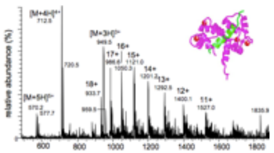
Structured feature selection

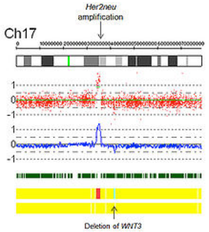
Jean-Philippe Vert

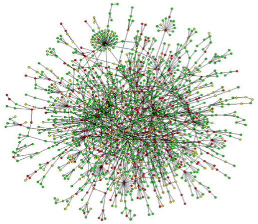
Jan 27, 2015

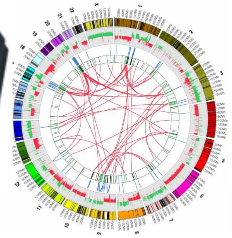
CBIO at work

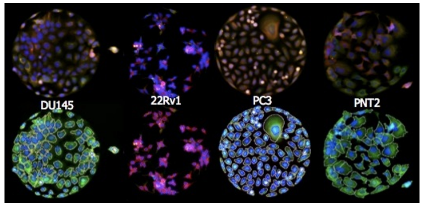
Rationale of the team

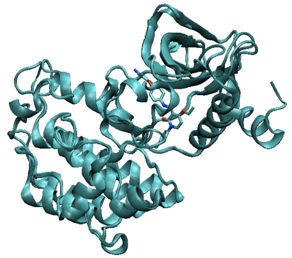


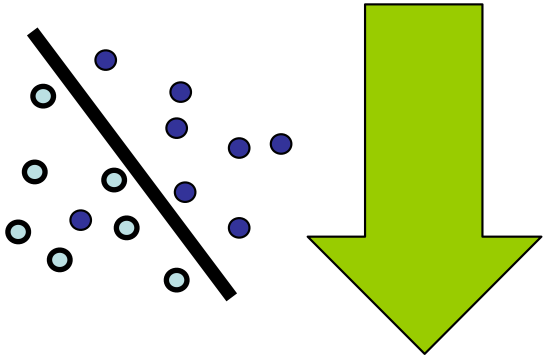




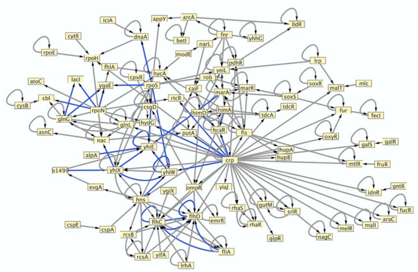




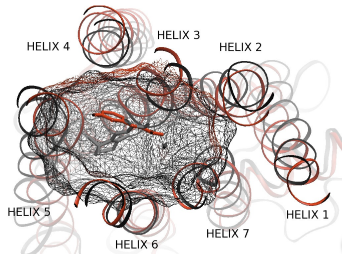




Machine learning



Mecanisms,
drug targets



Drug design

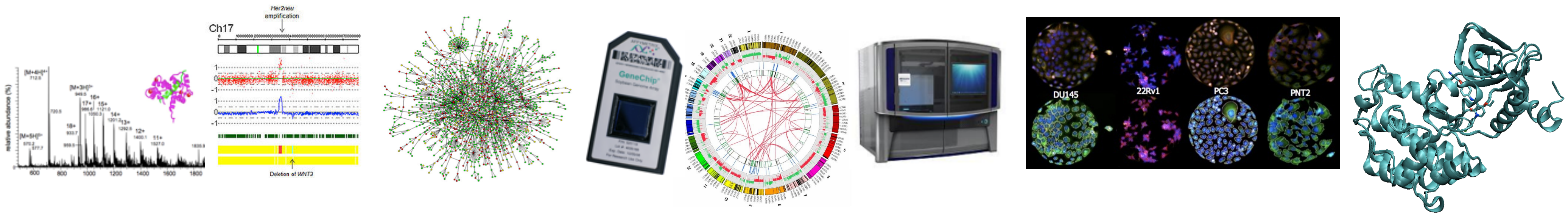
Personalized
medicine

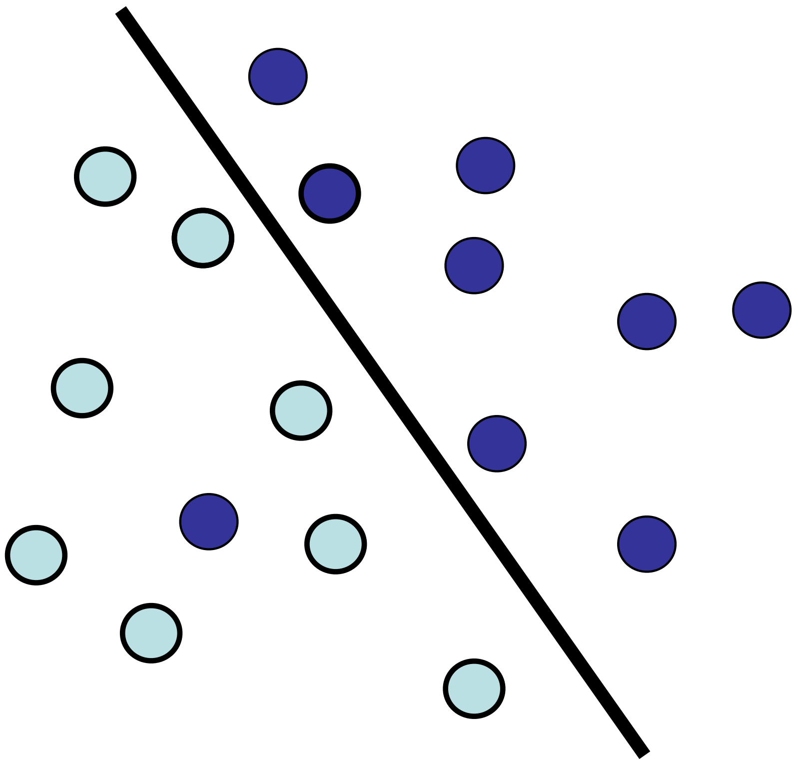
institut Curie

Instituts thématiques Inserm
Institut national de la santé et de la recherche médicale

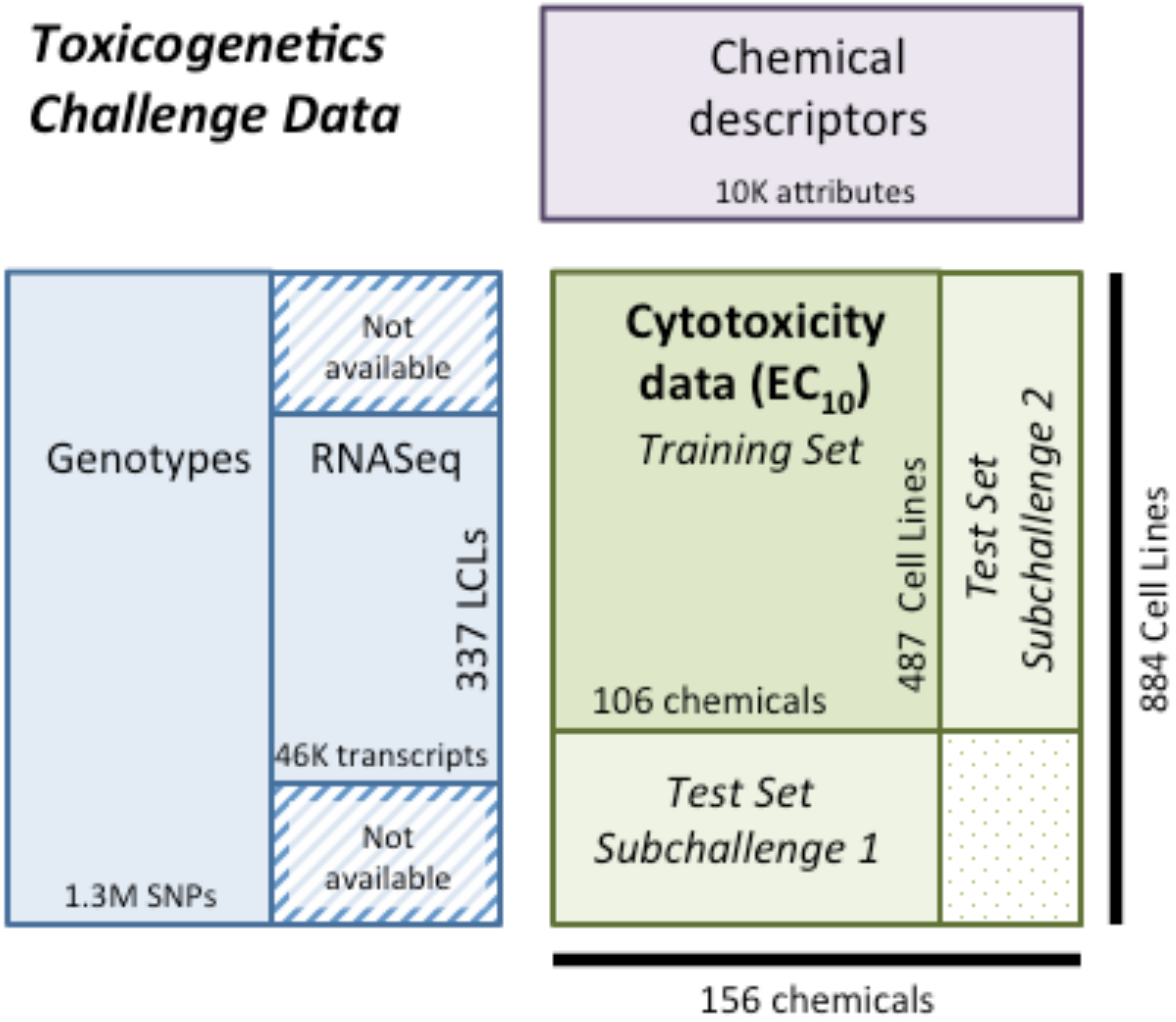
MINES ParisTech

Machine Learning?

