - \mathbf{X}_i\beta_i) + \lambda\sum_{i,j}|\beta_{j,i}|$$

+ We introduce **graph-guided fusion penalty**

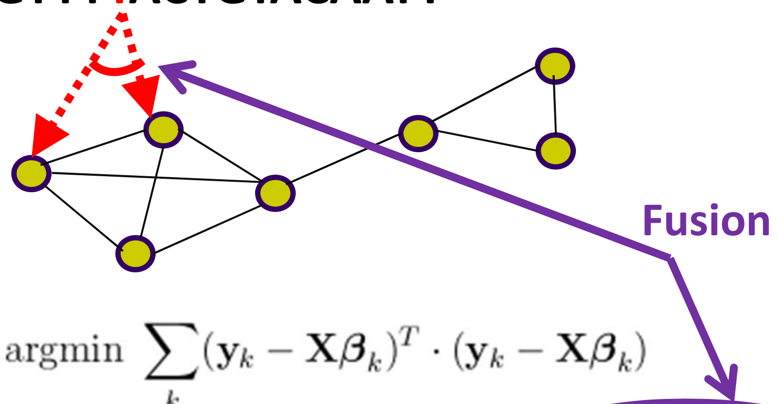
Multiple-trait Association: Graph-Constrained Fused Lasso

Step 1: Thresholded correlation graph of phenotypes



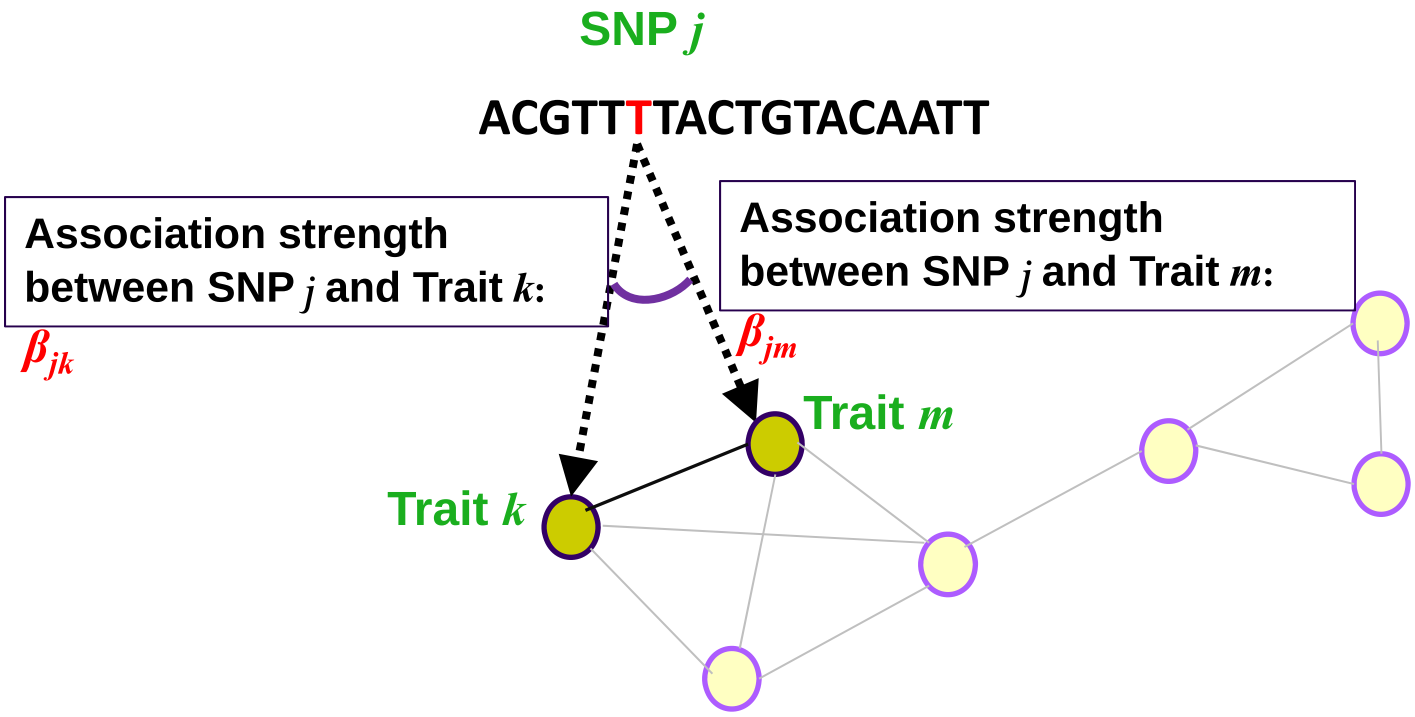
Step 2: Graph-constrained fused lasso

ACGTTTTACTGTACAATT



$$\hat{\mathbf{B}}^{GC} = \operatorname{argmin} \sum_k (\mathbf{y}_k - \mathbf{X}\boldsymbol{\beta}_k)^T \cdot (\mathbf{y}_k - \mathbf{X}\boldsymbol{\beta}_k) + \lambda \underbrace{\sum_k \sum_j |\beta_{jk}|}_{\text{Lasso Penalty}} + \gamma \underbrace{\sum_{(m,l)\in E} \sum_j |\beta_{jm} - \operatorname{sign}(r_{ml})\beta_{jl}|}_{\text{Graph-constrained fusion penalty}}$$

Fusion Penalty



SNP j

ACGTTTTACTGTACAATT

Association strength between SNP j and Trait k: β_{jk}

Association strength between SNP j and Trait m: β_{jm}

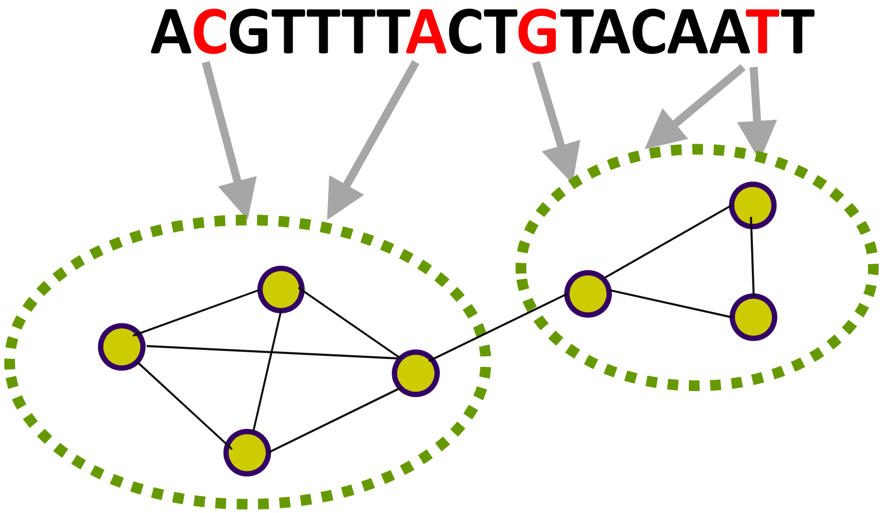
Trait m

Trait k

- Fusion Penalty: $| \beta_{jk} - \beta_{jm} |$
- For two correlated traits (connected in the network), the association strengths may have similar values.

Graph-Constrained Fused Lasso

Overall effect

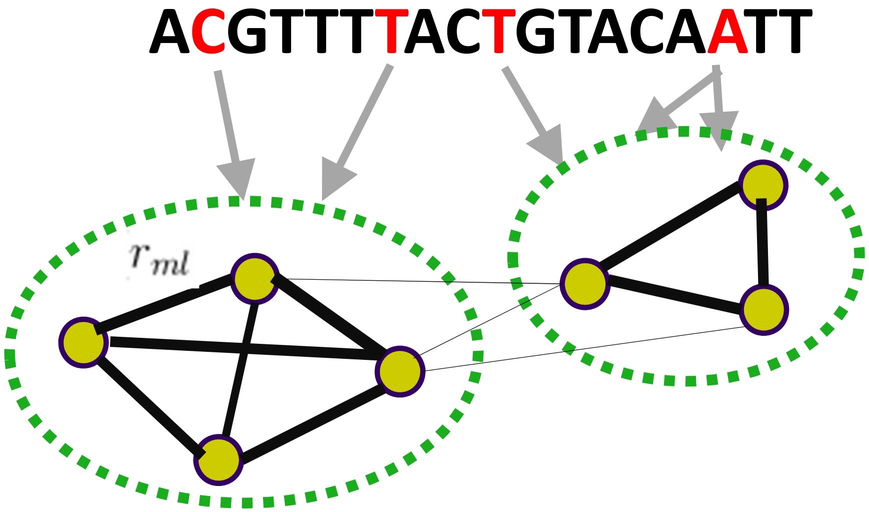


ACGTTTTACTGTACAATT

- Fusion effect propagates to the entire network
- Association between SNPs and subnetworks of traits

