

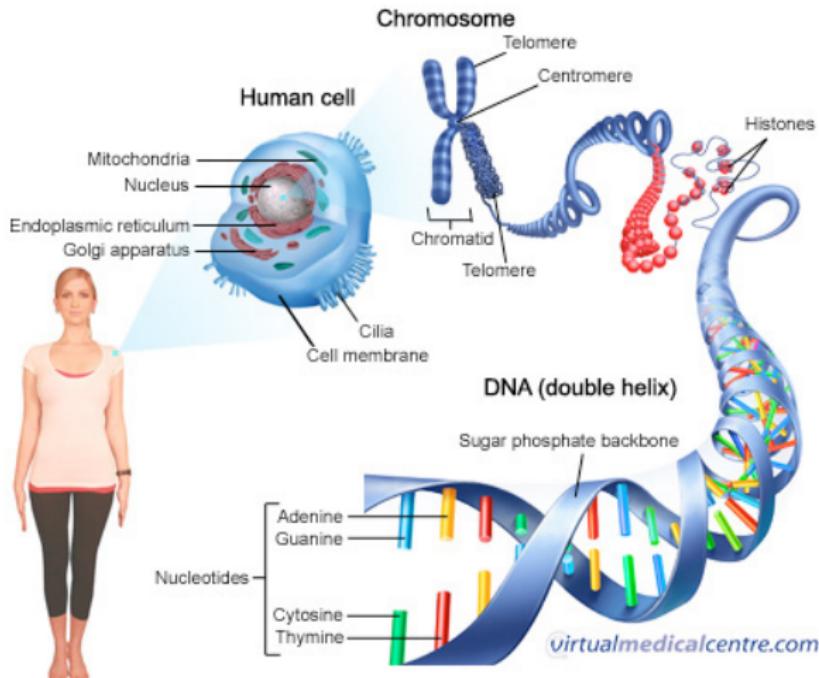
Machine Learning for Personalized Medicine

Jean-Philippe Vert



SeMoVi seminar, Grenoble, May 14, 2014

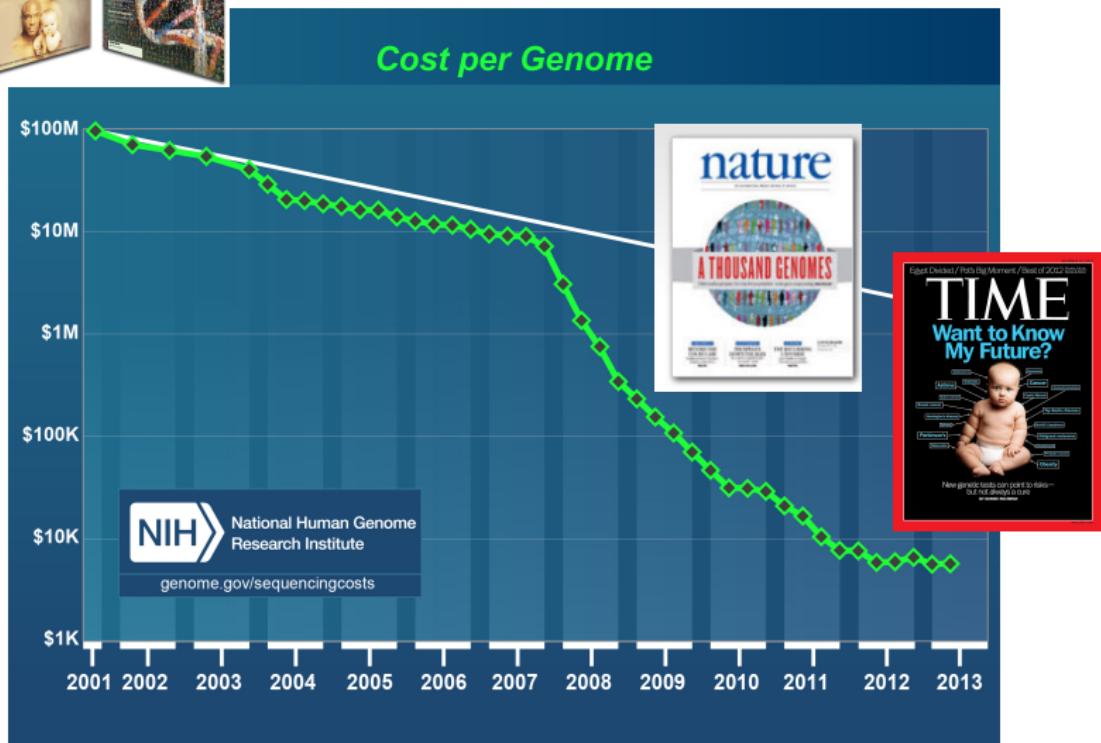
Complexity of life



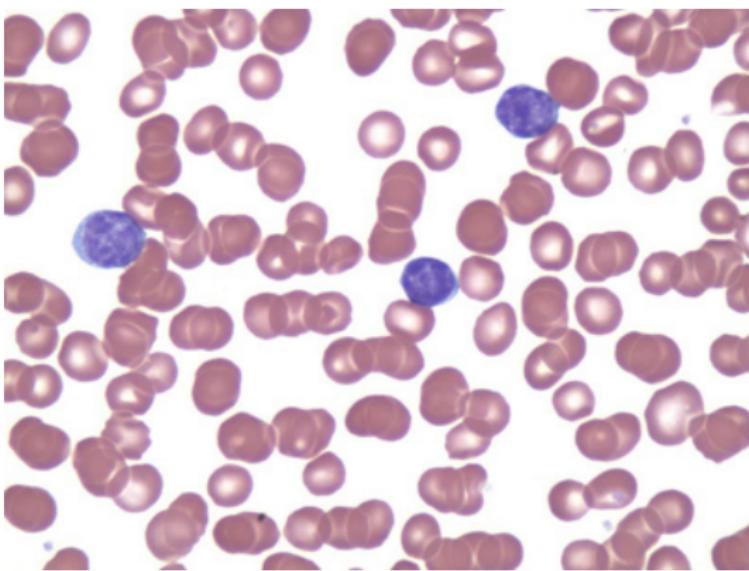
1 body = 10^{14} cells

1 cell = 6×10^9 ACGT coding for 20,000 genes

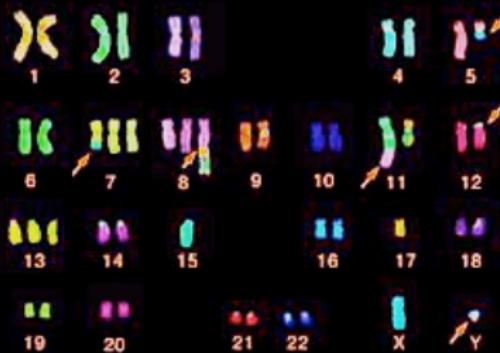
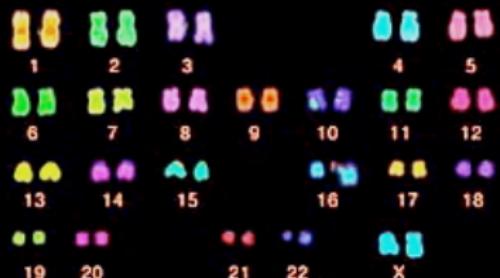
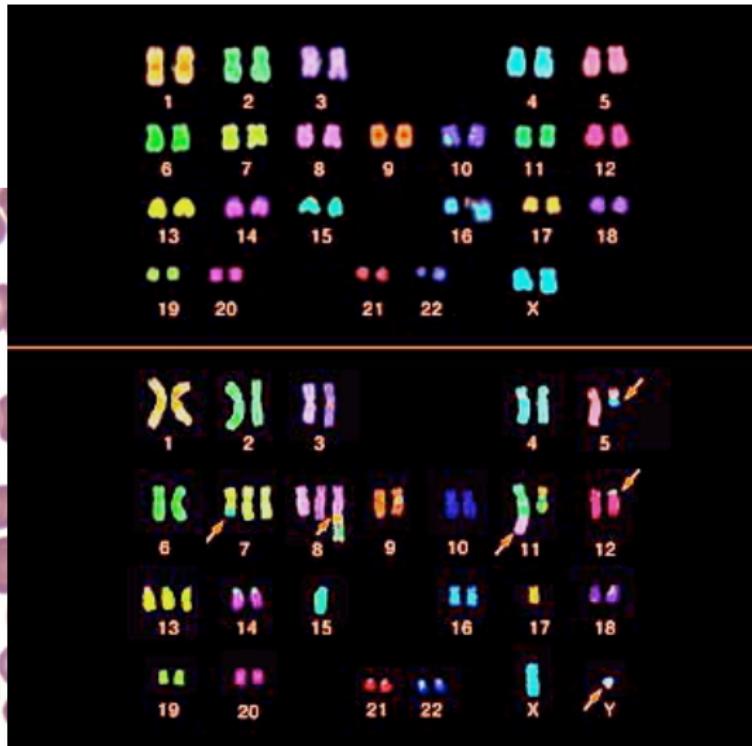
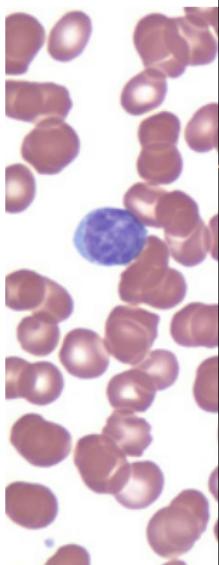
Sequencing revolution