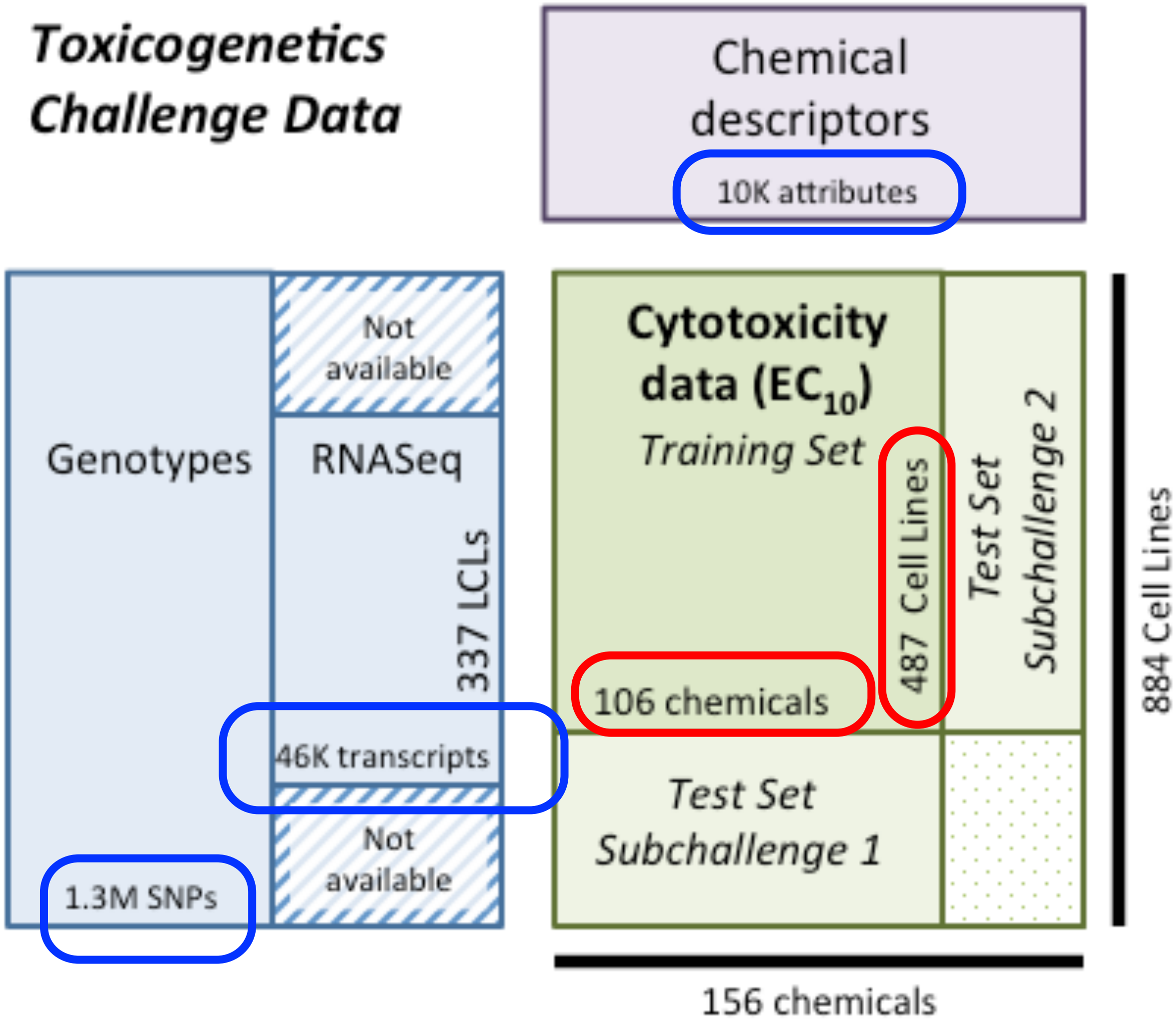




Example: Toxicogenetics / Pharmacogenomics



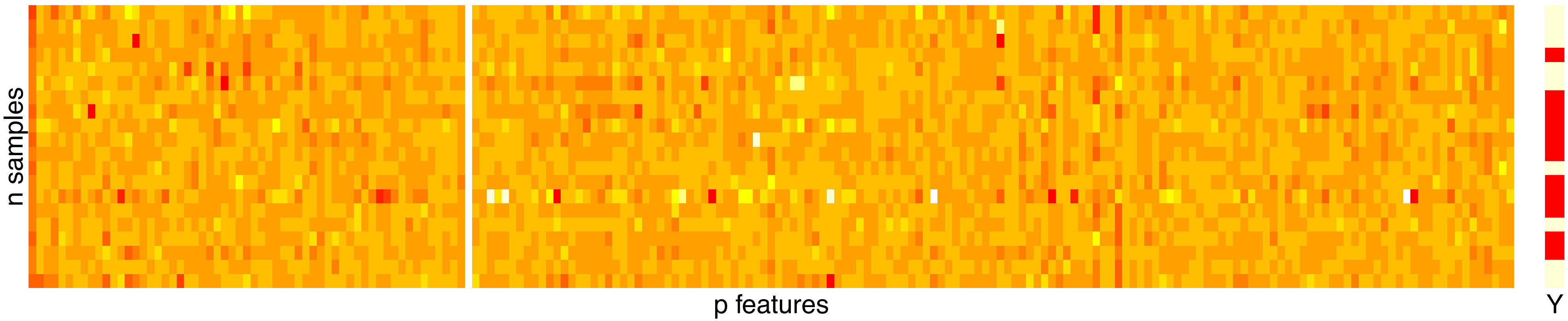
Problem: n << p



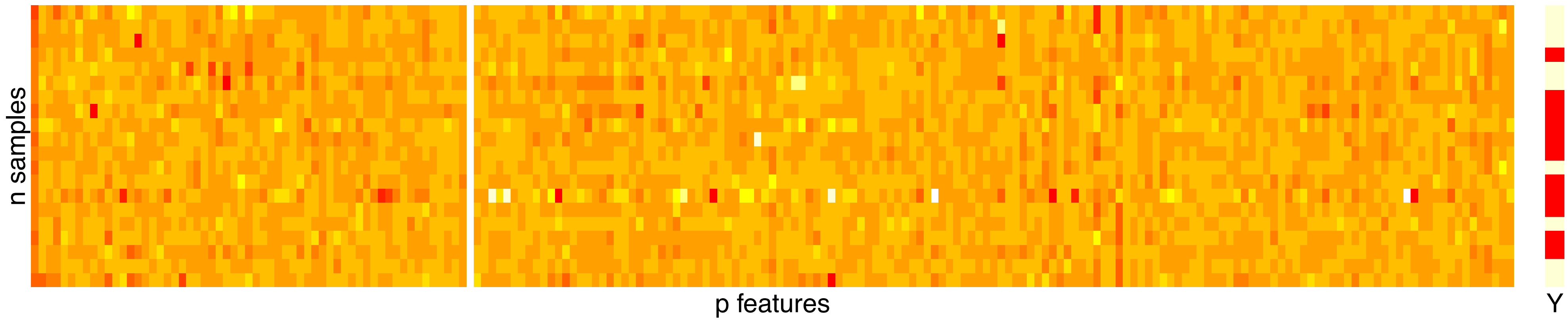
n = 5E4

p = 1E10

Example: Patient stratification



Problem again: n << p

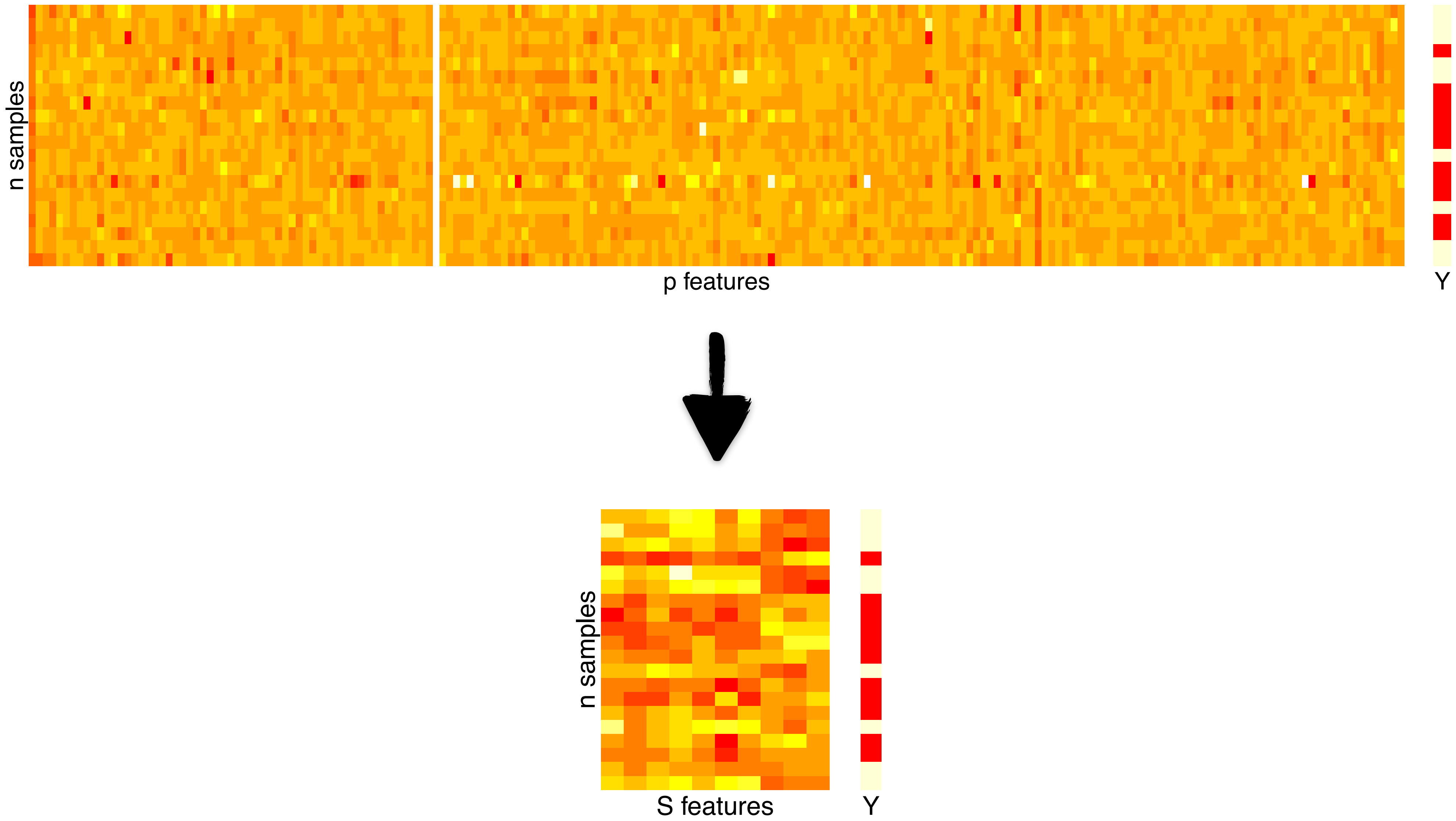


n = 1E2 ~ 1E4
(patients)

p = 1E4 ~ 1E7
(genes, mutations, copy numbers, …)

institutCurie Inserm MINES ParisTech

Feature Selection



Feature Selection techniques

1) **Filter methods**: test association between features and response one by one (eg: correlation, t-test, …)

2) **Wrapper methods**: search a subset of features such that the classifier works well (best subset selection, forward search, recursive feature elimination…)

3) **Embedded methods**: directly optimize sparse models (eg: lasso, elastic net, …)

But…

Gene expression profiling predicts clinical outcome of breast cancer