Multiple-trait Association: Graph-Weighted Fused Lasso

Overall effect



- Subnetwork structure is embedded as a densely connected nodes with large edge weights
- Edges with small weights are effectively ignored

© Eric Xing @ CMU, 2014

Estimating Parameters

- Quadratic programming formulation
  - Graph-constrained fused lasso

$$\hat{\mathbf{B}}^{\mathrm{GC}} = \operatorname{argmin} \sum_k (\mathbf{y}_k - \mathbf{X}\boldsymbol{\beta}_k)^T \cdot (\mathbf{y}_k - \mathbf{X}\boldsymbol{\beta}_k)$$

$$\text{s. t.} \quad \sum_k \sum_j |\beta_{jk}| \le s_1 \text{ and } \sum_{(m,l)\in E} \sum_j |\beta_{jm} - \mathrm{sign}(r_{ml})\beta_{jl}| \le s_2$$

  - Graph-weighted fused lasso

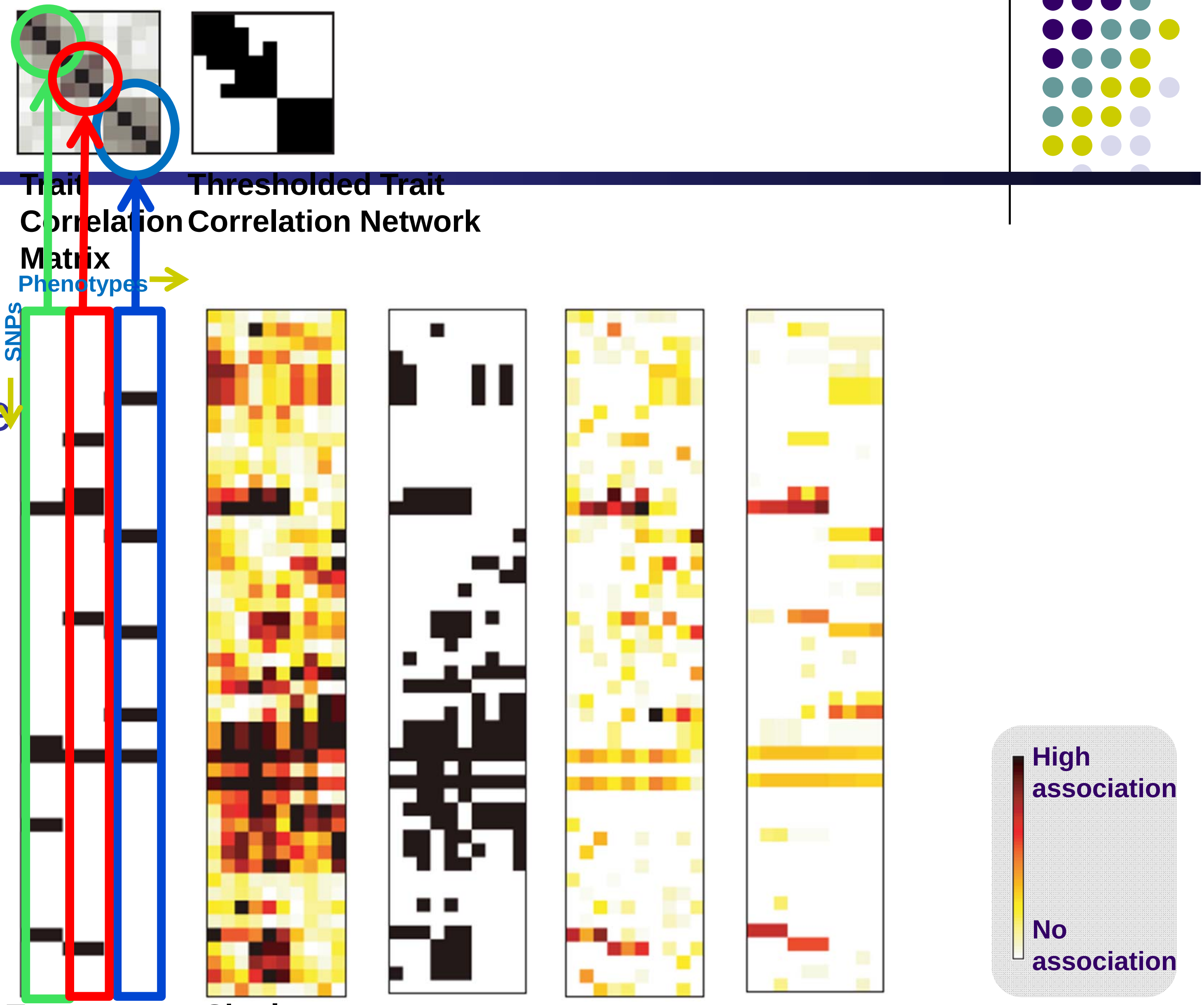
$$\hat{\mathbf{B}}^{\mathrm{GW}} = \operatorname{argmin} \sum_k (\mathbf{y}_k - \mathbf{X}\boldsymbol{\beta}_k)^T \cdot (\mathbf{y}_k - \mathbf{X}\boldsymbol{\beta}_k)$$

$$\text{s. t.} \quad \sum_k \sum_j |\beta_{jk}| \le s_1 \text{ and } \sum_{(m,l)\in E} f(r_{ml}) \sum_j |\beta_{jm} - \mathrm{sign}(r_{ml})\beta_{jl}| \le s_2$$

- Many publicly available software packages for solving convex optimization problems can be used