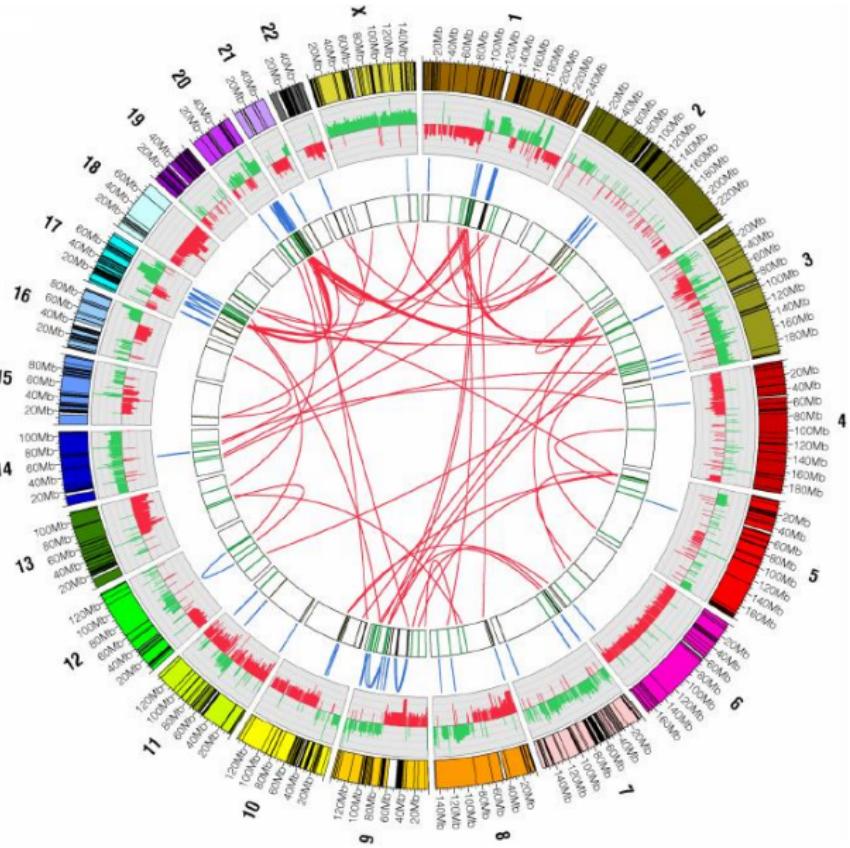
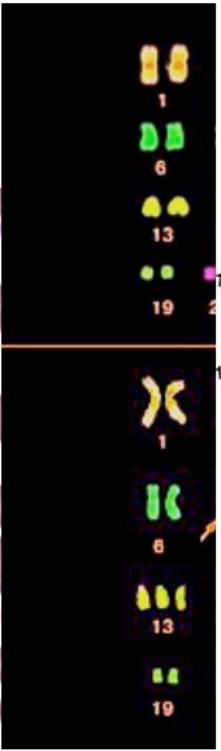
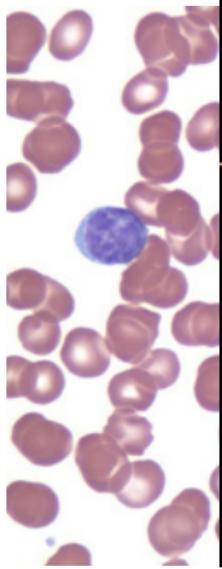
A cancer cell



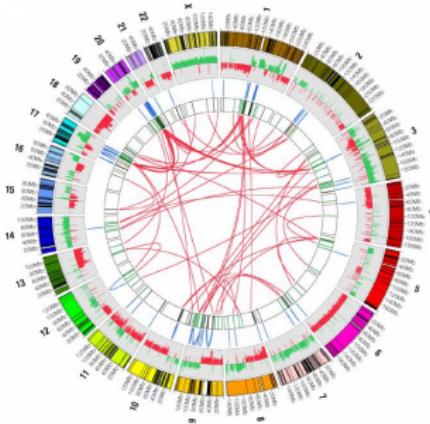
A cancer cell



A cancer cell

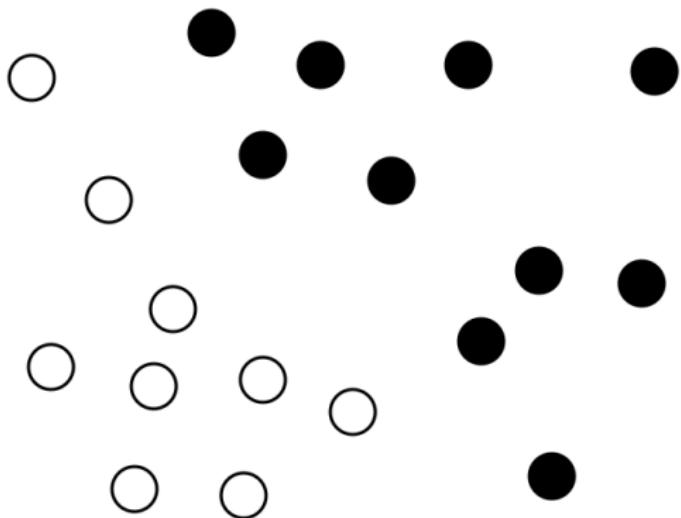


Opportunities

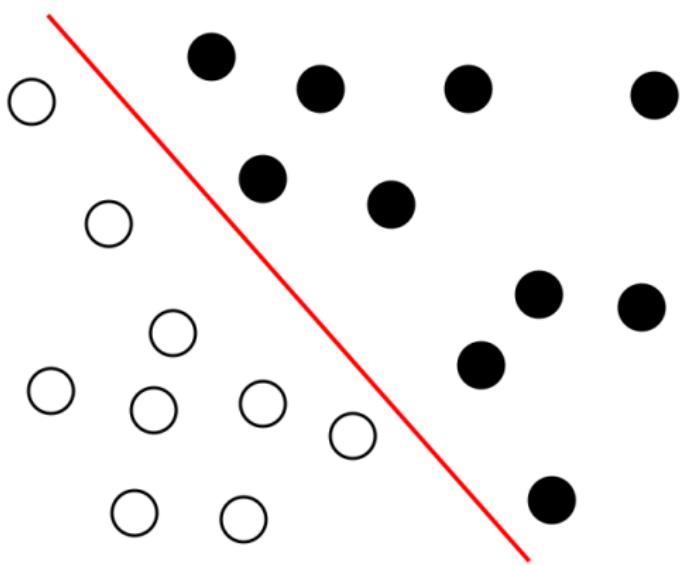


- What is your risk of developing a cancer? (*prevention*)
- After diagnosis and treatment, what is the risk of relapse? (*prognosis*)
- What specific treatment will cure your cancer? (*personalized medicine*)

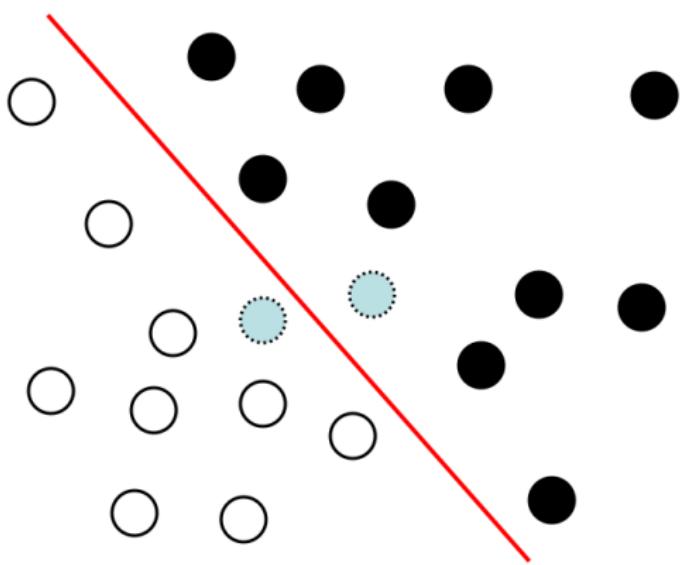
Machine learning formulation



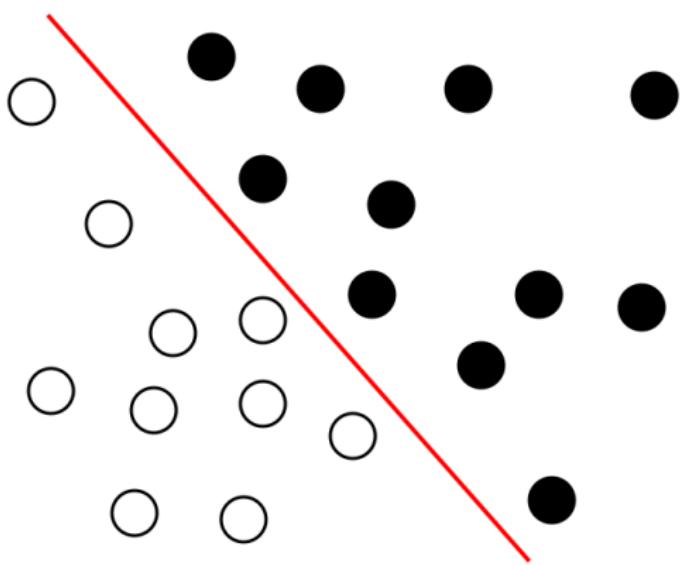
Machine learning formulation



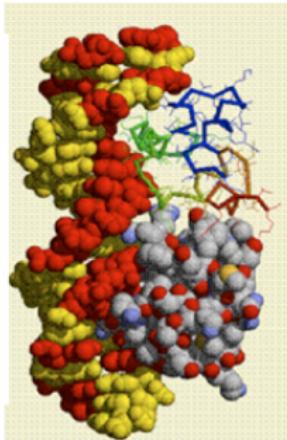
Machine learning formulation



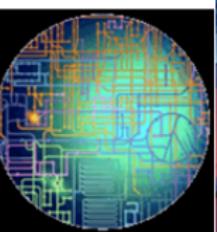
Machine learning formulation



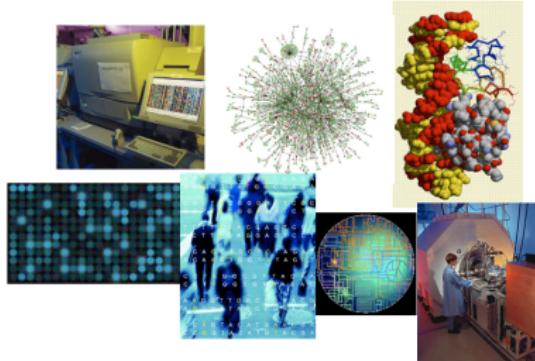
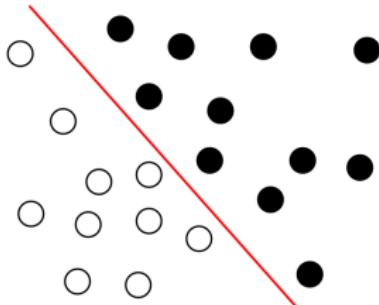
On real data...