Laura J. van 't Veer*†, Hongyue Dai†‡, Marc J. van de Vijver*†, Yudong D. He‡, Augustinus A. M. Hart*, Mao Mao‡, Hans L. Peterse*, Karin van der Kooy*, Matthew J. Marton‡, Anke T. Witteveen*, George J. Schreiber‡, Ron M. Kerkhoven*, Chris Roberts‡, Peter S. Linsley‡, René Bernards* & Stephen H. Friend‡

70 genes (Nature, 2002)

Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer

Yixin Wang, Jan G M Klijn, Yi Zhang, Anieta M Sieuwerts, Maxime P Look, Fei Yang, Dmitri Talantov, Mieke Timmermans, Marion E Meijer-van Gelder, Jack Yu, Tim Jatkoe, Els M J J Berns, David Atkins, John A Foekens

76 genes (Lancet, 2005)

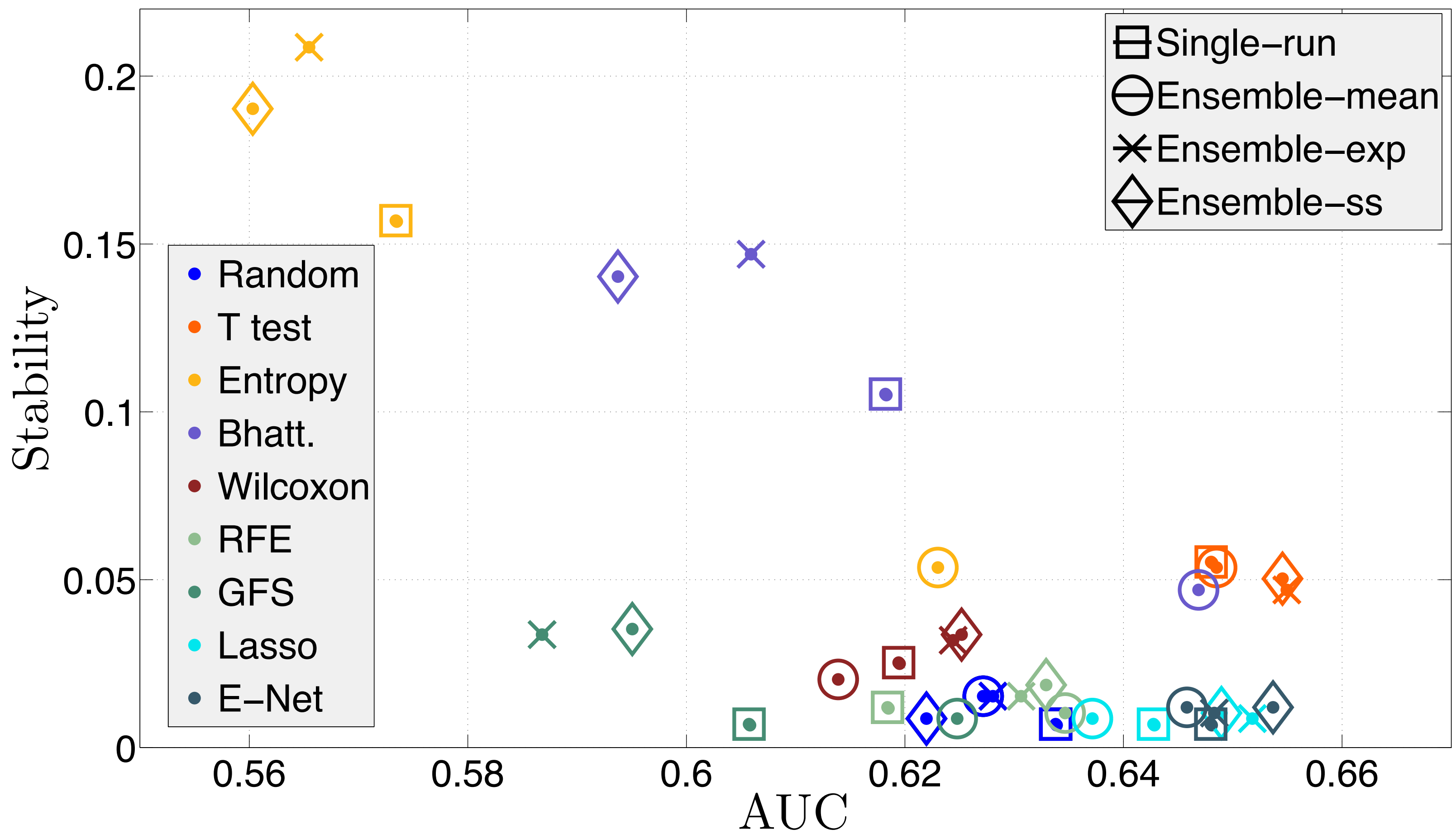
3 genes in common

institutCurie

Instituts thématiques Inserm
Institut national de la santé et de la recherche médicale

MINES ParisTech

and nothing seems to work better



(Haury et al., 2011)

Give up machine learning and go to Tahiti?