Simulation Results

- 50 SNPs taken from HapMap chromosome 7, CEU population
- 10 traits



Trait Correlation Matrix

Thresholded Trait Correlation Network

True Regression Coefficients

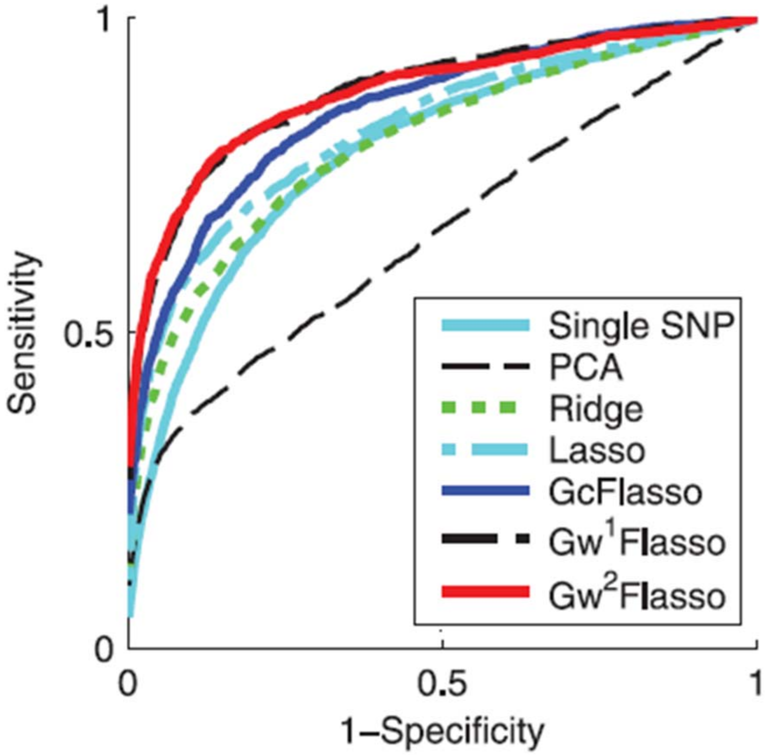
Single SNP-Single Trait Test

Significant at $\alpha = 0.01$

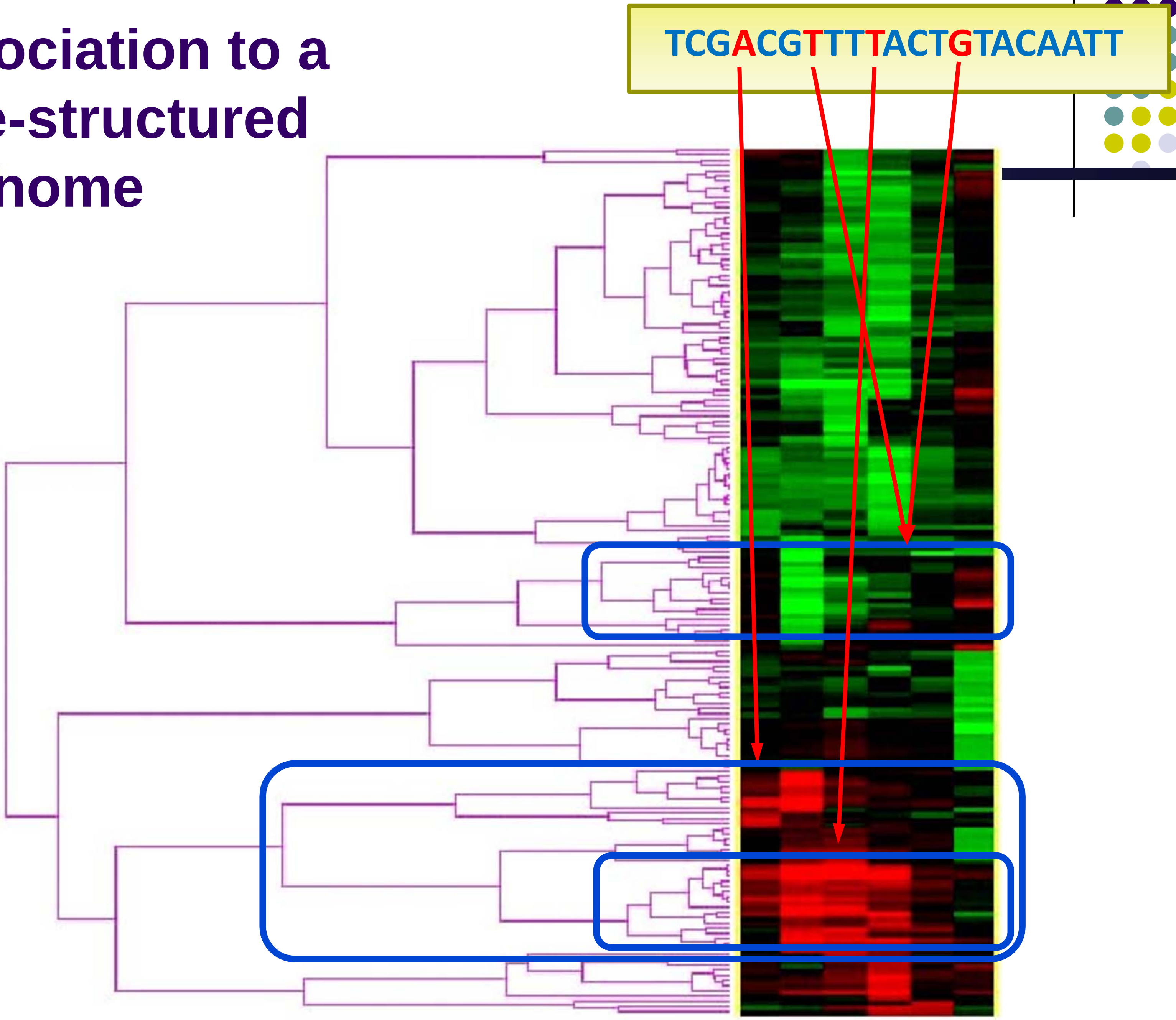
Lasso

Graph-guided Fused Lasso

Simulation Results

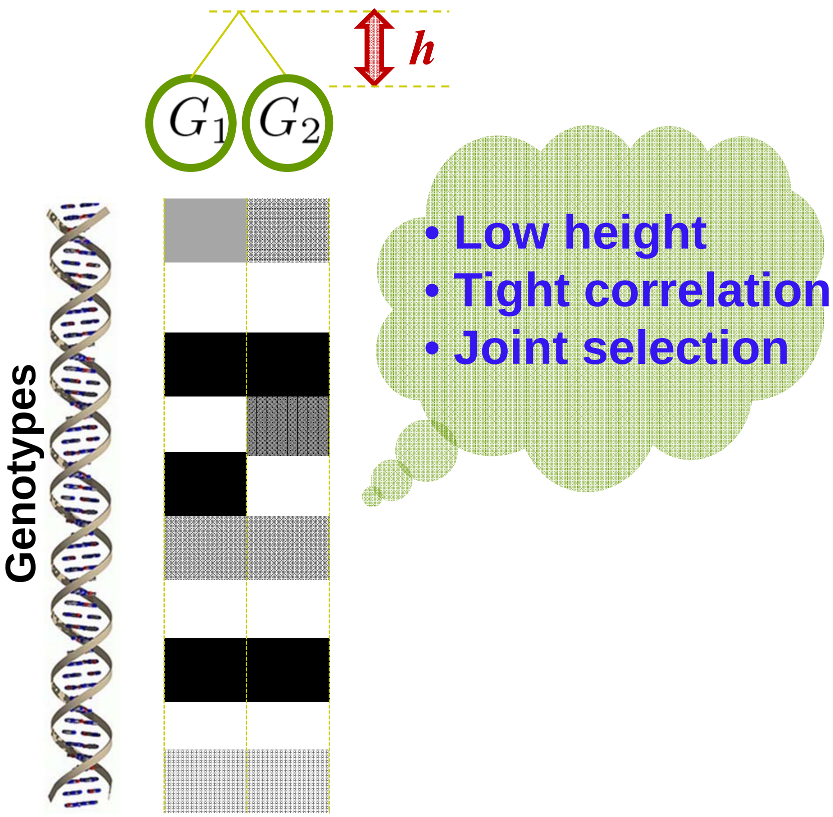


Association to a Tree-structured Phenome

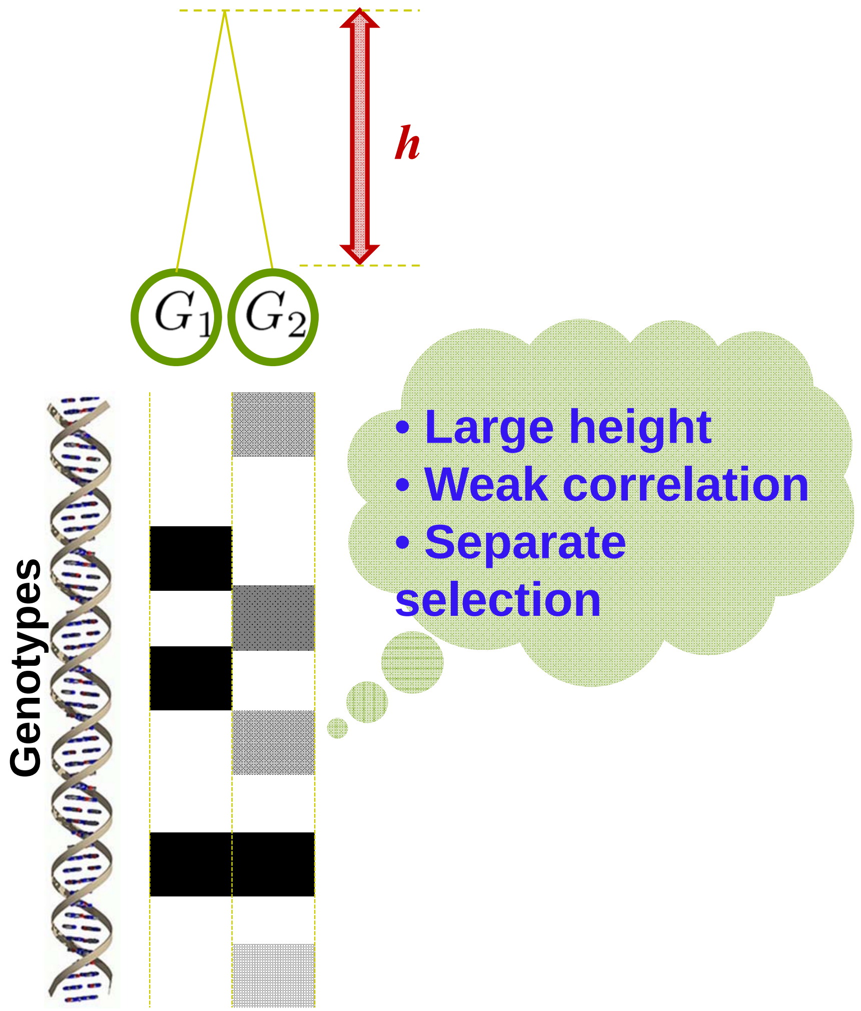


Tree-Guided Group Lasso

- In a simple case of two genes



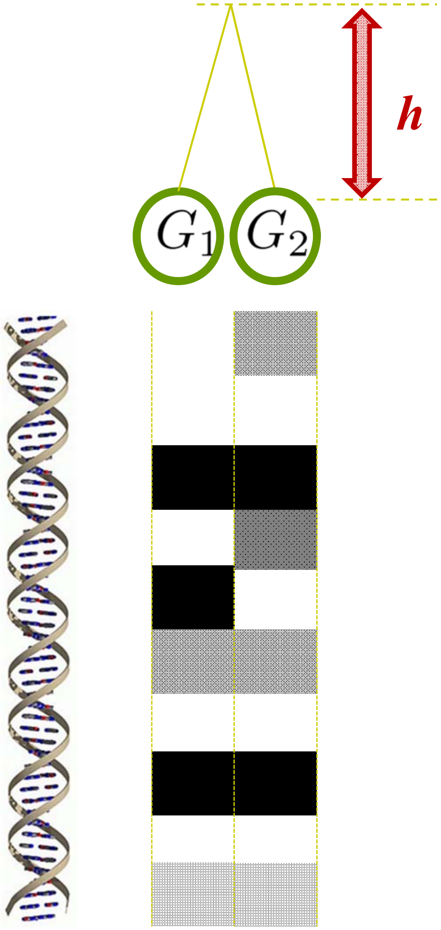
- Low height
- Tight correlation
- Joint selection