ATG C G G G G A G
T G G G T A T G C C G A G A
C G G T C G G G A A T C C C
G G T C G G G A A T C C C
T D T T G A C G A C T C C C
G C T T G A G G A C T C C C
G A G G T O G T G T A G A G A
G A A C T G G T T A T A G G T
C C T T G G G T T A C C A A
C T T T G G C T T A C C A C A C
A A G G T T G G C T T A C C A C
A G G G T A R C C G A A C G A
C C A G T A C A T G A A C G A
C C G G T A C A T G T A C G A

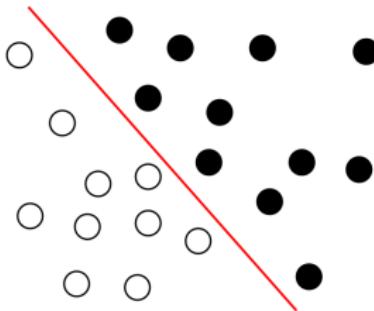


Challenges



- High dimension
- Few samples
- Structured data
- Heterogeneous data
- Prior knowledge
- Fast and scalable implementations
- Interpretable models

Learning with regularization



Learn

$$f_{\beta}(x) = \beta^T x$$

by solving

$$\min_{\beta \in \mathbb{R}^p} R(f_{\beta}) + \lambda \Omega(\beta)$$

- $R(f_{\beta})$ empirical risk
- $\Omega(\beta)$ penalty

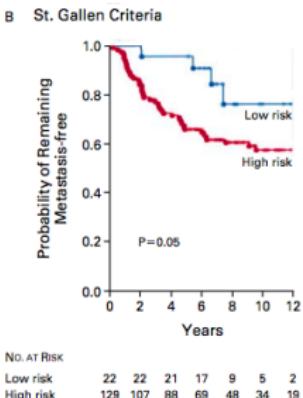
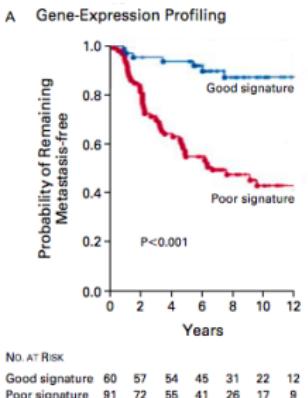
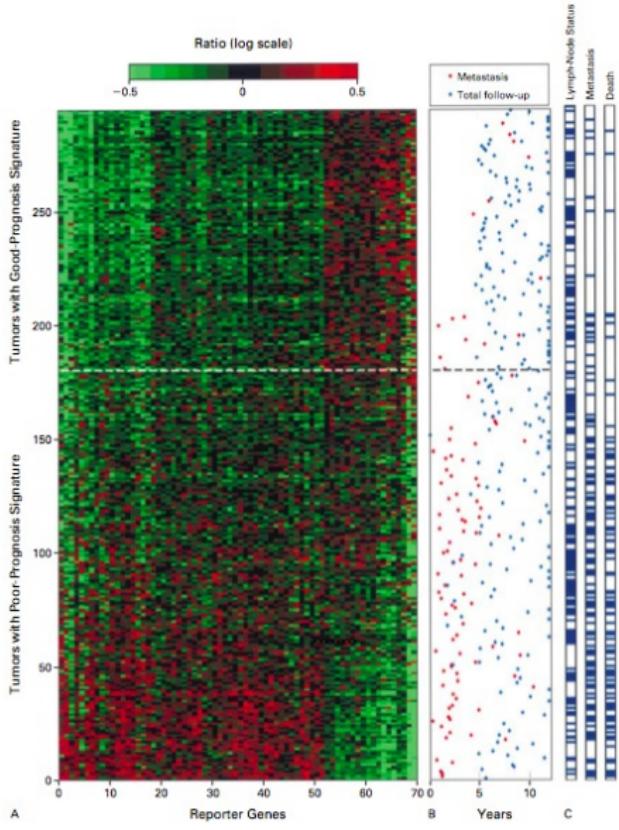
Outline

- 1 Learning molecular classifiers with network information
- 2 Kernel bilinear regression for toxicogenomics

Outline

- 1 Learning molecular classifiers with network information
- 2 Kernel bilinear regression for toxicogenomics

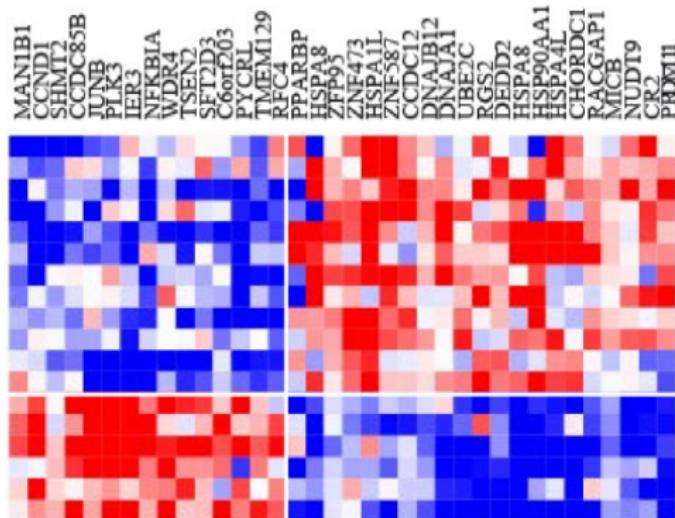
Breast cancer prognosis



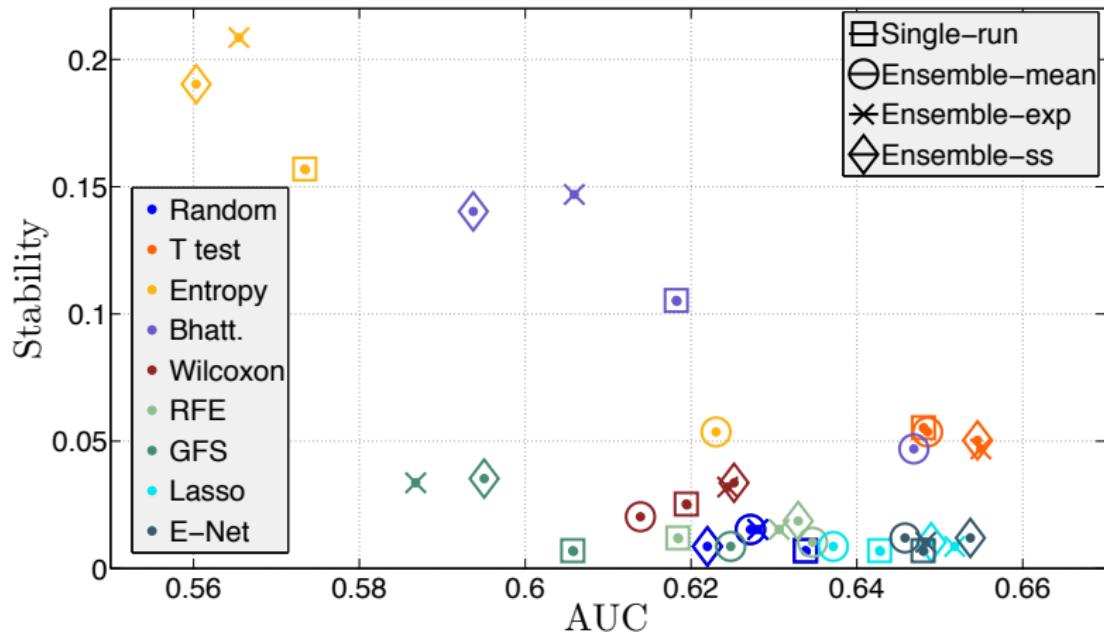
Gene selection, molecular signature

The idea

- We look for a **limited set** of genes that are sufficient for prediction.
- Selected genes should inform us about the underlying biology

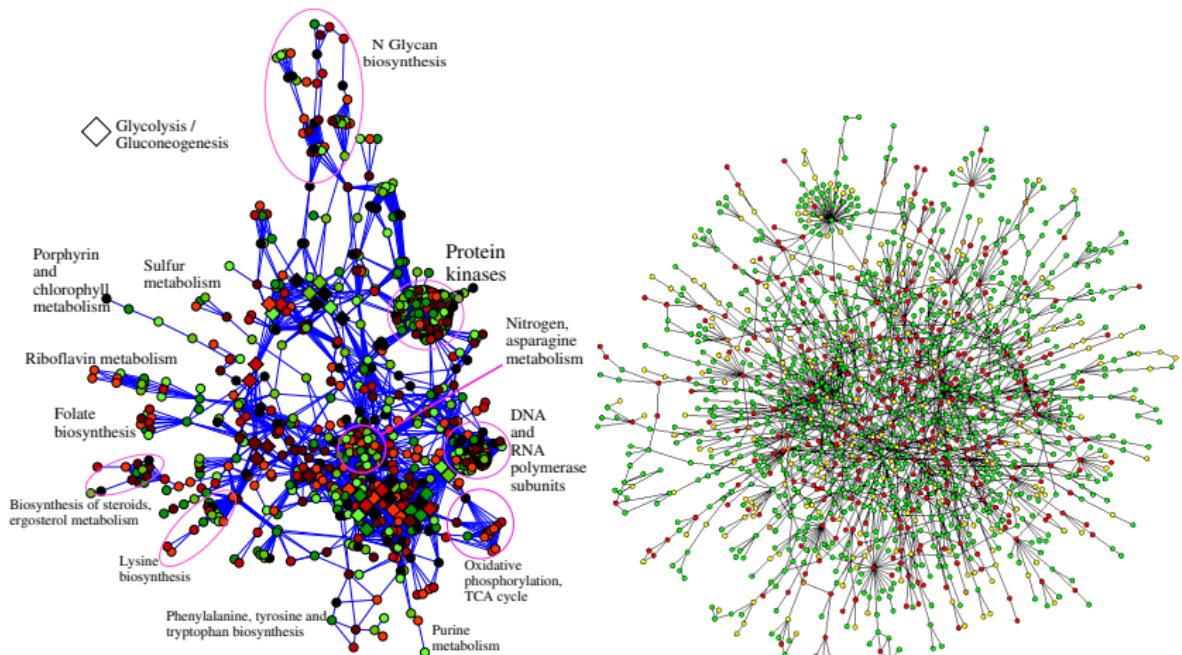


Lack of stability of signatures



Haurý et al. (2011)

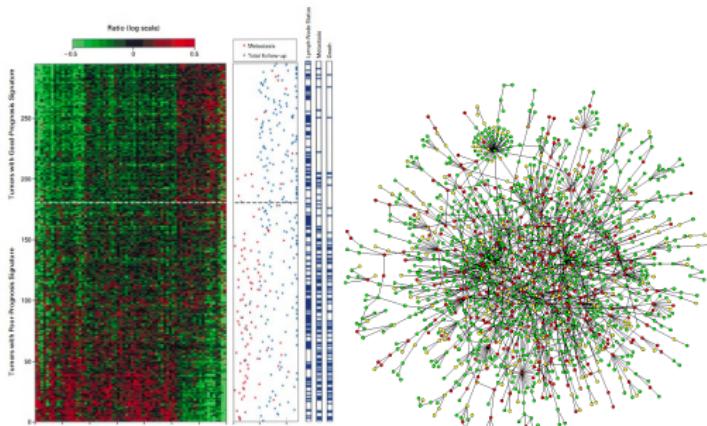
Gene networks



Gene networks and expression data

Motivation

- Basic biological functions usually involve the coordinated action of several proteins:
 - Formation of protein complexes
 - Activation of metabolic, signalling or regulatory pathways
 - Many pathways and protein-protein interactions are already known
 - Hypothesis: the weights of the classifier should be “coherent” with respect to this prior knowledge



Graph based penalty

$$f_{\beta}(x) = \beta^T x \quad \min_{\beta} R(f_{\beta}) + \lambda \Omega(\beta)$$

Prior hypothesis

Genes near each other on the graph should have **similar weights**.