institut Curie

Instituts thématiques Inserm

Institut national de la santé et de la recherche médicale

MINES ParisTech

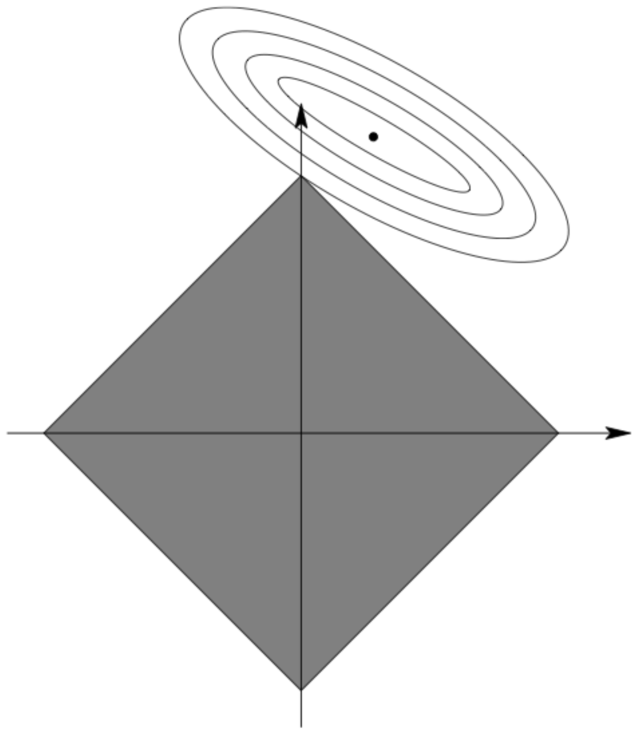
Sparsity with the LASSO

- Linear model
$f(x) = w1\ x1 + w2\ x2 + \ldots + wP\ xP$

- Sparse when wK=0 for many K's

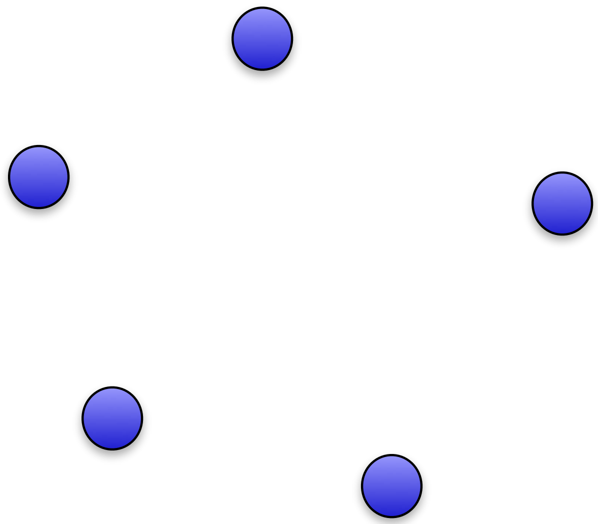
- Learn a sparse model by
minimize Error(w)
such that
w is in the grey box O

- O is convex -> efficient algorithm
- O has edges -> sparsity



Structured sparsity with atomic norms

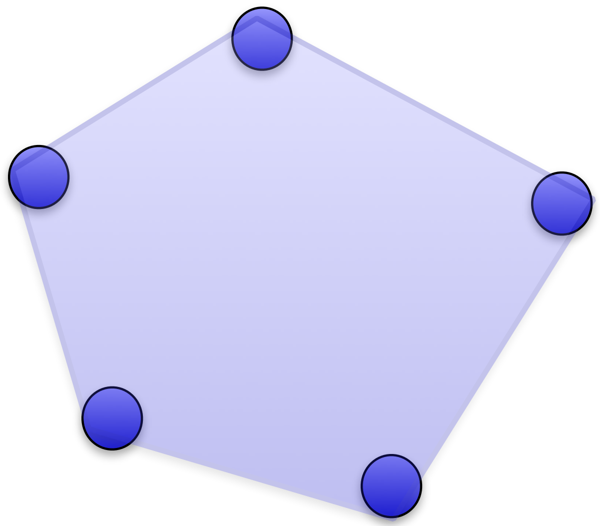
1) Choose a set of ATOMS



Structured sparsity with atomic norms

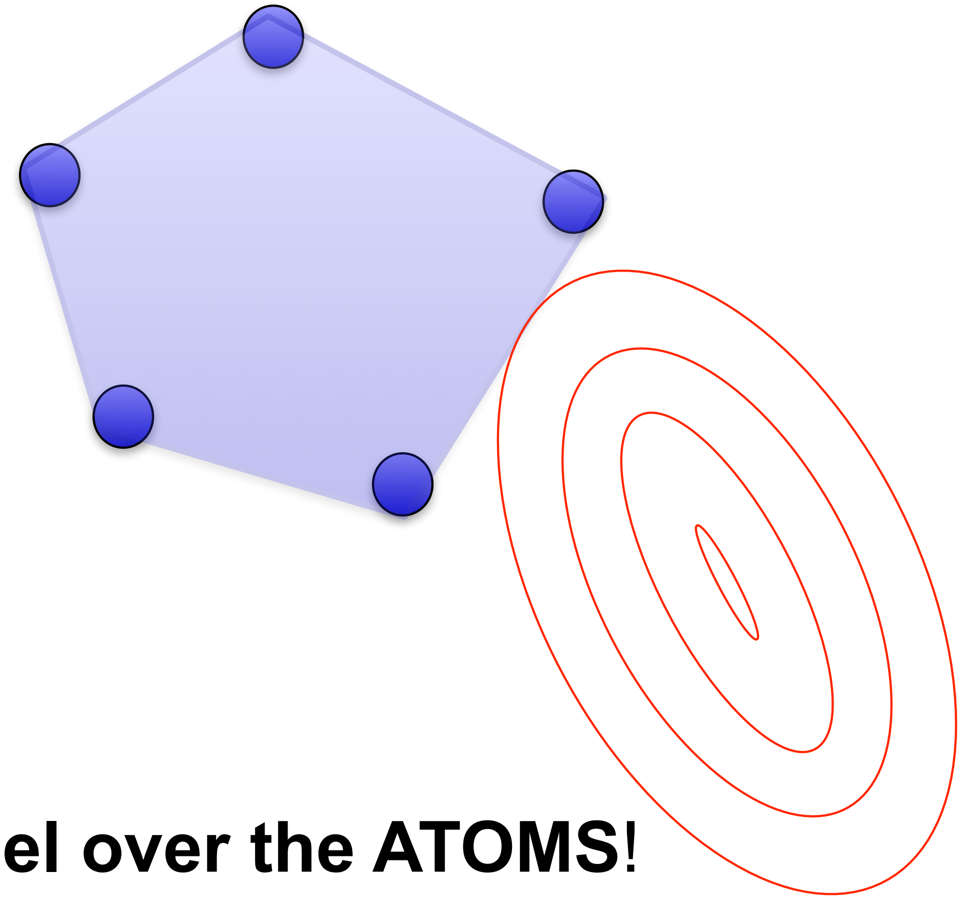
1) Choose a set of ATOMS

2) Take the convex hull O



Structured sparsity with atomic norms

1) Choose a set of ATOMS

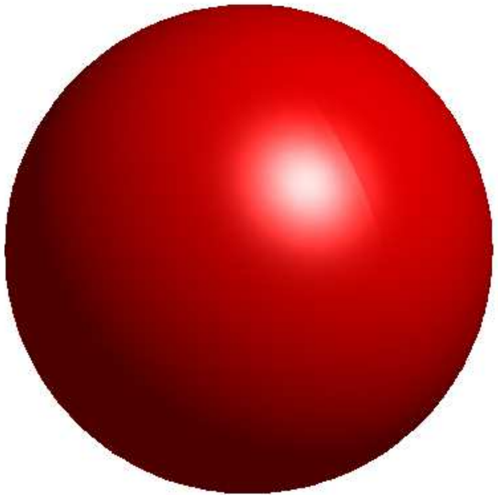


2) Take the convex hull

3) Minimize Error(w)
such that
w is in the convex hull

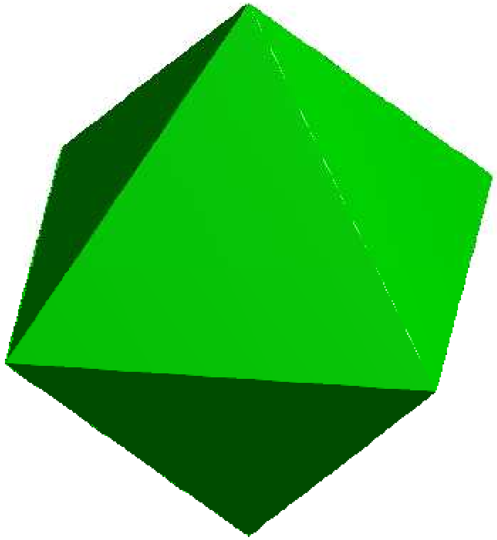
The solution is a **sparse model over the ATOMS**!

Quizz: where are the atoms?