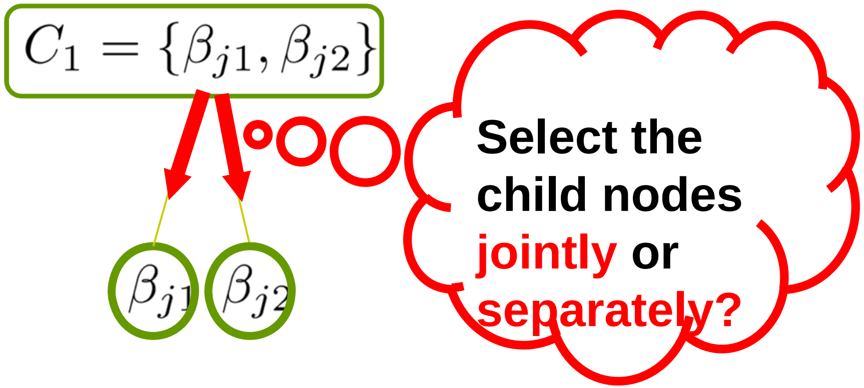


- Large height
- Weak correlation
- Separate selection

Tree-Guided Group Lasso

- In a simple case of two genes





Select the child nodes **jointly** or **separately?**

Tree-guided group lasso

$$\operatorname{argmin} \; (y - X\beta)' \cdot (y - X\beta) + \lambda \sum_j \left[h\left(|\beta_{j1}| + |\beta_{j2}|\right) + (1-h)\left(\sqrt{\beta_{j1}^2 + \beta_{j2}^2}\right) \right]$$

***L_1* penalty**

- Lasso penalty
- **Separate** selection

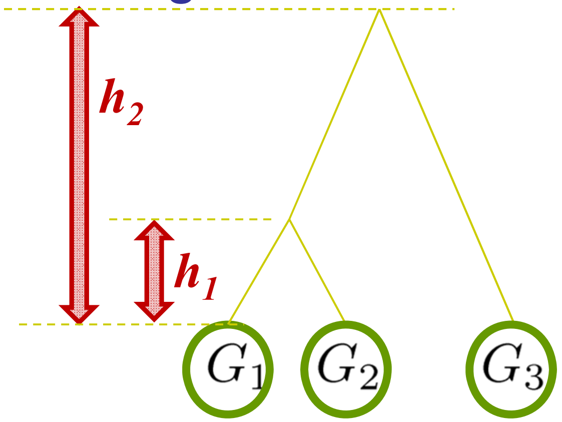
***L_2* penalty**

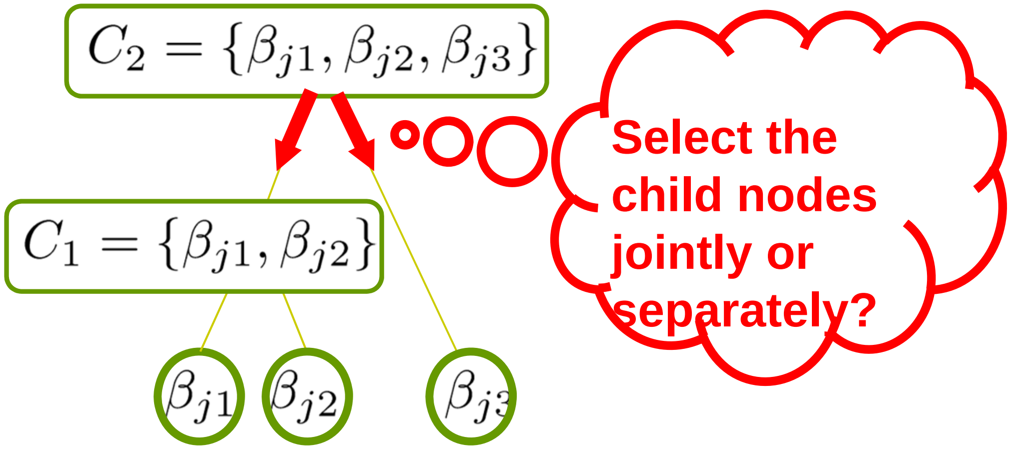
- Group lasso
- **Joint** selection

Elastic net

Tree-Guided Group Lasso

- For a general tree





$C_2 = \{\beta_{j1}, \beta_{j2}, \beta_{j3}\}$

$C_1 = \{\beta_{j1}, \beta_{j2}\}$

Select the child nodes jointly or separately?

Tree-guided group lasso

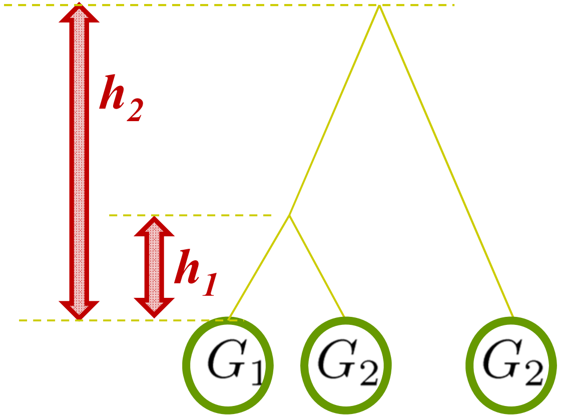
$$\operatorname{argmin}\ (y - X\beta)' \cdot (y - X\beta)$$
$$+\lambda \sum_j \Big[(1 - h_2)\Big(\sqrt{\beta_{j1}^2 + \beta_{j2}^2 + \beta_{j3}^2}\Big) + h_2\big(|C_1| + |\beta_{j3}|\big)\Big]$$

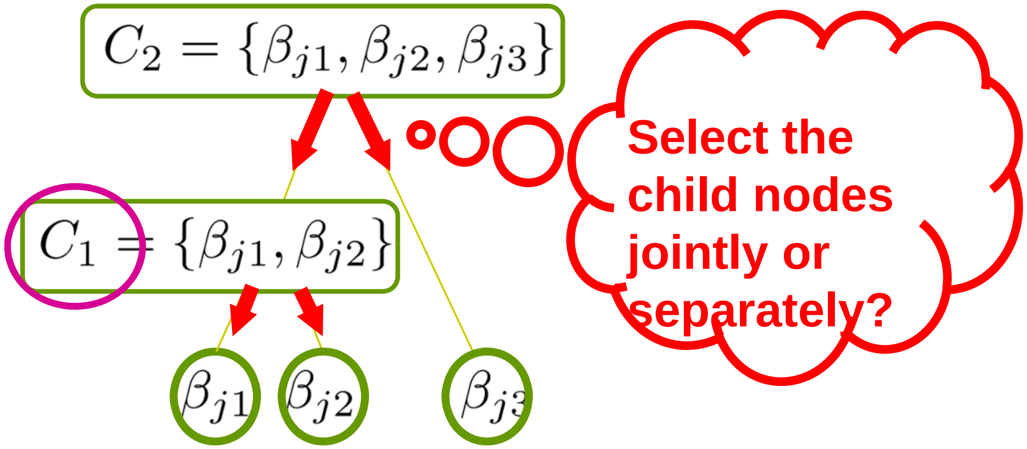
Joint selection

Separate selection

Tree-Guided Group Lasso

- For a general tree





$C_2 = \{\beta_{j1}, \beta_{j2}, \beta_{j3}\}$
$C_1 = \{\beta_{j1}, \beta_{j2}\}$
β_{j1}
β_{j2}
β_{j3}

Select the child nodes jointly or separately?