An idea (Rapaport et al., 2007)

$$\Omega(\beta) = \sum_{i \sim j} (\beta_i - \beta_j)^2,$$

$$\min_{\beta \in \mathbb{R}^p} R(f_{\beta}) + \lambda \sum_{i \sim j} (\beta_i - \beta_j)^2.$$

Graph based penalty

$$f_\beta(x) = \beta^\top x \quad \min_{\beta} R(f_\beta) + \lambda \Omega(\beta)$$

Prior hypothesis

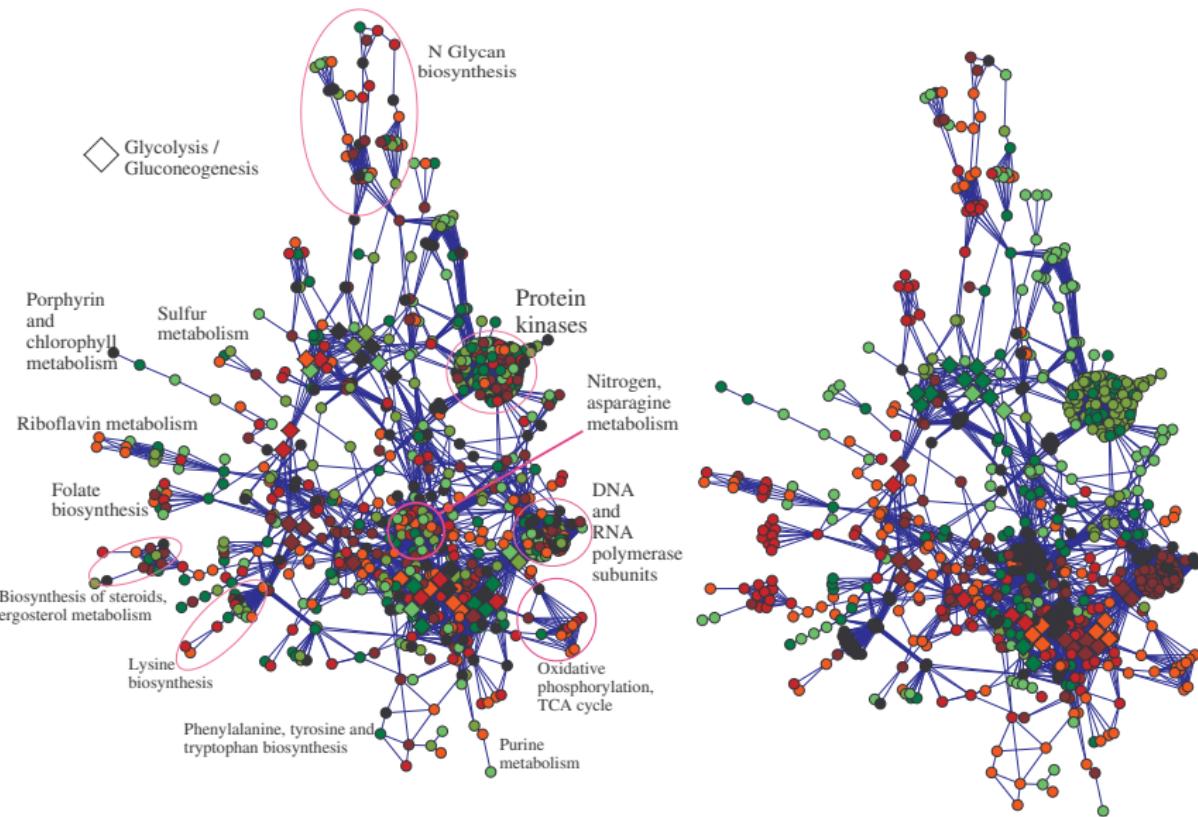
Genes near each other on the graph should have **similar weights**.

An idea (Rapaport et al., 2007)

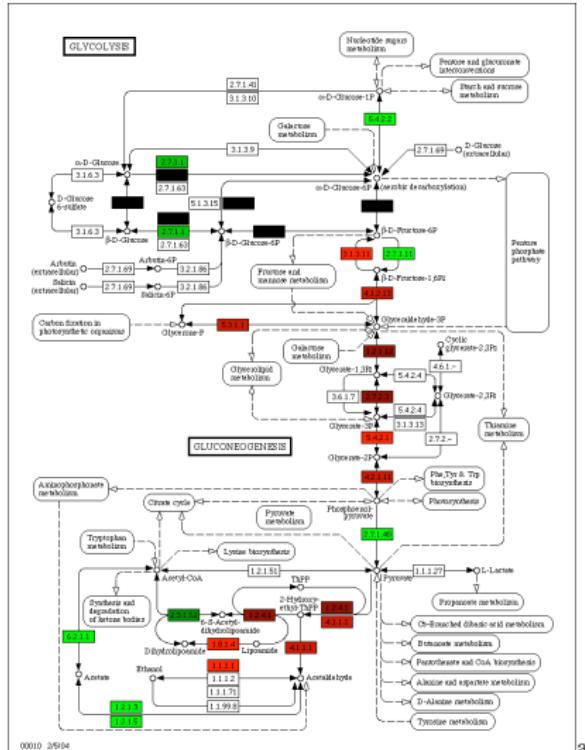
$$\Omega(\beta) = \sum_{i \sim j} (\beta_i - \beta_j)^2,$$

$$\min_{\beta \in \mathbb{R}^p} R(f_\beta) + \lambda \sum_{i \sim j} (\beta_i - \beta_j)^2.$$

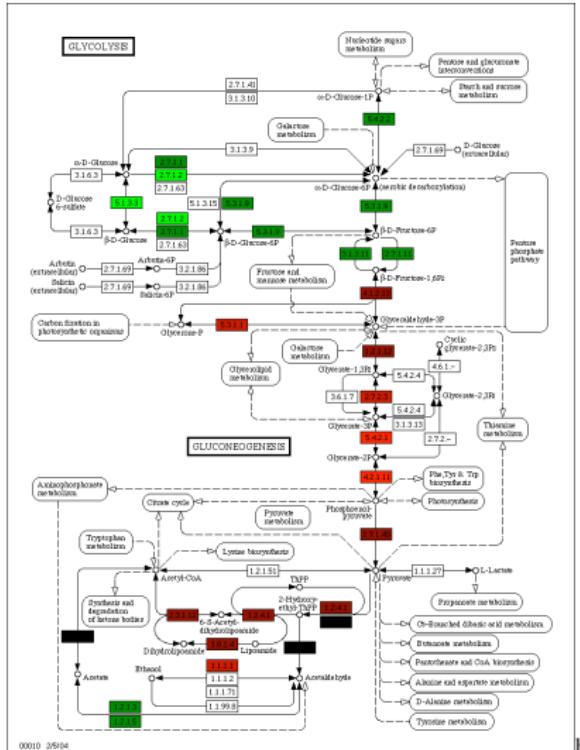
Classifiers



Classifier



a)



b)

Spectral penalty as a kernel

Theorem

The function $f(x) = \beta^\top x$ where β is solution of

$$\min_{\beta \in \mathbb{R}^p} \frac{1}{n} \sum_{i=1}^n \ell(\beta^\top \mathbf{x}_i, y_i) + \lambda \sum_{i \sim j} (\beta_i - \beta_j)^2$$

is equal to $g(x) = \gamma^\top \Phi(x)$ where γ is solution of

$$\min_{\gamma \in \mathbb{R}^p} \frac{1}{n} \sum_{i=1}^n \ell(\gamma^\top \Phi(x_i), y_i) + \lambda \gamma^\top \gamma,$$

and where

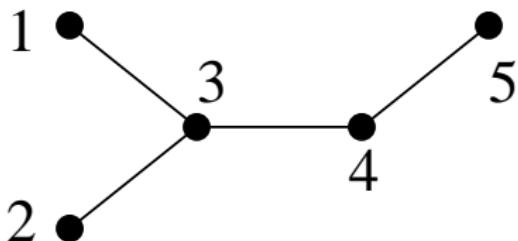
$$\Phi(x)^\top \Phi(x') = x^\top K_G x'$$

for $K_G = L^*$, the pseudo-inverse of the graph Laplacian.

Graph Laplacian

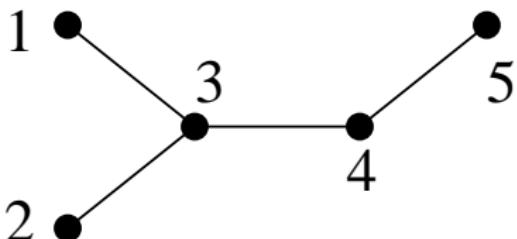
Definition

The Laplacian of the graph is the matrix $L = D - A$.



$$L = D - A = \begin{pmatrix} 1 & 0 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ -1 & -1 & 3 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

Pseufo-inverse of the Laplacian



$$L^* = \begin{pmatrix} 0.88 & -0.12 & 0.08 & -0.32 & -0.52 \\ -0.12 & 0.88 & 0.08 & -0.32 & -0.52 \\ 0.08 & 0.08 & 0.28 & -0.12 & -0.32 \\ -0.32 & -0.32 & -0.12 & 0.48 & 0.28 \\ -0.52 & -0.52 & -0.32 & 0.28 & 1.08 \end{pmatrix}$$

Other penalties with kernels

$$\Phi(x)^\top \Phi(x') = x^\top K_G x'$$

with:

- $K_G = (c + L)^{-1}$ leads to

$$\Omega(\beta) = c \sum_{i=1}^p \beta_i^2 + \sum_{i \sim j} (\beta_i - \beta_j)^2.$$

- The diffusion kernel:

$$K_G = \exp_M(-2tL).$$

penalizes high frequencies of β in the Fourier domain.