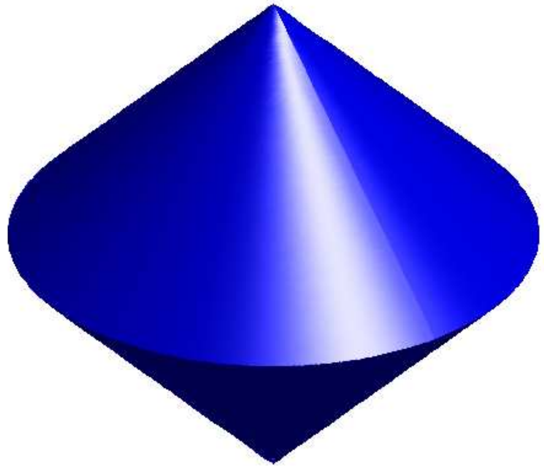
$\|w\|_2$

Ridge



$\|w\|_1$

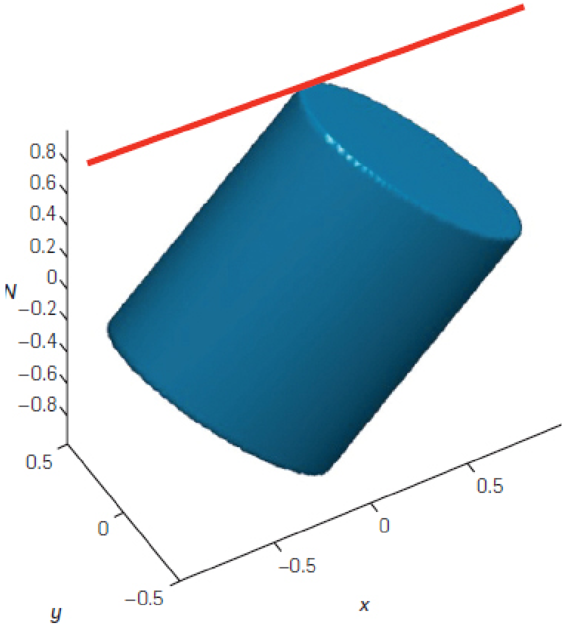
Lasso



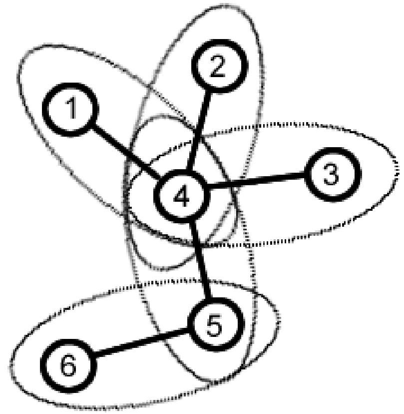
$\sqrt{w_1^2 + w_2^2} + |w_3|$

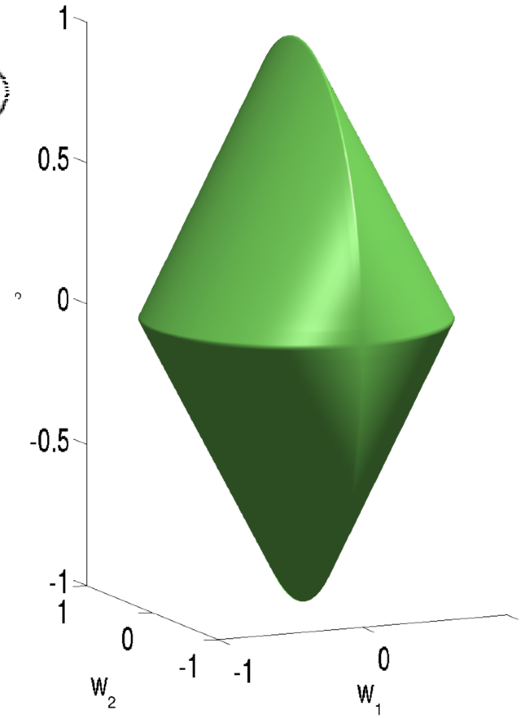
Group Lasso

Quizz (cont.)



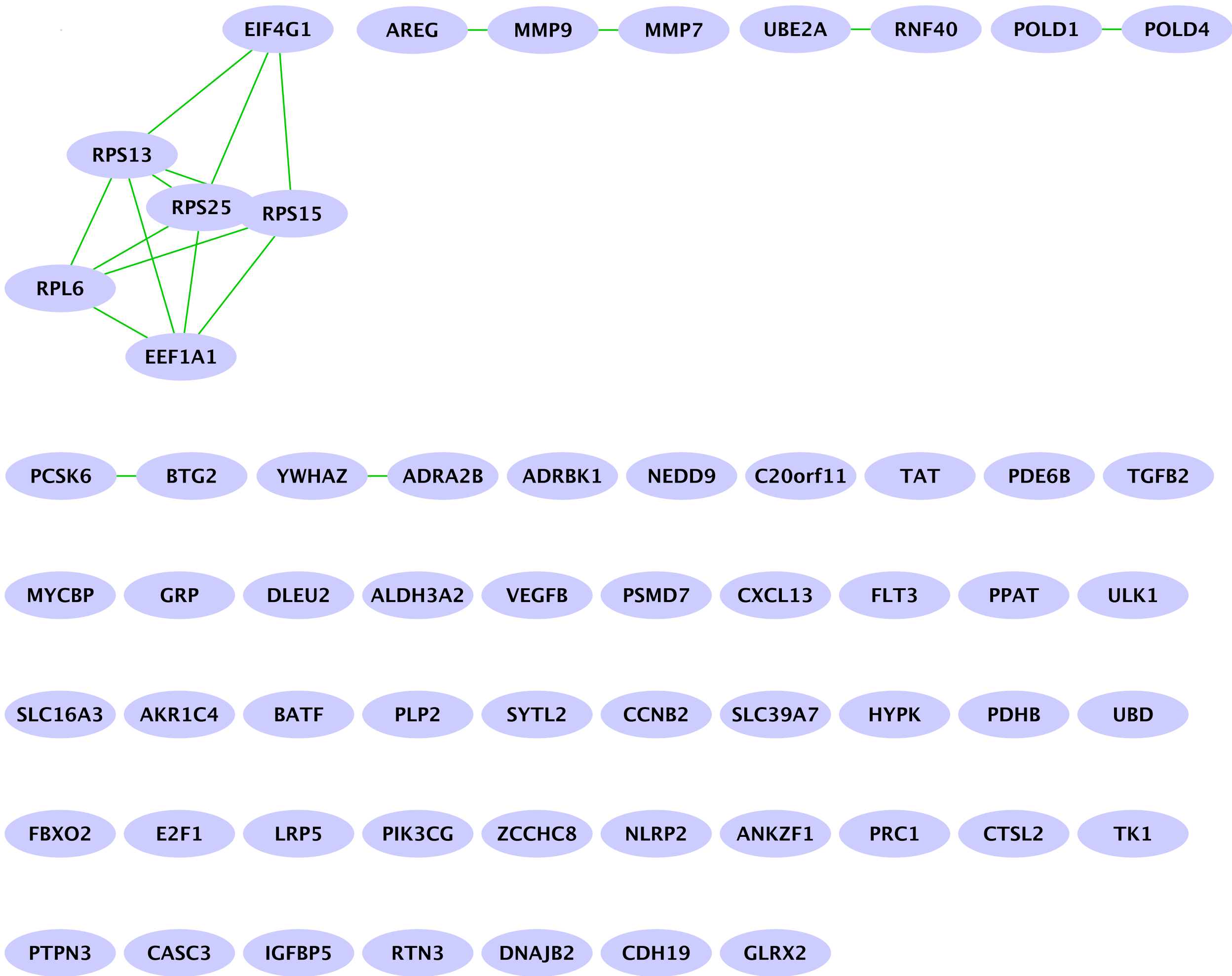
Trace norm
to learn matrices with small rank