Tree-guided group lasso

$$\operatorname{argmin}\ (y - X\beta)' \cdot (y - X\beta)$$

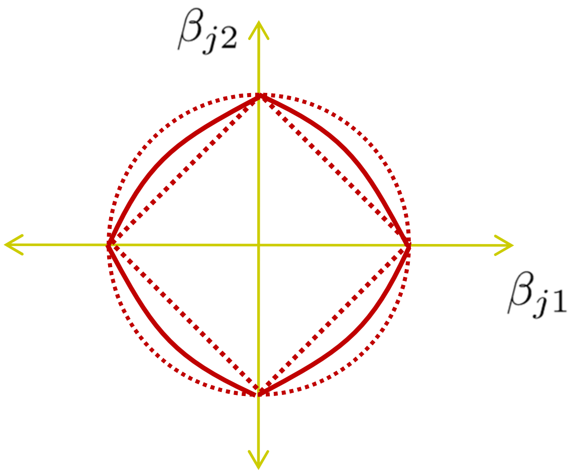
$$+\lambda \sum_j \left[(1-h_2)\left(\sqrt{\beta_{j1}^2 + \beta_{j2}^2 + \beta_{j3}^2}\right) + h_2\left(|C_1| + |\beta_{j3}|\right)\right]$$

$$|C_1| = (1-h_1)\underbrace{\left(\sqrt{\beta_{j1}^2 + \beta_{j2}^2}\right)}_{\text{Joint selection}} + h_1\underbrace{\left(|\beta_{j1}| + |\beta_{j2}|\right)}_{\text{Separate selection}}$$

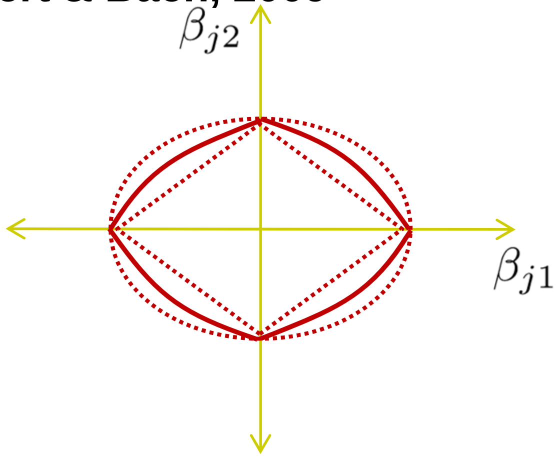
Balanced Shrinkage

Proposition 1 *For each of the k-th output (gene), the sum of the weights w_v for all nodes $v \in V$ in T whose group G_v contains the k-th output (gene) as a member equals one. In other words, the following holds:*

$$\sum_{v:k\in G_v} w_v = \prod_{m\in Ancestors(v_k)} h_m + \sum_{l\in Ancestors(v_k)} (1-h_l) \prod_{m\in Ancestors(v_l)} h_m = 1.$$



Previously, in Jenatton, Audibert & Bach, 2009



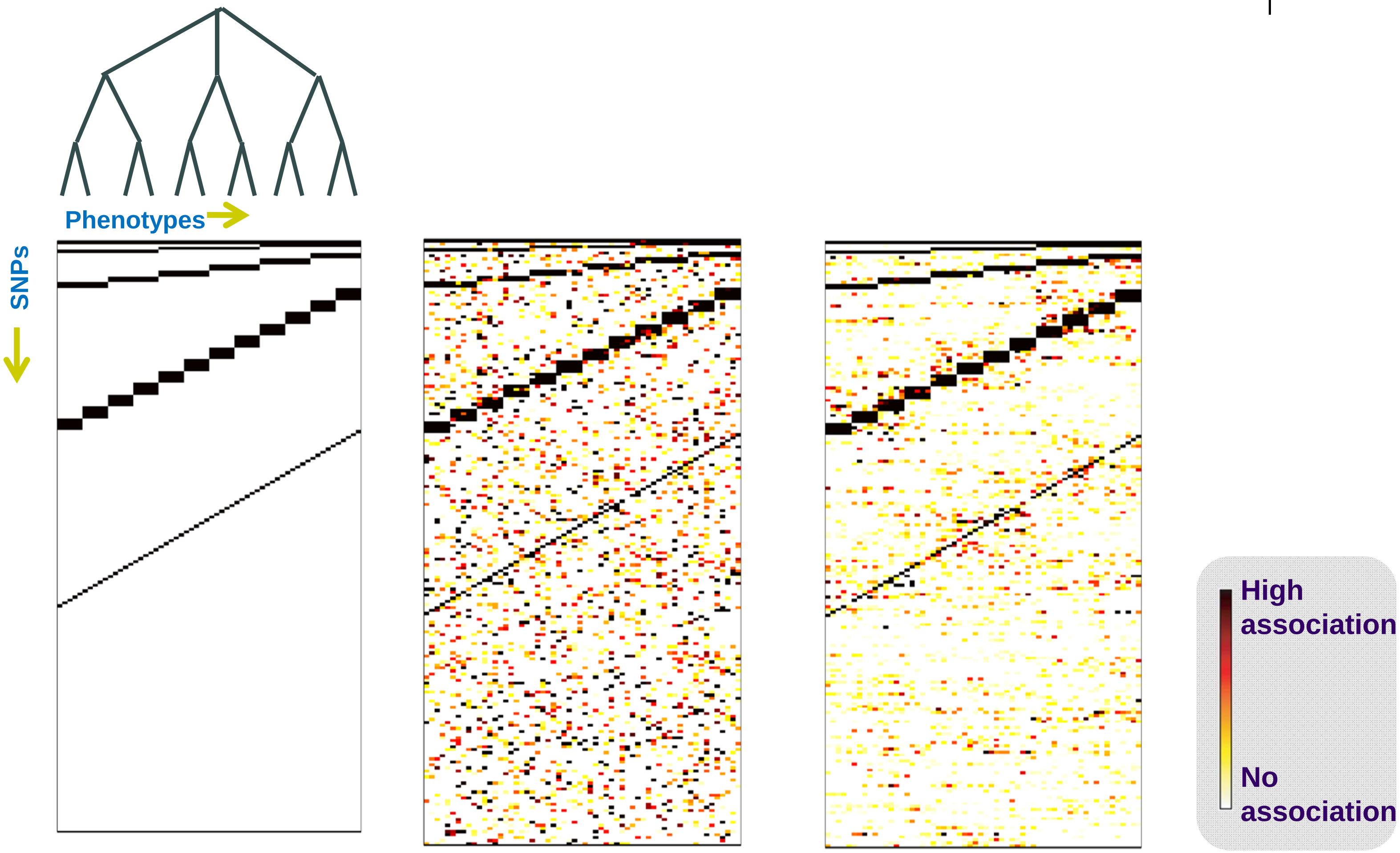
Estimating Parameters

- Second-order cone program

$$\hat{\mathrm{B}}^T \quad = \quad \mathrm{argmin} \quad \sum_k (\mathbf{y}_k - \mathbf{X}\beta_k)^T \cdot (\mathbf{y}_k - \mathbf{X}\beta_k) + \lambda \sum_j \sum_{v \in V} w_v \|\beta^j_{G_v}\|_2$$

  - Many publicly available software packages for solving convex optimization problems can be used