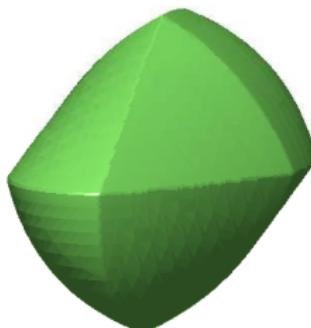
Other penalties without kernels

- Gene selection + Piecewise constant on the graph

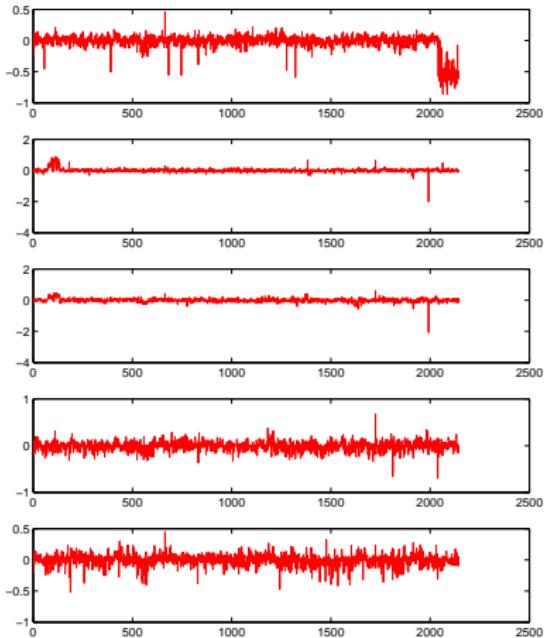
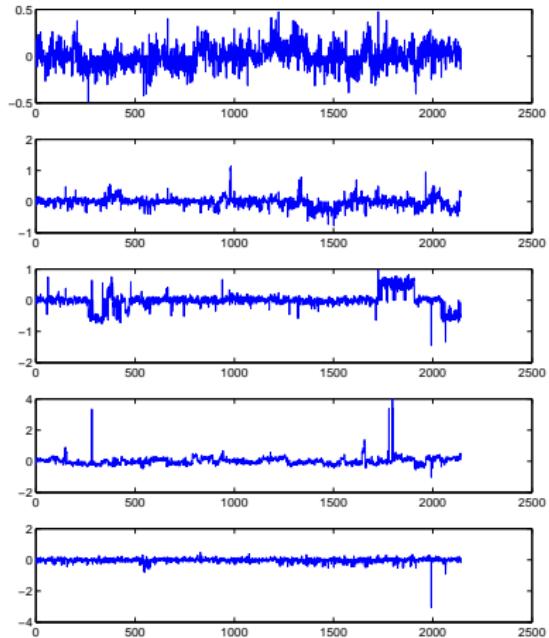
$$\Omega(\beta) = \sum_{i \sim j} |\beta_i - \beta_j| + \sum_{i=1}^p |\beta_i|$$

- Gene selection + smooth on the graph

$$\Omega(\beta) = \sum_{i \sim j} (\beta_i - \beta_j)^2 + \sum_{i=1}^p |\beta_i|$$



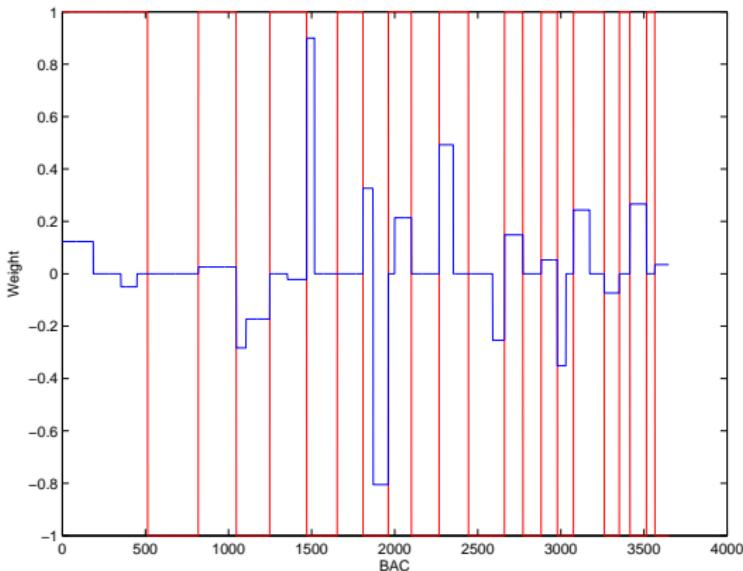
Example: classification of DNA copy number profiles



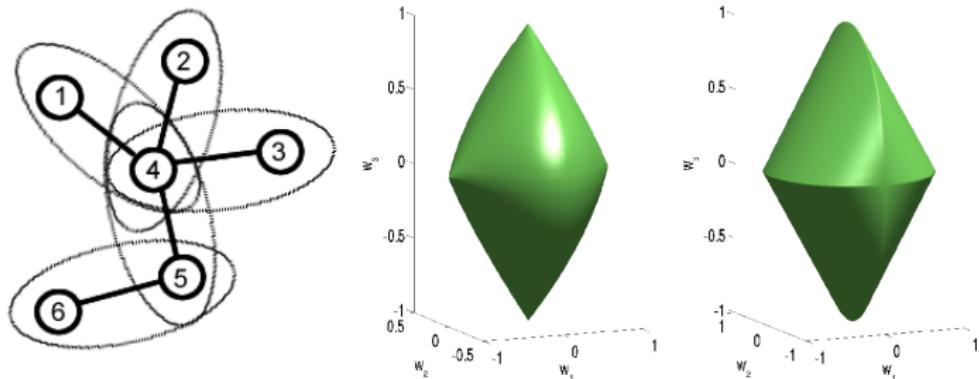
Aggressive (left) vs non-aggressive (right) melanoma

Fused lasso solution (Rapaport et al., 2008)

$$\Omega(\beta) = \sum_{i \sim j} |\beta_i - \beta_j| + \sum_{i=1}^p |\beta_i|$$



Graph-based structured feature selection

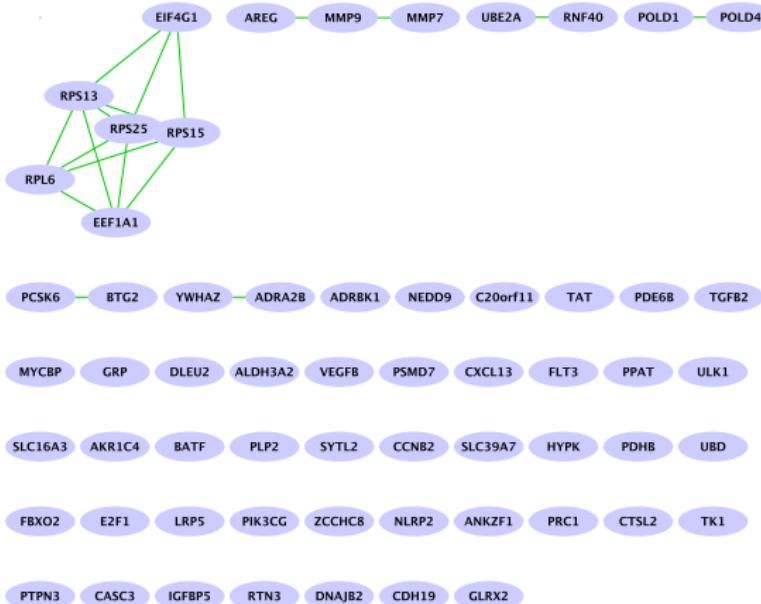


Graph lasso(s)

$$\Omega_1(\beta) = \sum_{i \sim j} \sqrt{\beta_i^2 + \beta_j^2}, \quad (\text{Jenatton et al., 2009})$$

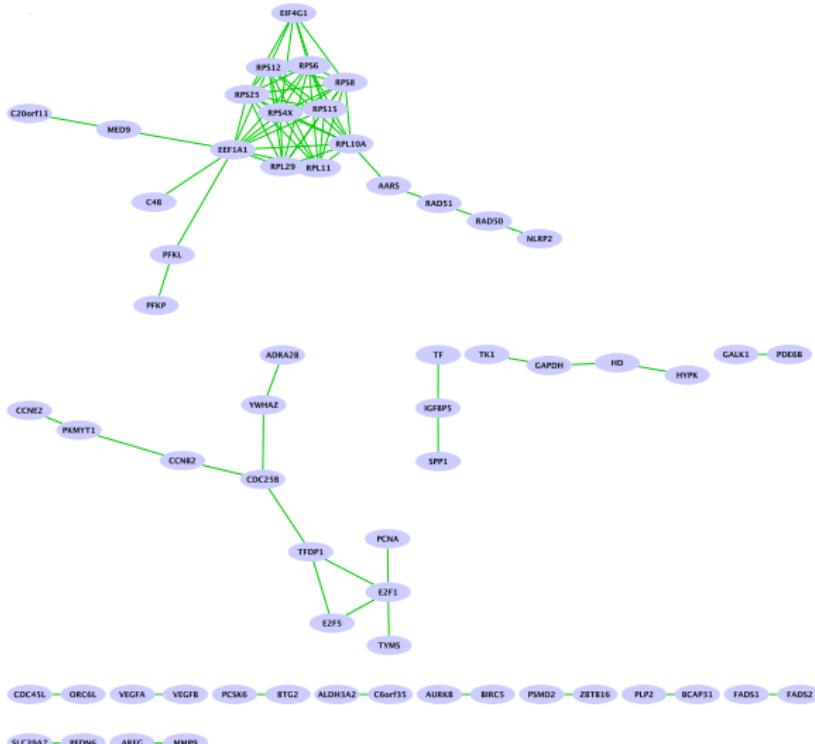
$$\Omega_2(\beta) = \sup_{\alpha \in \mathbb{R}^p : \forall i \sim j, \|\alpha_i^2 + \alpha_j^2\| \leq 1} \alpha^\top \beta. \quad (\text{Jacob et al., 2008})$$

Lasso signature (accuracy 0.61)



Breast cancer prognosis

Graph Lasso signature (accuracy 0.64)

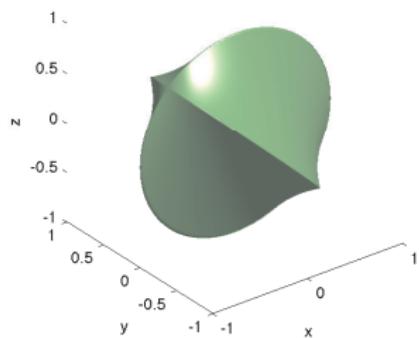
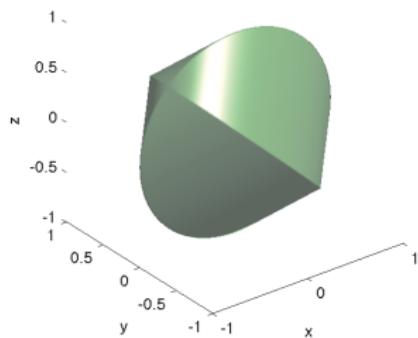
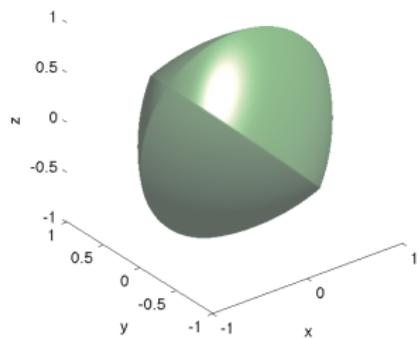