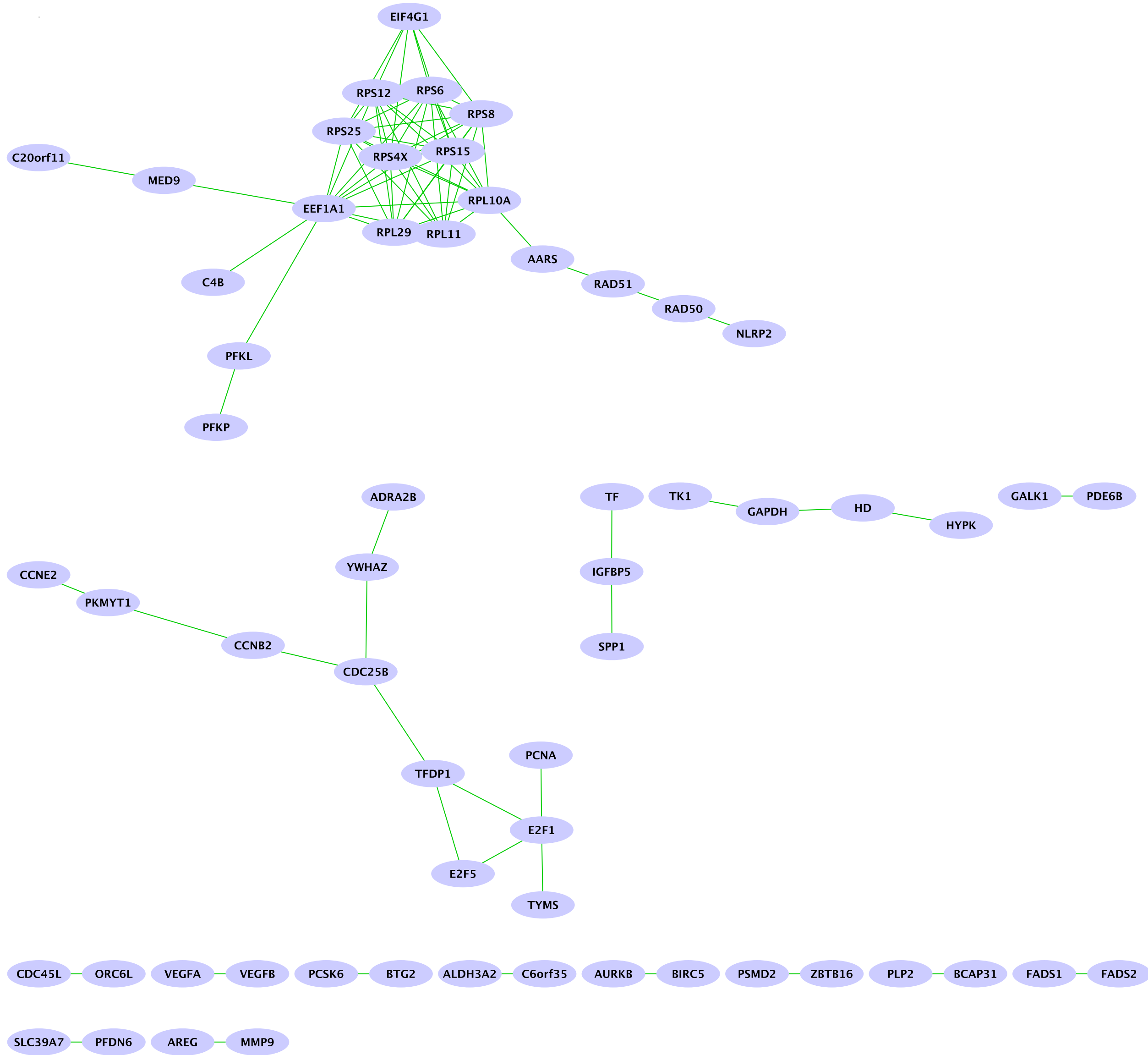


Graph Lasso (Jacob et al. 2009)
to select features that tend to be connected over a given network

Breast cancer prognosis signature with Lasso (accuracy=61%)



Breast cancer prognosis signature with Graph Lasso (accuracy=64%)



institutCurie

Instituts thématiques Inserm
Institut national de la santé et de la recherche médicale

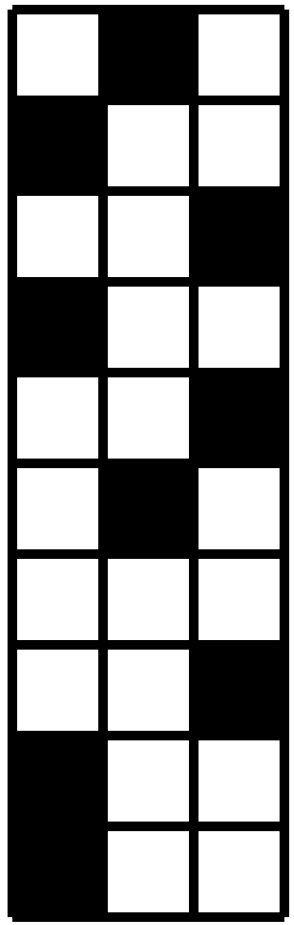
MINES ParisTech

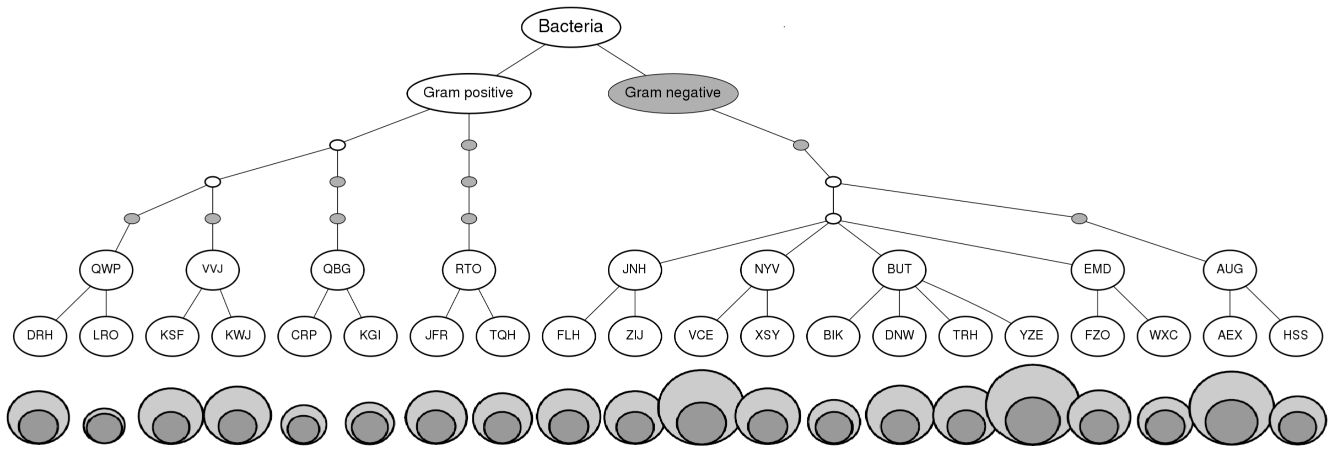
Learning sparse models with disjoint support ?

Motivation

- Multiclass or multi-task classification problems
- Eg: predict identity and emotion from a face
- Eg: cascade of classifiers

$X =$





institut Curie

Instituts thématiques Inserm

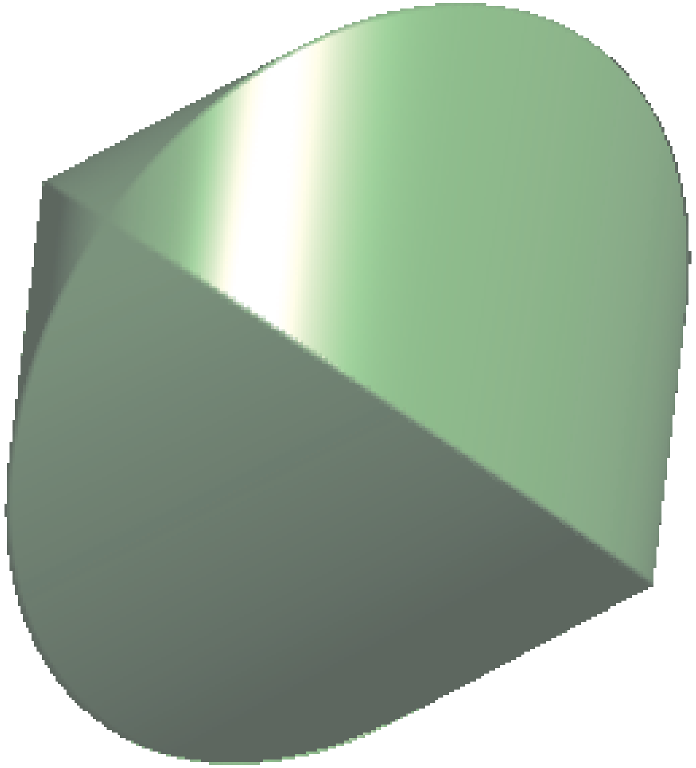
Institut national de la santé et de la recherche médicale

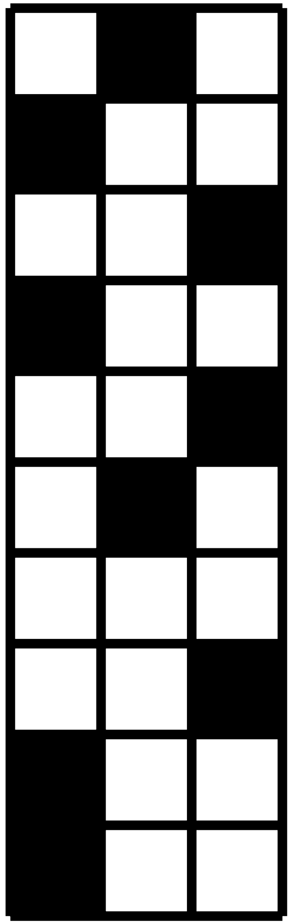
MINES ParisTech

An atomic norm (ECML 2014)