Illustration with Simulated Data

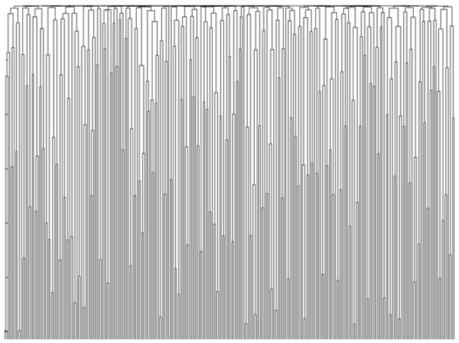


True association strengths

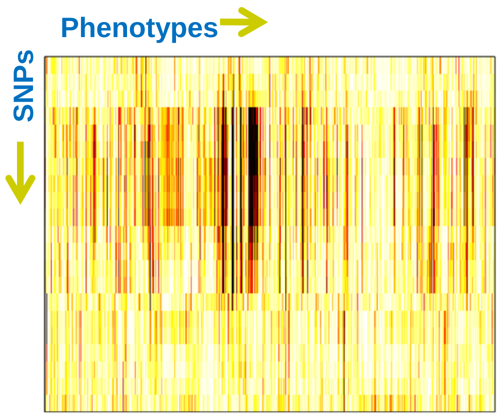
Lasso

Tree-guided group lasso

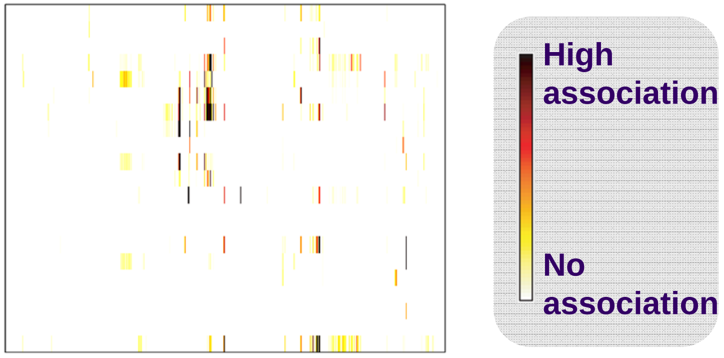
Yeast eQTL Analysis



Hierarchical clustering tree

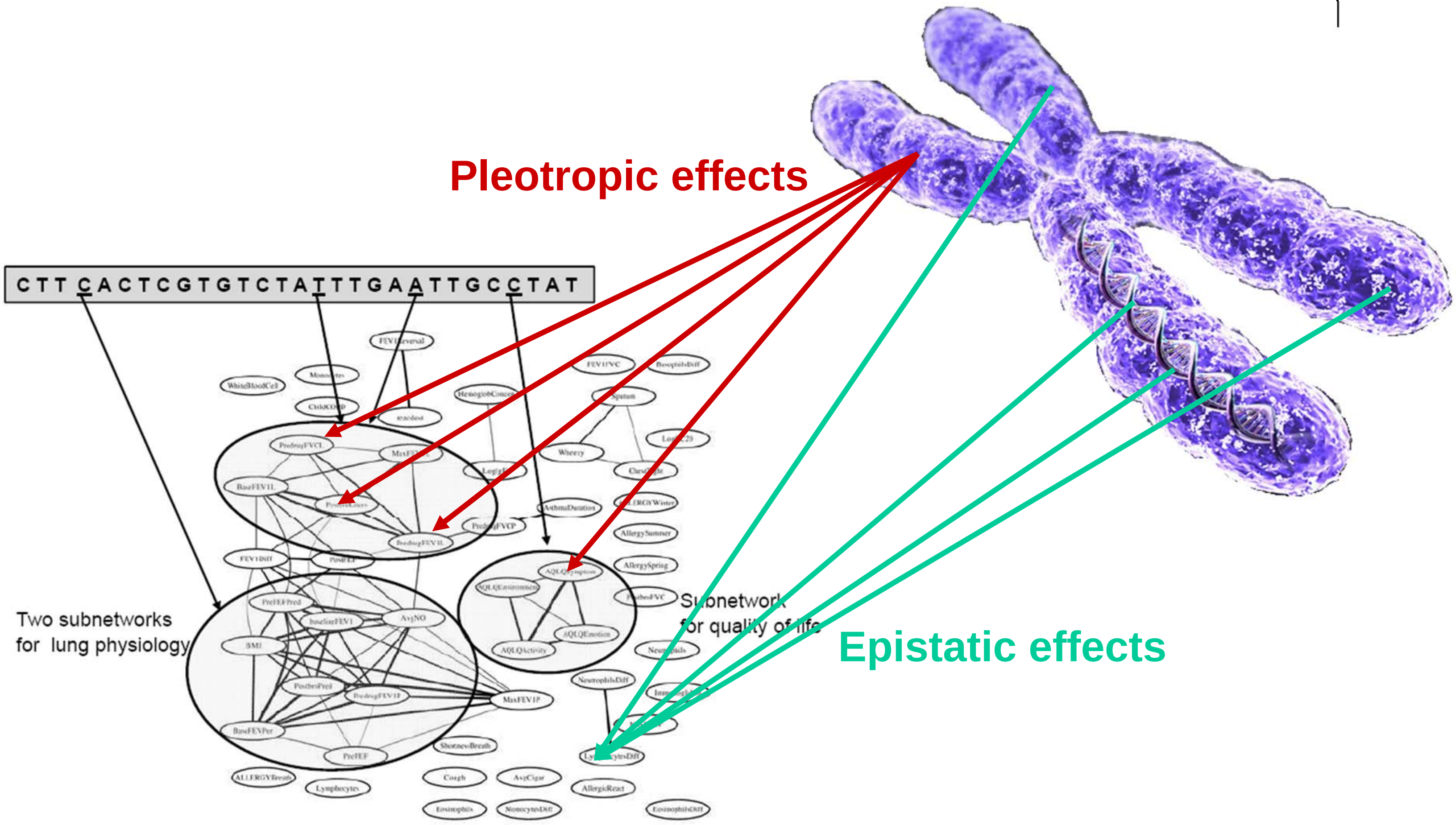


Single-Marker
Single-Trait Test



Tree-guided
group lasso

Ultimately …



Pleotropic effects

Epistatic effects

CTT CACTCGTGTCTATTTGAATTGCCTAT
Two subnetworks for lung physiology
Subnetwork for quality of life

Structured Input/Output-Lasso

[Lee, Zhu and Xing, submitted 2010]

$$\beta_{io-lasso} = \arg\min_{\beta} \sum_{k=1}^{K}\sum_{i=1}^{N}\left(Y_i^k - \sum_{j=1}^{p}\beta_j^k X_{ij} - \sum_{(r,s)\in U}\beta_{rs}^k Z_{i,rs}\right) + \lambda_1 \sum_{j=1}\sum_{k=1}\left|\beta_j^k\right|$$

$$+ \lambda_2 \sum_{k}\sum_{m}\sqrt{\sum_{(r,s)\in S_m}{\beta_{rs}^k}^2}$$

$$+ \lambda_3 \sum_{j}\sqrt{\sum_{k}{\beta_j^k}^2}$$

$$+ \lambda_4 \sum_{k}\sum_{(r.s)\in U}\left|\beta_{rs}^k\right|$$

Output structure: group selection of SNPs associated with multiple correlated traits

Input structure: group selection of correlated SNPs

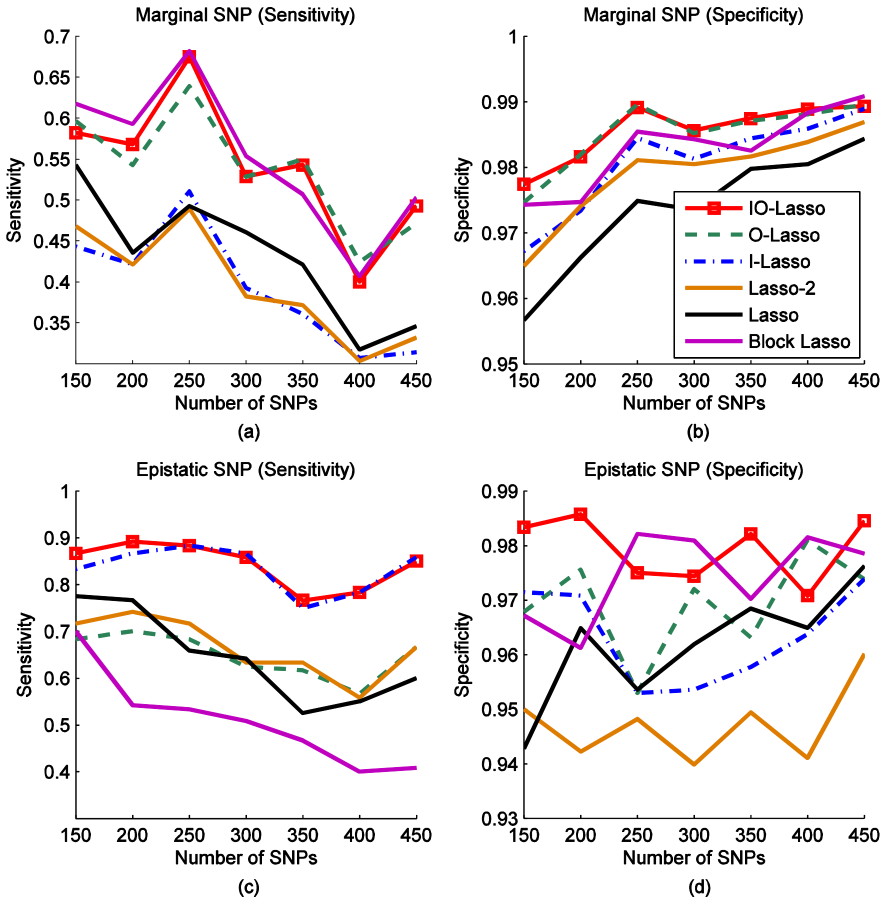
Lasso penalty: within group sparsity

U : genetic interaction networks

$\mathrm{S_m}$: $\mathrm{m^{th}}$ cluster in SNP network

This full model incorporates input/output structure of the dataset as well as epistatic effects guided by genetic interaction networks