Breast cancer prognosis

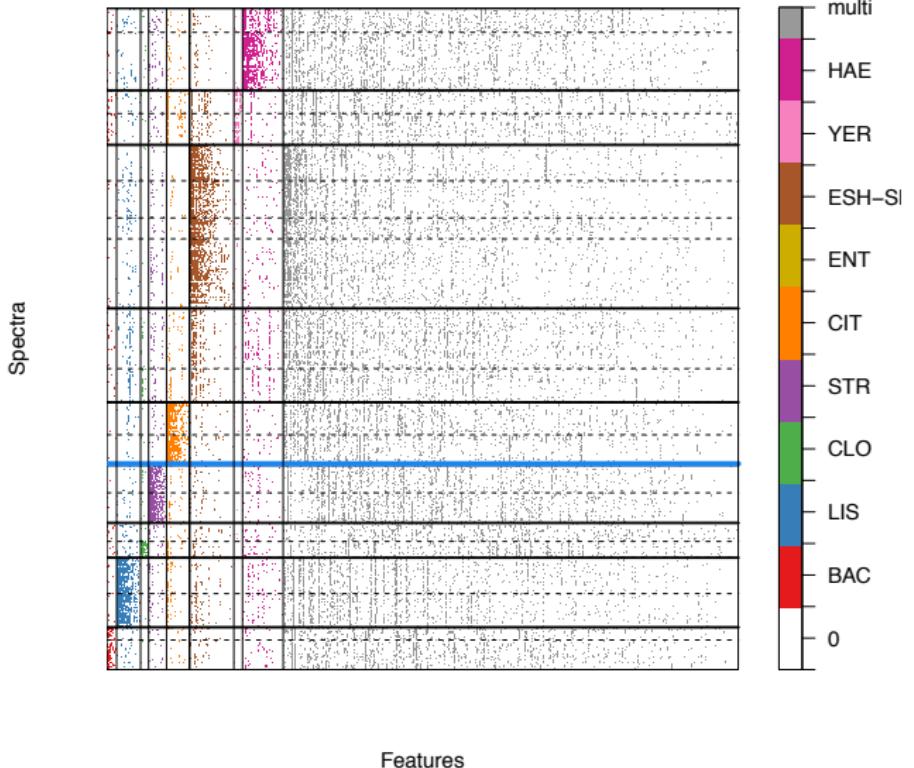
Disjoint feature selection (Vervier et al., 2014)

$$W = (w_i)_{i \in V} \in \mathbb{R}^{p \times V}$$

$$\Omega(W) = \min_{-H \leq W \leq H} \sum_{i \sim j} K_{ij} |h_i^\top h_j|$$



Example: multiclass classification of MS spectra



(Vervier et al, 2013, unpublished)

Outline

- 1 Learning molecular classifiers with network information
- 2 Kernel bilinear regression for toxicogenomics

Pharmacogenomics / Toxicogenomics



DREAM8 Toxicogenetics challenge

Toxicogenetics Challenge Data

Chemical
descriptors

10K attributes

Genotypes		Cytotoxicity data (EC_{10})		884 Cell Lines	
		Training Set			
1.3M SNPs	337 LCLs	487 Cell Lines	Test Set Subchallenge 2		
Not available	RNASeq	106 chemicals	Test Set		
46K transcripts	Not available	156 chemicals	Subchallenge 1		

Genotypes from the 1000 genome project
RNASeq from the Geuvadis project

Bilinear regression

- Cell line X , chemical Y , toxicity Z .
- Bilinear regression model:

$$Z = f(X, Y) + b(Y) + \epsilon,$$

- Estimation by kernel ridge regression:

$$\min_{f \in \mathcal{H}, b \in \mathbb{R}^p} \sum_{i=1}^n \sum_{j=1}^p (f(x_i, y_j) + b_j - z_{ij})^2 + \lambda \|f\|^2,$$

Solving in $O(\max(n, p)^3)$

Theorem 1. Let $Z \in \mathbb{R}^{n \times p}$ be the response matrix, and $K_X \in \mathbb{R}^{n \times n}$ and $K_Y \in \mathbb{R}^{p \times p}$ be the kernel Gram matrices of the n cell lines and p chemicals, with respective eigenvalue decompositions $K_X = U_X D_X U_X^\top$ and $K_Y = U_Y D_Y U_Y^\top$. Let $\gamma = U_X^\top \mathbf{1}_n$ and $S \in \mathbb{R}^{n \times p}$ be defined by $S_{ij} = 1 / (\lambda + D_X^i D_Y^j)$, where D_X^i (resp. D_Y^i) denotes the i -th diagonal term of D_X (resp. D_Y). Then the solution (f^*, b^*) of (2) is given by

$$b^* = U_Y \text{Diag} \left(S^\top \gamma^{\circ 2} \right)^{-1} \left(S^\top \circ \left(U_Y^\top Z^\top U_X \right) \right) \gamma \quad (3)$$

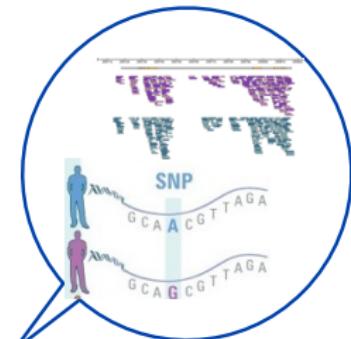
and

$$\forall (x, y) \in \mathcal{X} \times \mathcal{Y}, \quad f^*(x, y) = \sum_{i=1}^n \sum_{j=1}^p \alpha_{i,j}^* K_X(x_i, x) K_Y(y_i, y), \quad (4)$$

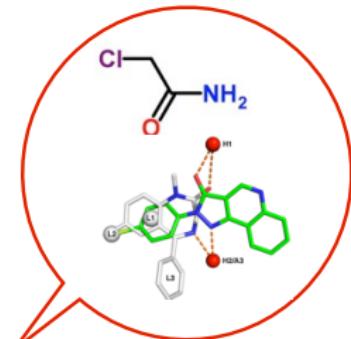
where

$$\alpha^* = U_X \left(S \circ \left(U_X^\top \left(Z - \mathbf{1}_n b^{*\top} \right) U_Y \right) \right) U_Y^\top. \quad (5)$$

Kernel Trick

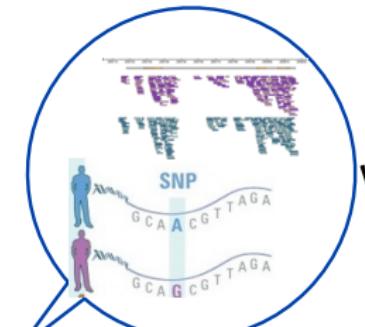


cell line descriptors

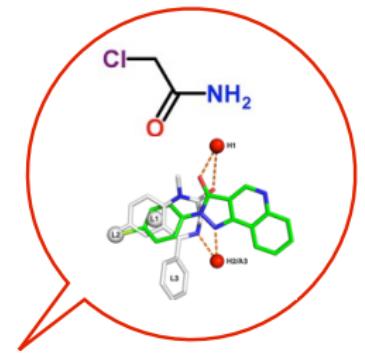


drug descriptors

Kernel Trick

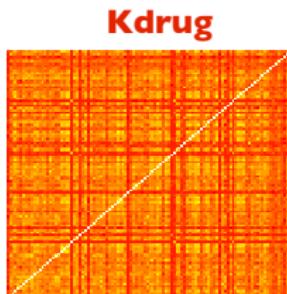
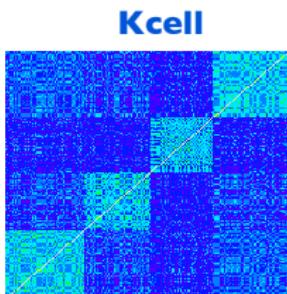


cell line descriptors

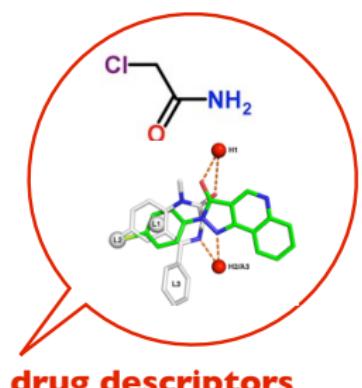
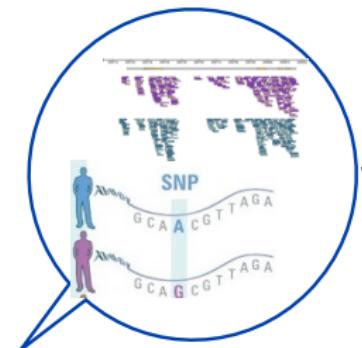


drug descriptors

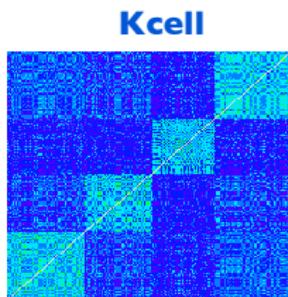
kernelized →



Kernel Trick



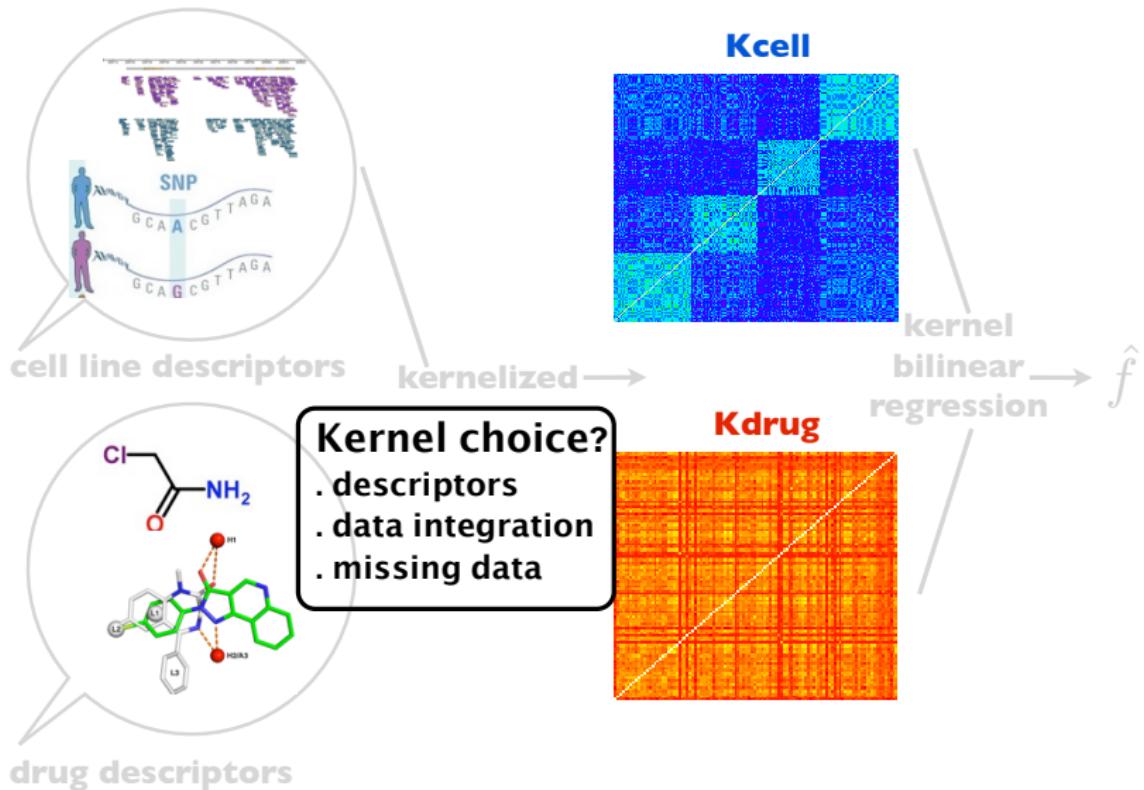
kernelized →



kernel bilinear regression → \hat{f}



Kernel Trick



Kernel choice

① K_{cell} :