K. Vervier

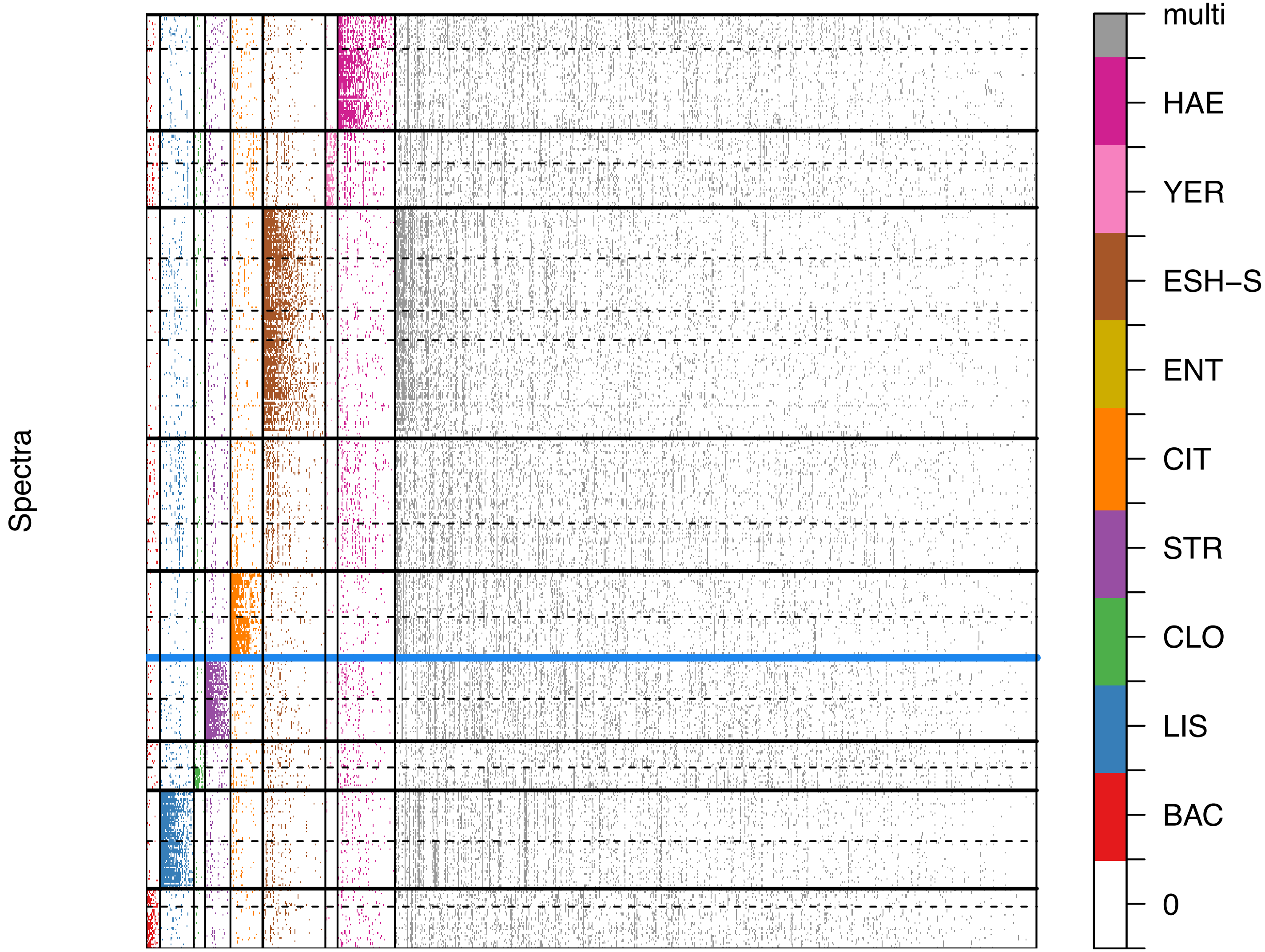
A. d'Aspremont



$X =$ 

$$\Omega_K(X) = \sum_{i=1}^{p} K_{ii} \| x_i \|^2 + \sum_{i \neq j} K_{ij} \left| x_i^\top x_j \right|$$

Application: Microbial identification from MALDI-TOF MS spectra



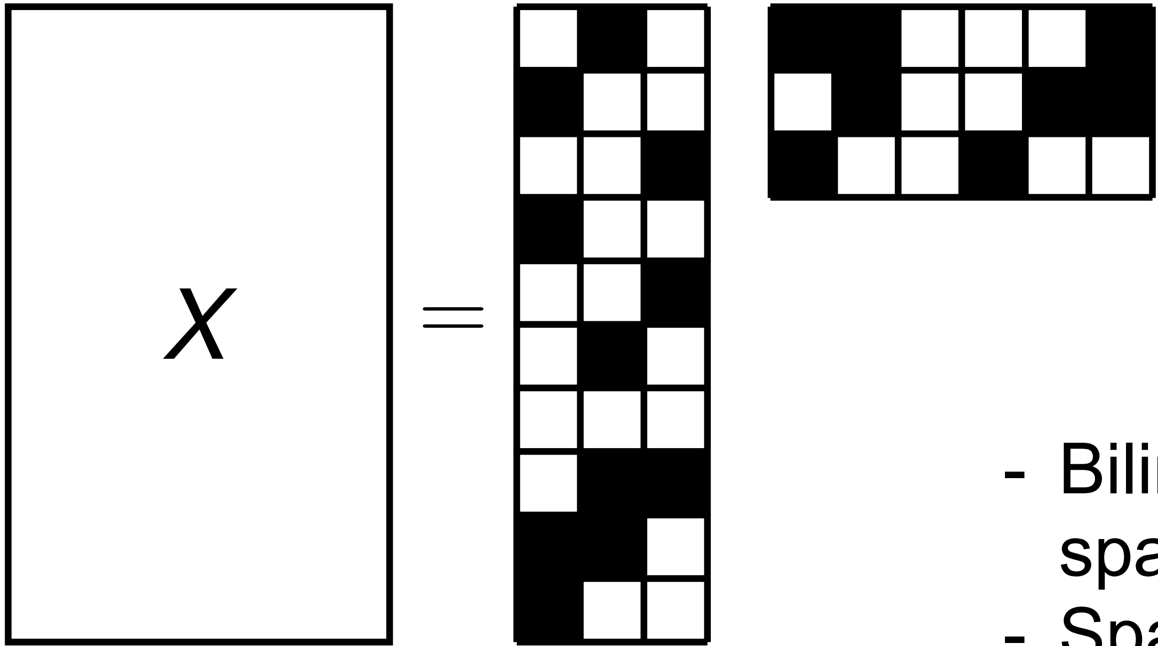
institut Curie

Instituts thématiques Inserm

Institut national de la santé et de la recherche médicale

MINES ParisTech

Learning low-rank matrices with sparse factors ?



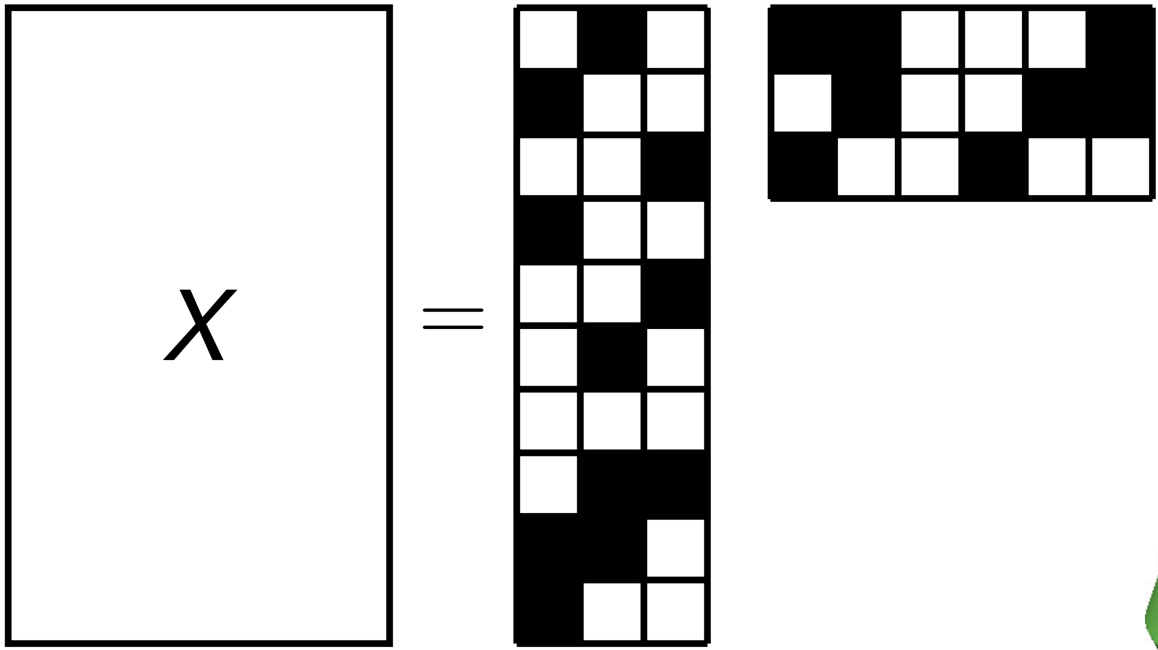
$$X = \sum_{i=1}^{r} u_i v_i^\top$$

- Bilinear regression with sparse latent factors
- Sparse PCA
- Sparse CCA
- Hidden clique problem
- Community detection in networks