Sensitivity and Specificity varying the number of SNPs



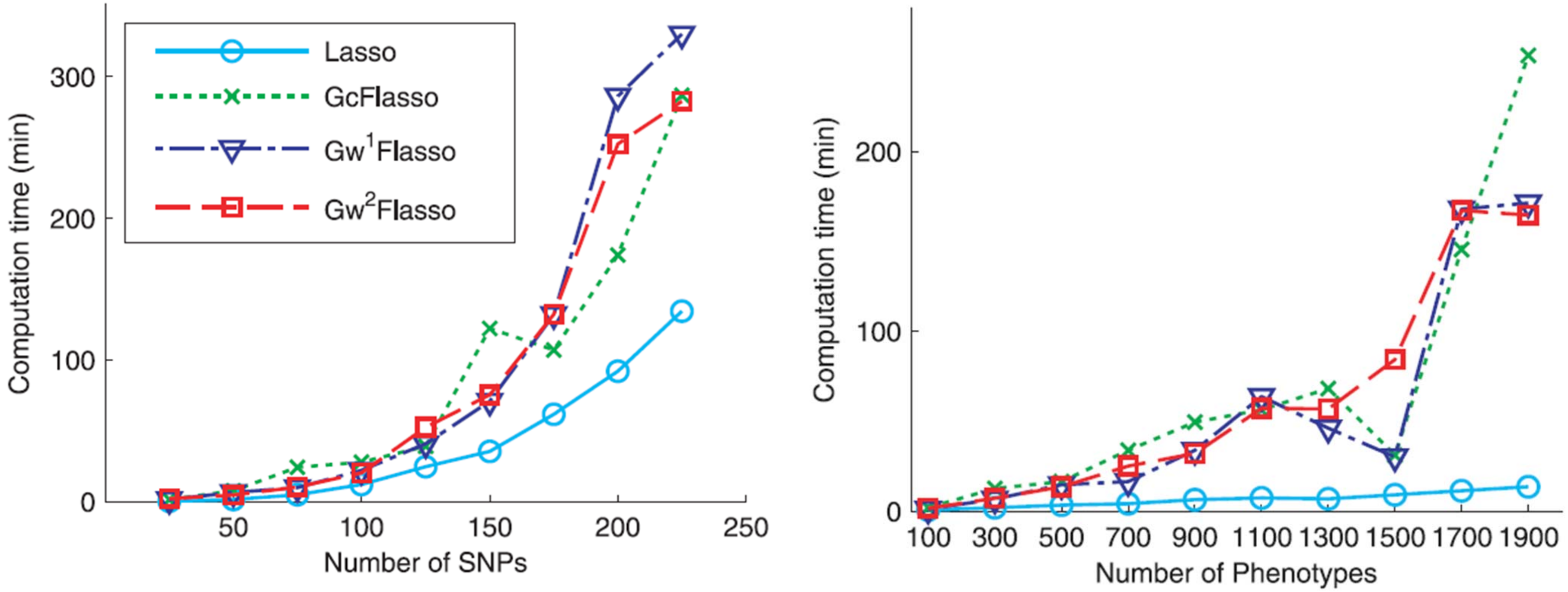
❑ Marginal SNP: Methods taking advantage of output structures outperforms others.

❑ Epistatic SNP: Methods taking advantage of input structures outperforms others.

❑ IO-Lasso outperforms other methods for detecting both marginal & epsitatic eQTLs

❖ For each number of SNPs, we show the average of the performance with 5 different simulated data

Computation Time



Proximal Gradient Descent

Original Problem:

$$\arg\min_{\boldsymbol{\beta}\in\mathbb{R}^J} f(\boldsymbol{\beta}) \equiv \frac{1}{2}\|\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\|_2^2 + \Omega(\boldsymbol{\beta})$$

$$\Omega(\boldsymbol{\beta}) = \max_{\boldsymbol{\alpha}\in\mathcal{Q}} \boldsymbol{\alpha}^T C\boldsymbol{\beta}$$

Approximation Problem:

$$\arg\min_{\boldsymbol{\beta}\in\mathbb{R}^J} \widetilde{f}(\boldsymbol{\beta}) \equiv \frac{1}{2}\|\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\|_2^2 + f_\mu(\boldsymbol{\beta})$$

$$f_\mu(\boldsymbol{\beta}) = \max_{\boldsymbol{\alpha}\in\mathcal{Q}} \boldsymbol{\alpha}^T C\boldsymbol{\beta} - \mu d(\boldsymbol{\alpha})$$

Gradient of the Approximation:

$$\nabla\widetilde{f}(\boldsymbol{\beta}) = \mathbf{X}^T(\mathbf{X}\boldsymbol{\beta}-\mathbf{y}) + C^T\boldsymbol{\alpha}^*$$

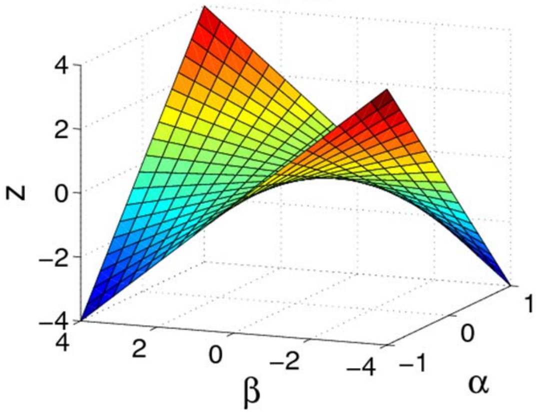
$$\boldsymbol{\alpha}^* = \arg\max_{\boldsymbol{\alpha}\in\mathcal{Q}} \boldsymbol{\alpha}^T C\boldsymbol{\beta} - \mu d(\boldsymbol{\alpha})$$

$\nabla\widetilde{f}(\boldsymbol{\beta})$ is Lipschitz continuous with the Lipschitz constant L

$$L = \lambda_{\max}(\mathbf{X}^T\mathbf{X}) + L_\mu$$

Geometric Interpretation

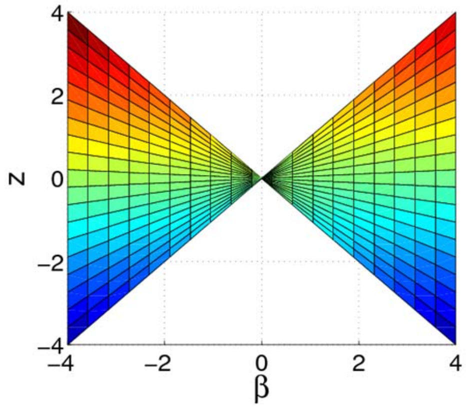
- Smooth approximation



$$z(\alpha,\beta) = \alpha\beta$$

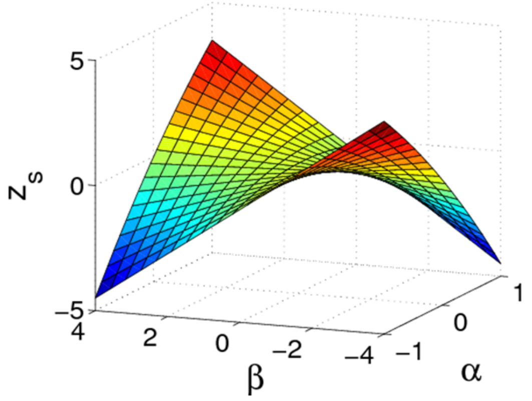
Projection onto
$z - \beta$ Plane

$\longrightarrow$



$$f_0(\beta) = \max_{\alpha\in[-1,1]} z(\alpha,\beta) = |\beta|$$

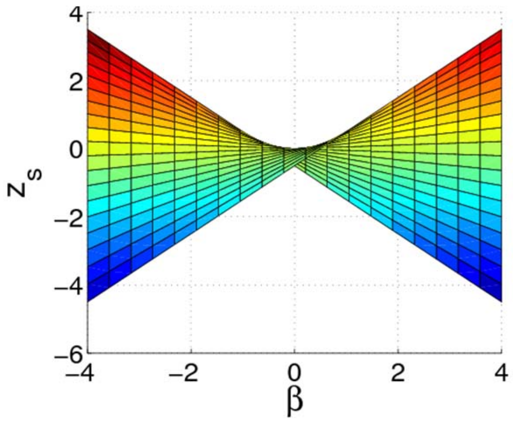
Uppermost Line Nonsmooth ☹



$$z_s(\alpha,\beta) = \alpha\beta - \frac{1}{2}\alpha^2$$

Projection onto
$z_s - \beta$ Plane

$\longrightarrow$



$$f_1(\beta) = \max_{\alpha\in[-1,1]} z_s(\alpha,\beta)$$

Uppermost Line Smooth ☺

Convergence Rate

Theorem: If we require $f(\boldsymbol{\beta}^t) - f(\boldsymbol{\beta}^*) \leq \epsilon$ and set $\mu = \frac{\epsilon}{2D}$, the number of iterations is upper bounded by:

$$t \leq \sqrt{\frac{4\|\boldsymbol{\beta}^*\|_2^2}{\epsilon}\left(\lambda_{\max}(\mathbf{X}^T\mathbf{X}) + \frac{2D\|\Gamma\|^2}{\epsilon}\right)} = O(\frac{1}{\epsilon})$$