- ⇒ 29 cell line kernels tested
- ⇒ 1 kernel that *integrate all information*
- ⇒ deal with missing data

② K_{drug} :

- ⇒ 48 drug kernels tested
- ⇒ multi-task kernels

Kernel choice

① K_{cell} :

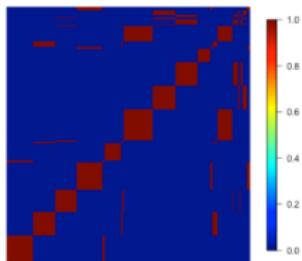
- ⇒ 29 cell line kernels tested
- ⇒ 1 kernel that *integrate all information*
- ⇒ deal with missing data

② K_{drug} :

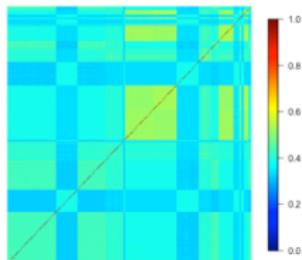
- ⇒ 48 drug kernels tested
- ⇒ multi-task kernels

Cell line data integration

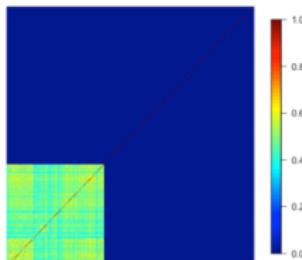
Covariates
. linear kernel



SNPs
. 10 gaussian kernels

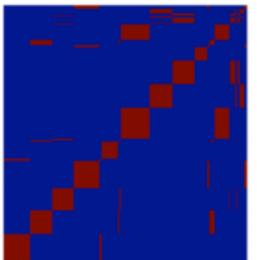


RNA-seq
. 10 gaussian kernels

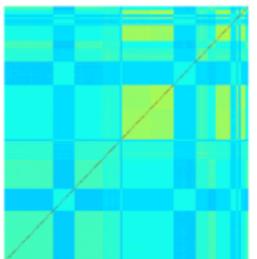


Cell line data integration

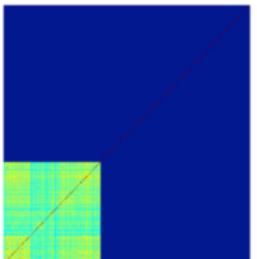
Covariates
. linear kernel



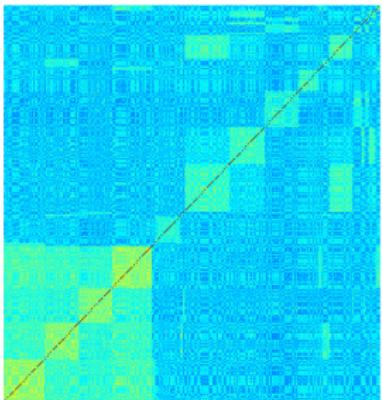
SNPs
. 10 gaussian kernels



RNA-seq
. 10 gaussian kernels

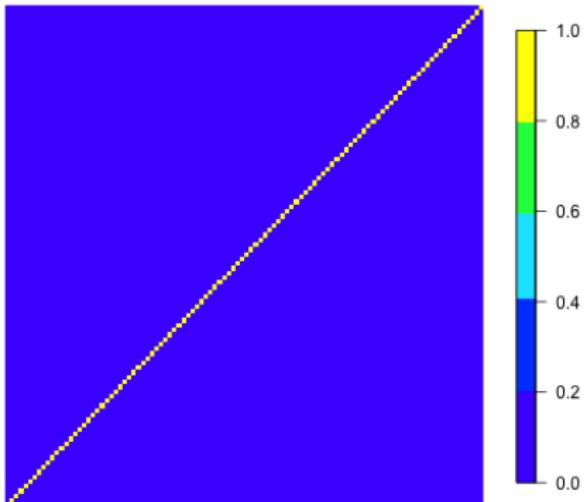


Integrated kernel



Multi-task drug kernels

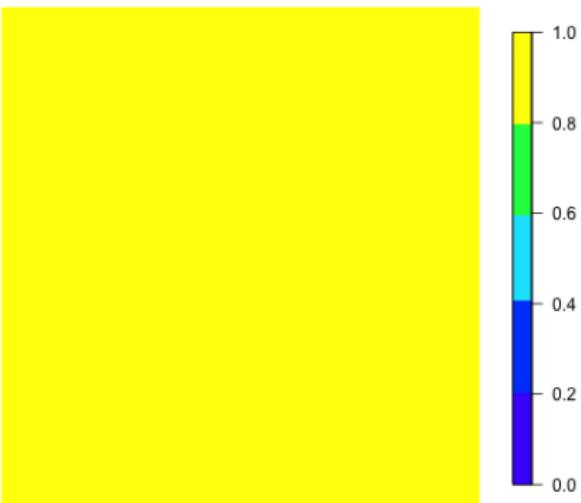
- ① **Dirac**
- ② Multi-Task
- ③ Feature-based
- ④ Empirical
- ⑤ Integrated



independent regression for each drug

Multi-task drug kernels

- ① Dirac
- ② **Multi-Task**
- ③ Feature-based
- ④ Empirical
- ⑤ Integrated



sharing information across drugs

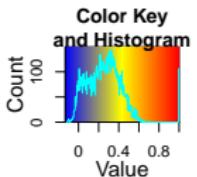
Multi-task drug kernels

- ① Dirac
- ② Multi-Task
- ③ **Feature-based**
- ④ Empirical
- ⑤ Integrated

Linear kernel and 10 gaussian kernels based on features:

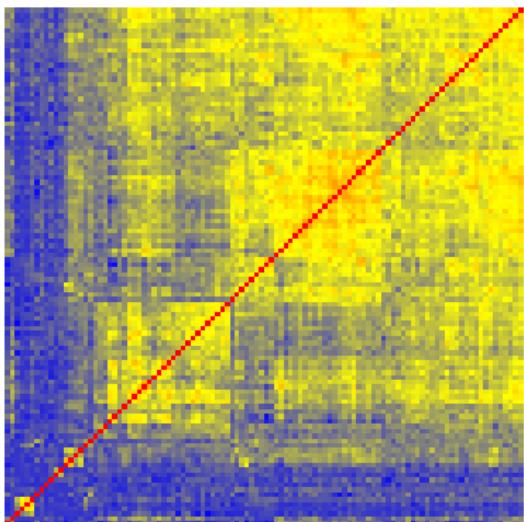
- CDK (160 descriptors) and SIRMS (9272 descriptors)
- Graph kernel for molecules (2D walk kernel)
- Fingerprint of 2D substructures (881 descriptors)
- Ability to bind human proteins (1554 descriptors)

Multi-task drug kernels



Empirical correlation

- ① Dirac
- ② Multi-Task
- ③ Feature-based
- ④ **Empirical**
- ⑤ Integrated



Multi-task drug kernels

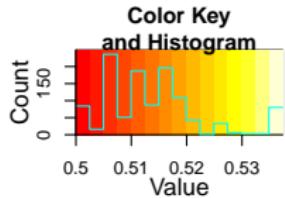
- ① Dirac
- ② Multi-Task
- ③ Feature-based
- ④ Empirical
- ⑤ **Integrated**

$$K_{int} = \sum_i K_i$$

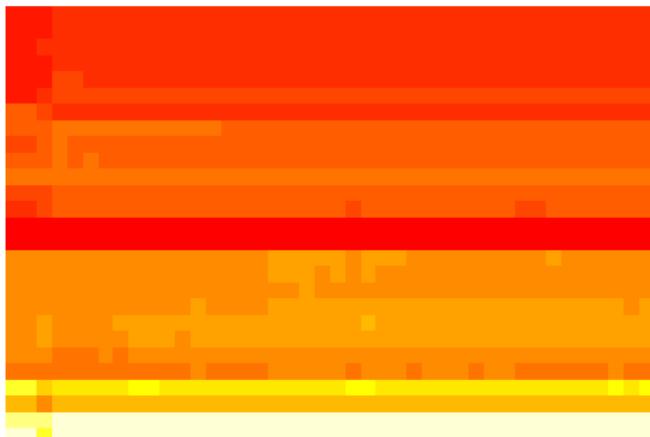
Integrated kernel:

- Combine all information on drugs

29x48 kernel combinations: CV results

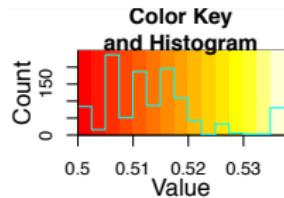


CI

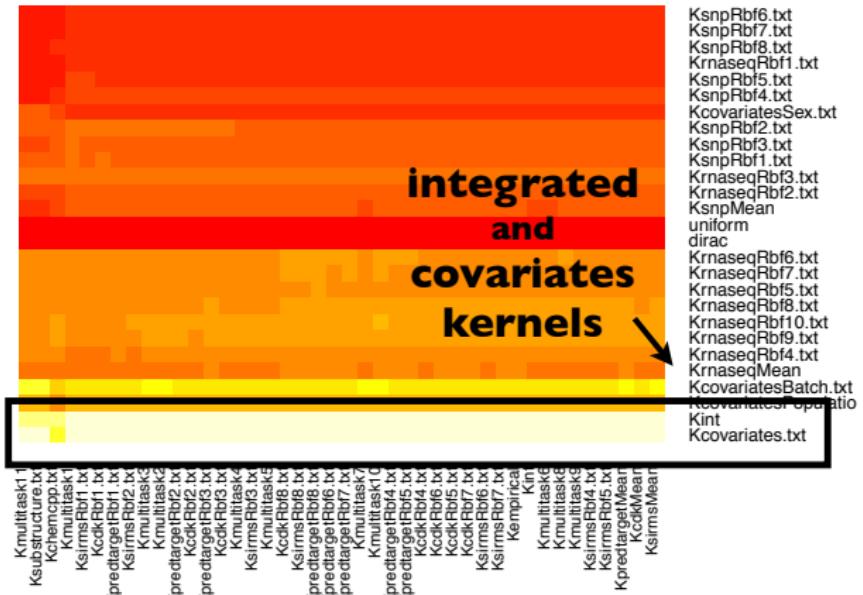


Kmultitask11
Ksubstructure.txt
Kchemerpp.txt
Kmultitask1
KsimrsRbf1.txt
Kcdkrbf1.txt
KpredtargetRbf1.txt
KsimrsRbf2.txt
Kmultitask3
Kmultitask2
KpredtargetRbf2.txt
Kcdkrbf2.txt
KpredtargetRbf3.txt
Kcdkrbf3.txt
Kmultitask4
KsimrsRbf3.txt
Kmultitask5
Kcdkrbf5.txt
KsimrsRbf8.txt
KpredtargetRbf8.txt
Kcdkrbf8.txt
KpredtargetRbf10.txt
Kcdkrbf10.txt
Kmultitask7
KpredtargetRbf11.txt
Kcdkrbf11.txt
KpredtargetRbf12.txt
Kcdkrbf12.txt
Kmultitask9
KsimrsRbf12.txt
KpredtargetRbf12.txt
Kcdkrbf12.txt
KpredtargetRbf13.txt
Kcdkrbf13.txt
Kmultitask10
KpredtargetRbf14.txt
Kcdkrbf14.txt
KpredtargetRbf15.txt
Kcdkrbf15.txt
KpredtargetRbf16.txt
Kcdkrbf16.txt
KpredtargetRbf17.txt
Kcdkrbf17.txt
Kmultitask15
KsimrsRbf17.txt
Kempirical
Kint
Kmultitask6
Kmultitask8
Kmultitask9
KsimrsRbf14.txt
KsimrsRbf15.txt
KpredtargetMean
KcdkrMean
KsimrsMean