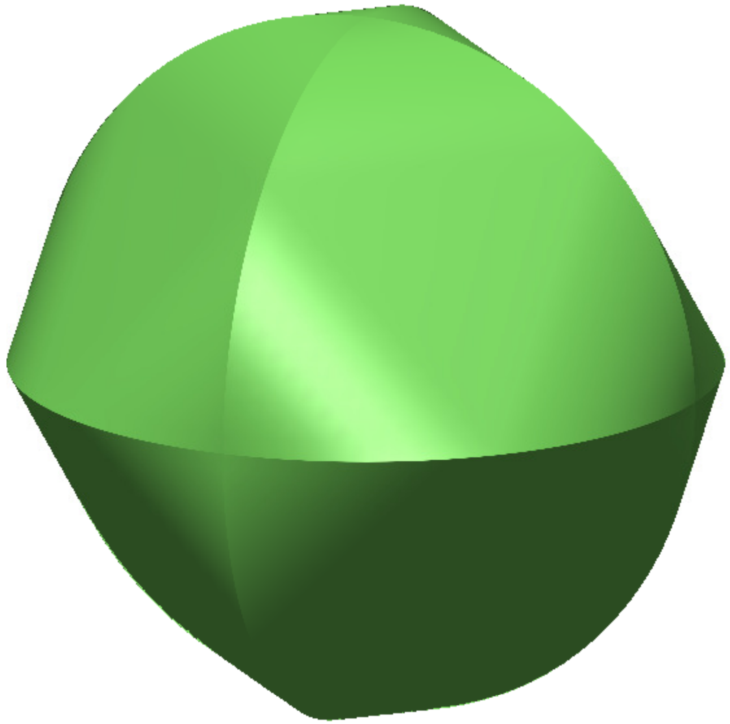
An atomic norm (NIPS 2014)

E. Richard

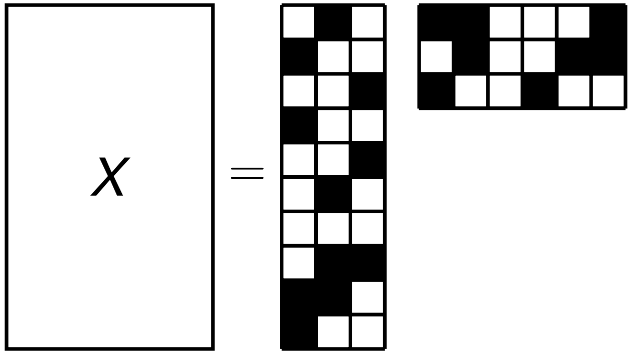
G. Obozinski



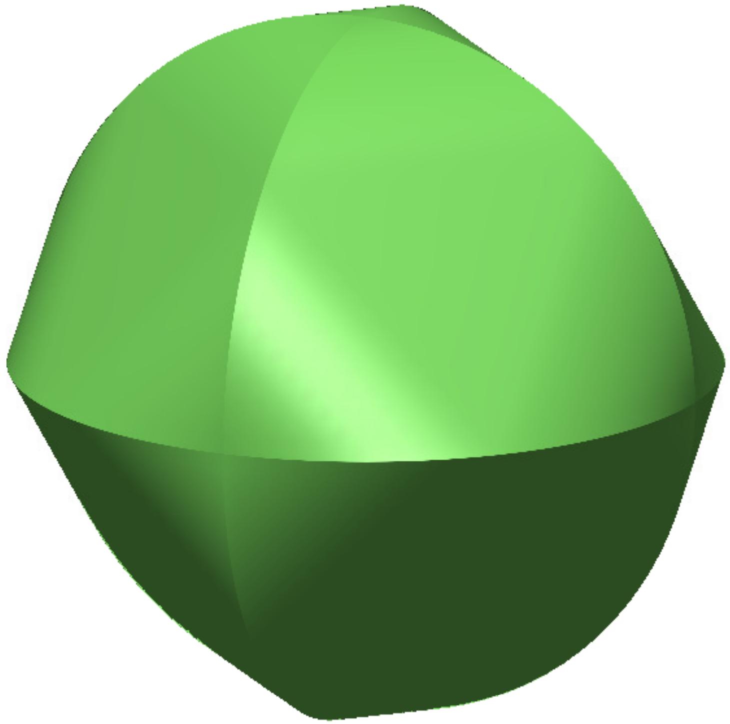
$$X = \sum_{i=1}^{r} u_i v_i^\top$$



An atomic norm (NIPS 2014)



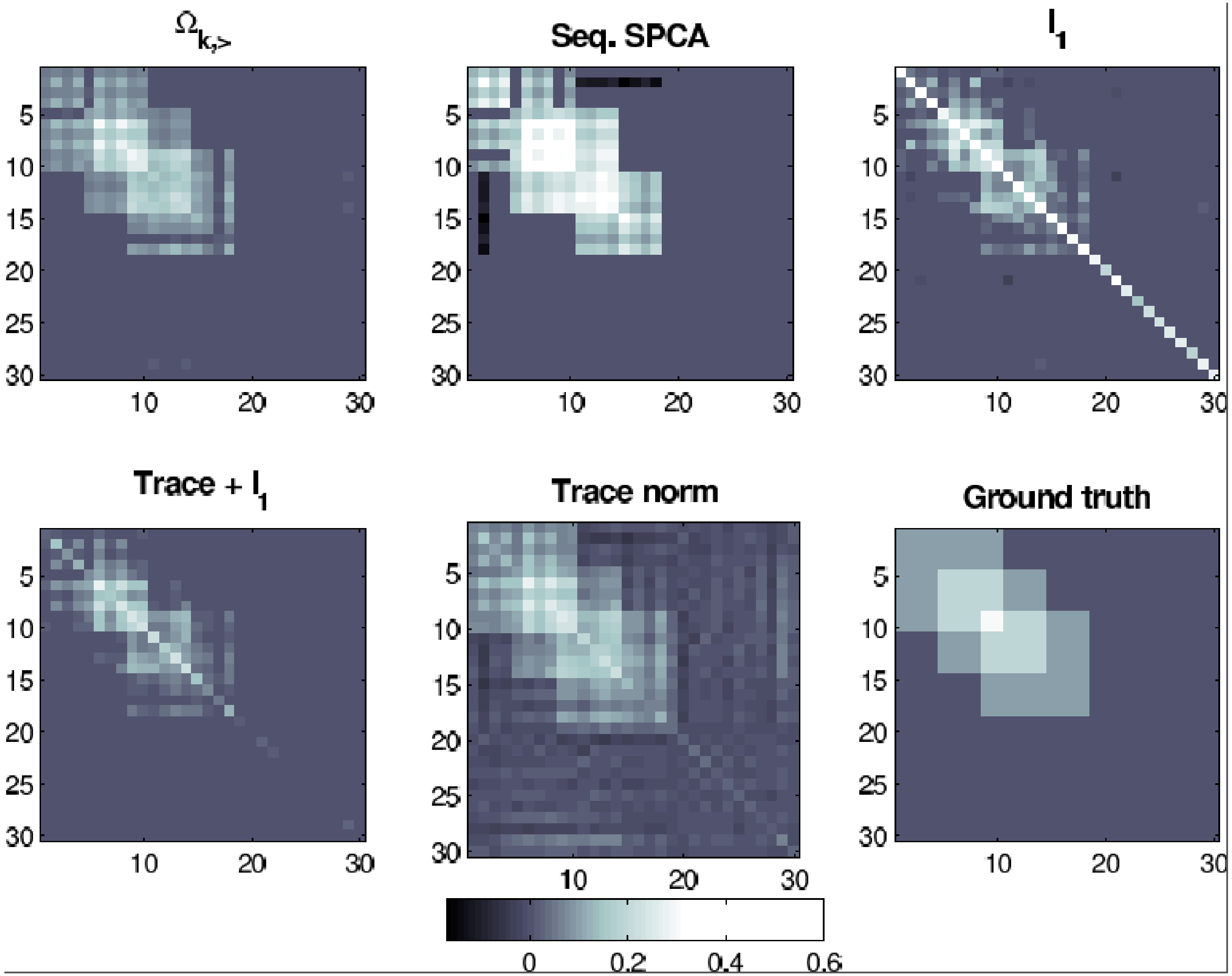
$$X = \sum_{i=1}^{r} u_i v_i^\top$$



Theorem
Learning with this norm is « statistically optimal » to infer sparse low-rank matrices

But
Convex but NP-hard

Preliminary results on sparse PCA



Sample covariance	Trace	ℓ_1	Trace + ℓ_1	Sequential	$\Omega_{k,\succeq}$
4.20 ± 0.02	0.98 ± 0.01	2.07 ± 0.01	0.96 ± 0.01	0.93 ± 0.08	$\mathbf{0.59 \pm 0.03}$

Conclusion
Make your Atomic norm !

http://www.homemade-gifts-made-easy.com/make-paper-lanterns.html

institutCurie

Instituts thématiques Inserm
Institut national de la santé et de la recherche médicale

MINES ParisTech