Remarks: state of the art IPM method for for SOCP converges at a rate $O(\frac{1}{\epsilon^2})$

Multi-Task Time Complexity

- Pre-compute: $\mathbf{X}^T\mathbf{X},\ \mathbf{X}^T\mathbf{Y}:\quad O(J^2N + JKN)$
- Per-iteration Complexity (computing gradient)

Tree:

IPM for SOCP	$O\left(J^2(K+\lvert\mathcal{G}\rvert)^2(KN+J(\sum_{g\in\mathcal{G}}\lvert g\rvert))\right)$
Proximal-Gradient	$O(J^2K + J\sum_{g\in\mathcal{G}}\lvert g\rvert)$

Graph:

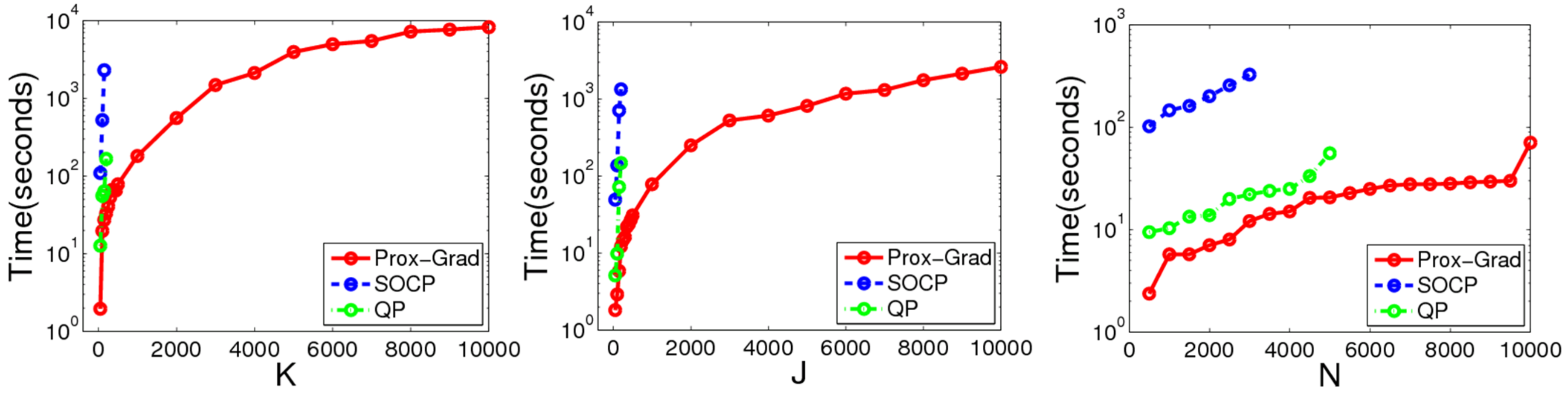
IPM for SOCP	$O\left(J^2(K+\lvert E\rvert)^2(KN+JK+J\lvert E\rvert)\right)$
Proximal-Gradient	$O(J^2K + J\lvert E\rvert)$

Proximal-Gradient: Independent of Sample Size

Linear in #.of Tasks

Experiments

- Multi-task Graph Structured Sparse Learning (GFlasso)

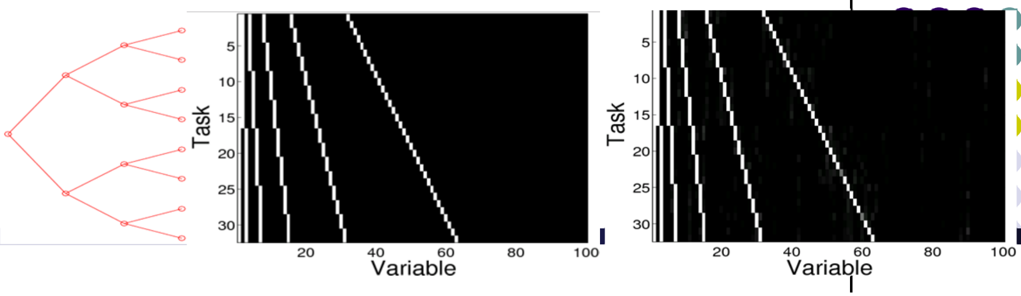


$N = 500, J = 100$ $N = 1000, K = 50$ $J = 100, K = 50$

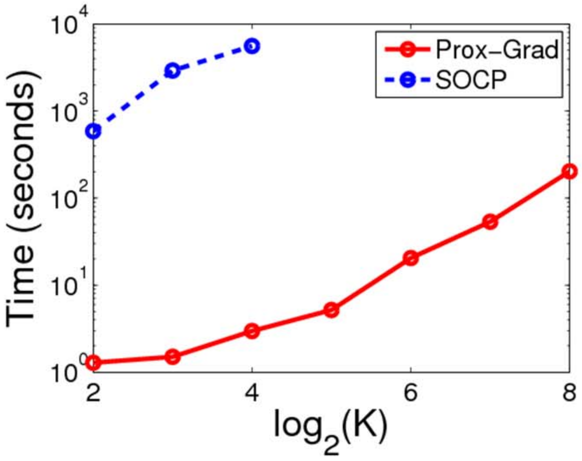
$$\mu = 10^{-4}, \rho = 0.5$$

© Eric Xing @ CMU, 2014

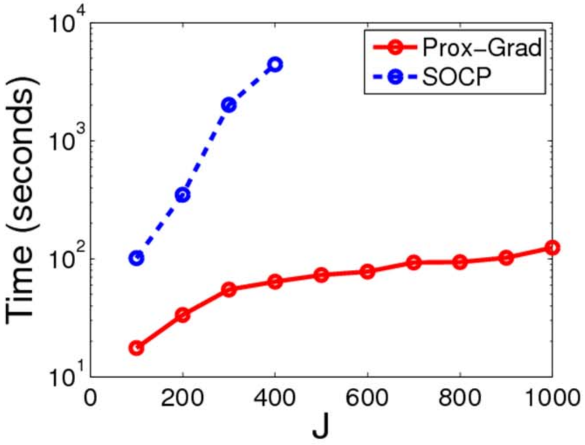
Experiments



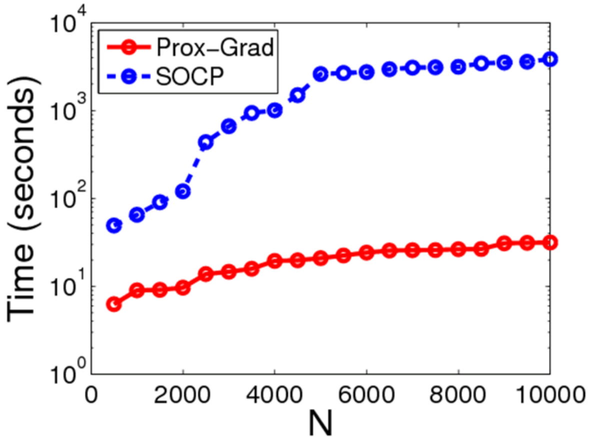
- Multi-task Tree-Structured Sparse Learning (TreeLasso)



$$N = 1000, J = 600$$



$$N = 1000, K = 32$$



$$J = 100, K = 32$$

$$\epsilon = 0.1$$

Conclusions

- Novel statistical methods for joint association analysis to correlated phenotypes
  - Graph-structured phenome : graph-guided fused lasso
  - Tree-structured phenome : tree-guided group lasso

- Advantages
  - Greater power to detect weak association signals
  - Fewer false positives
  - Joint association to multiple correlated phenotypes