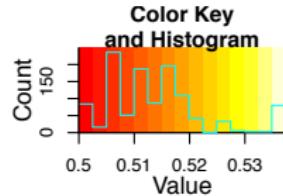
29x48 kernel combinations: CV results



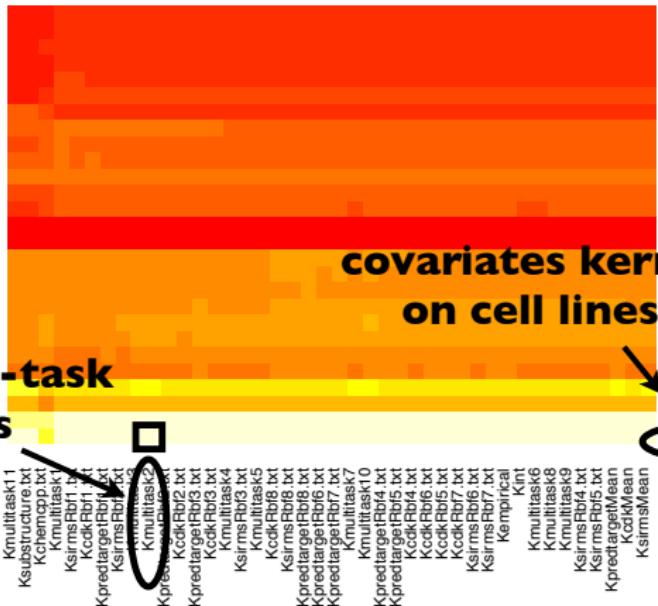
CI



29x48 kernel combinations: CV results



CI

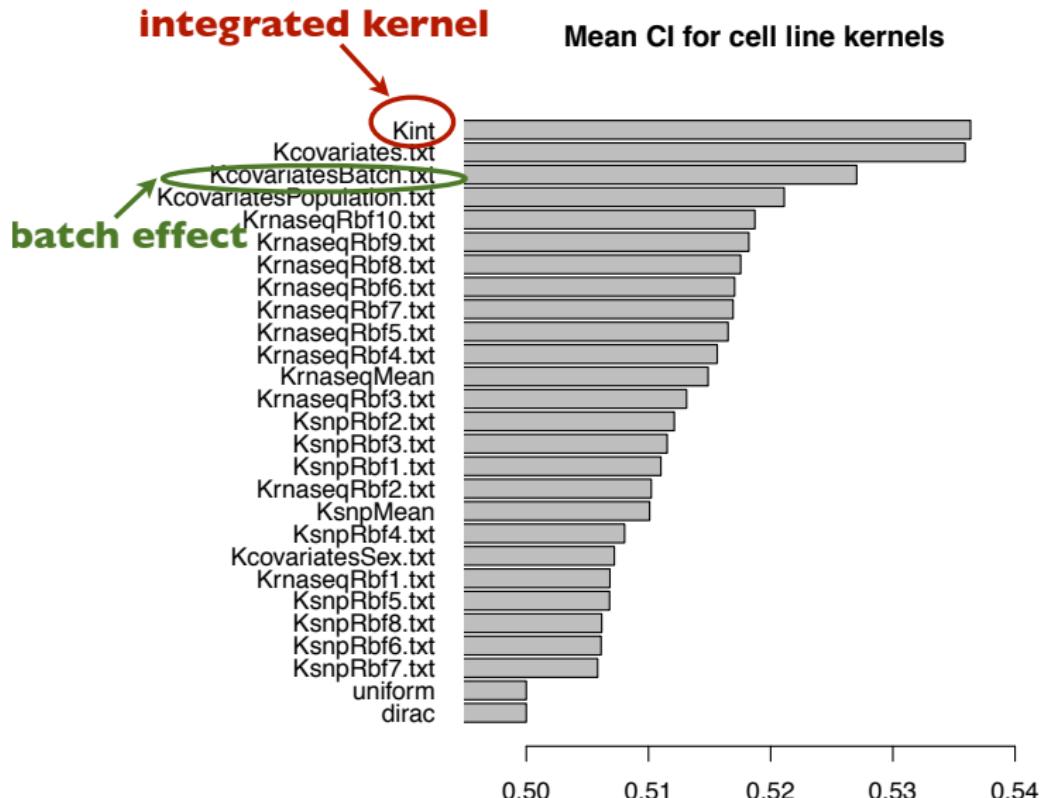


sightly multi-task on drugs

covariates kernel on cell lines

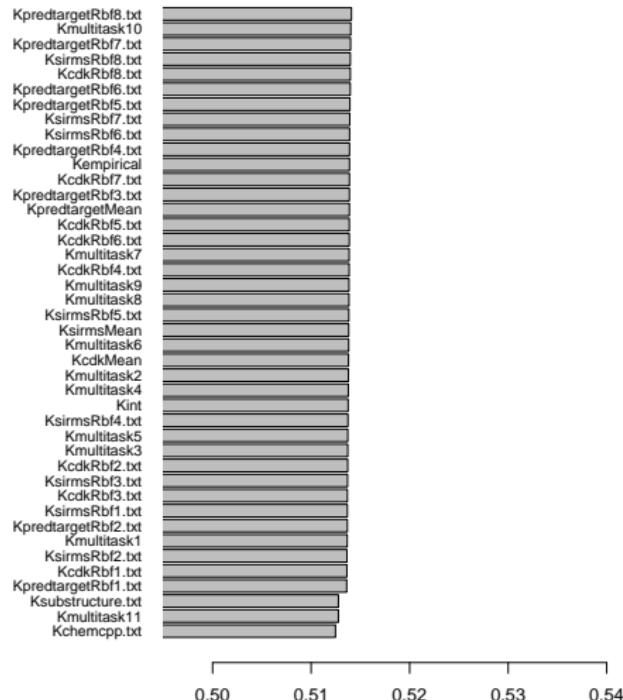
KsnP Rbf6.txt
KsnP Rbf7.txt
KsnP Rbf8.txt
KrnaseqRbf1.txt
KsnP Rbf5.txt
KsnP Rbf4.txt
KcovariatesSex.txt
KsnP Rbf2.txt
KsnP Rbf3.txt
KsnP Rbf1.txt
KrnaseqRbf3.txt
KrnaseqRbf2.txt
KsnP Mean
uniform
dirac
KrnaseqRbf6.txt
KrnaseqRbf7.txt
KrnaseqRbf5.txt
KrnaseqRbf8.txt
KrnaseqRbf10.txt
KrnaseqRbf9.txt
KrnaseqRbf4.txt
KrnaseqMean
KcovariatesBatch.txt
KcovariatesPopulatio
Kcovariates.txt

Kernel on cell lines: CV results



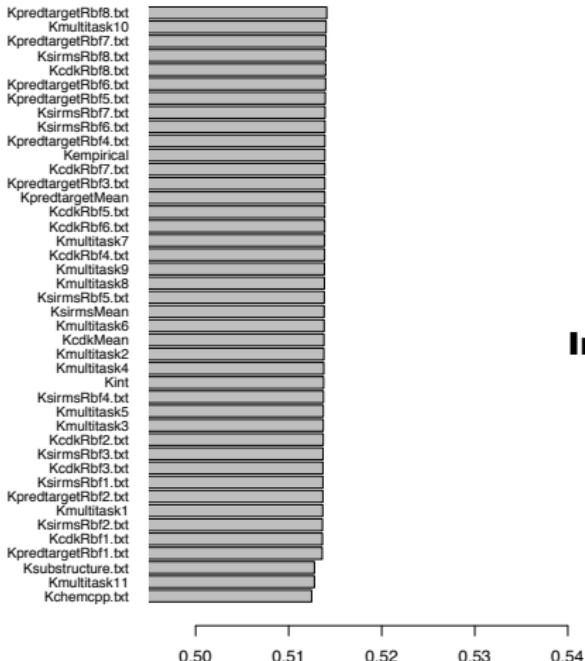
Kernel on drugs: CV results

Mean CI for chemicals kernels

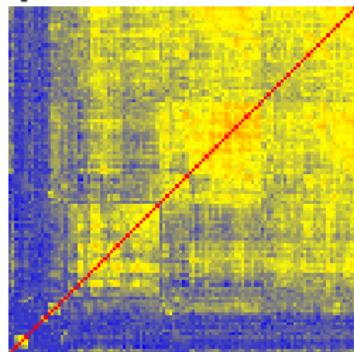


Final Submission (ranked 2nd)

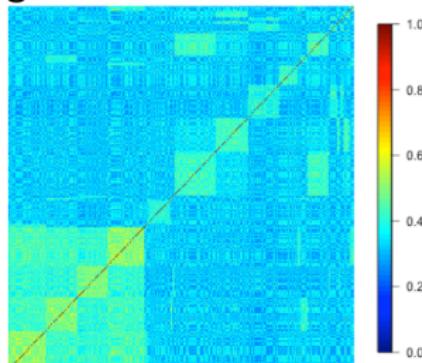
Mean CI for chemicals kernels



Empirical kernel on drugs



Integrated kernel on cell lines



Thanks

Alexandre d'Aspremont, Emmanuel Barillot, Anne-Claire Haury,
Laurent Jacob, Pierre Mahé, Guillaume Obozinski, Franck Rapaport,
Jean-Baptiste Veyrieras, Andrei Zynoviev, ... and all CBIO



European Research Council

