Plateforme de Bioinformatique

Jonathan Séguin

Geneviève Boucher

Outil de visualisation de profils d'expression de gènes

BioHackUdem '15

Problématique

Les outils pour analyser de manière exploratoire une collection de profils d’expression de gènes (quelques centaines d'échantillons) sont limités.

Projet

Visualisation interactive donnant un maximum d'information sur une collection de profils d'expression de gènes (RNA-Seq ou microarray).

Approches traditionelles

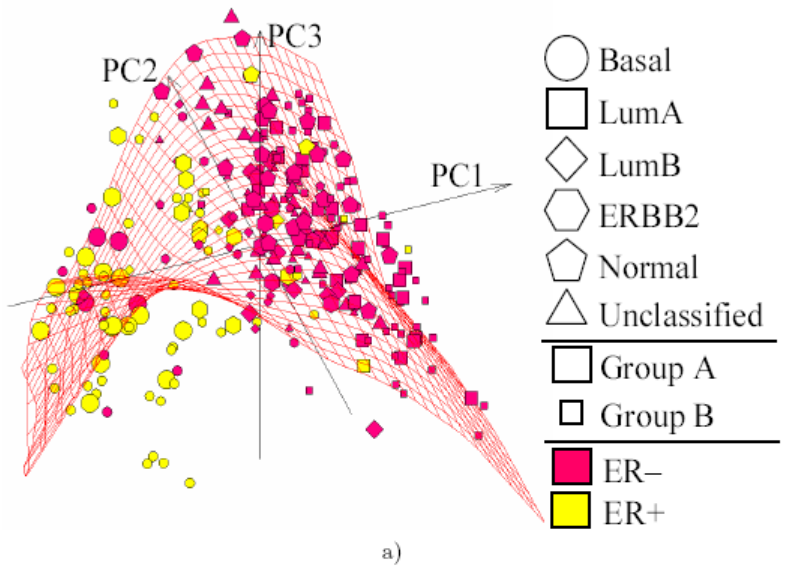
PCA / MDS

- Représentation lourde (20,000 points)
- Discrimination limitée en 2D
- Tendance à mal exploiter l'espace disponible

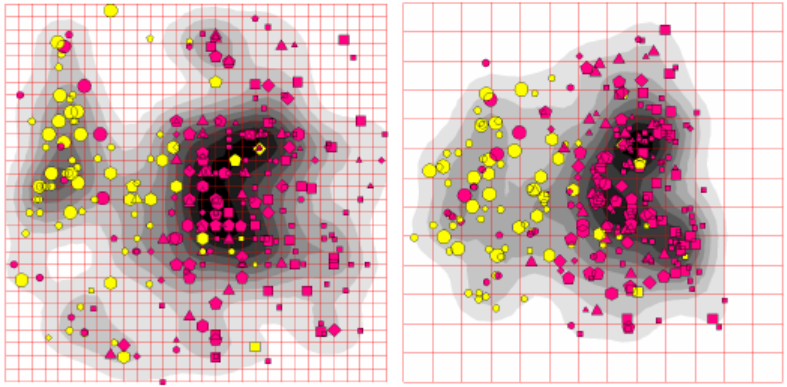
Clustering Hiéarchique / Heat Map

- Tendance à sur-interpréter l'ordre
- Le dendrogramme est difficile à visualiser pour 20,000 gènes

PCA

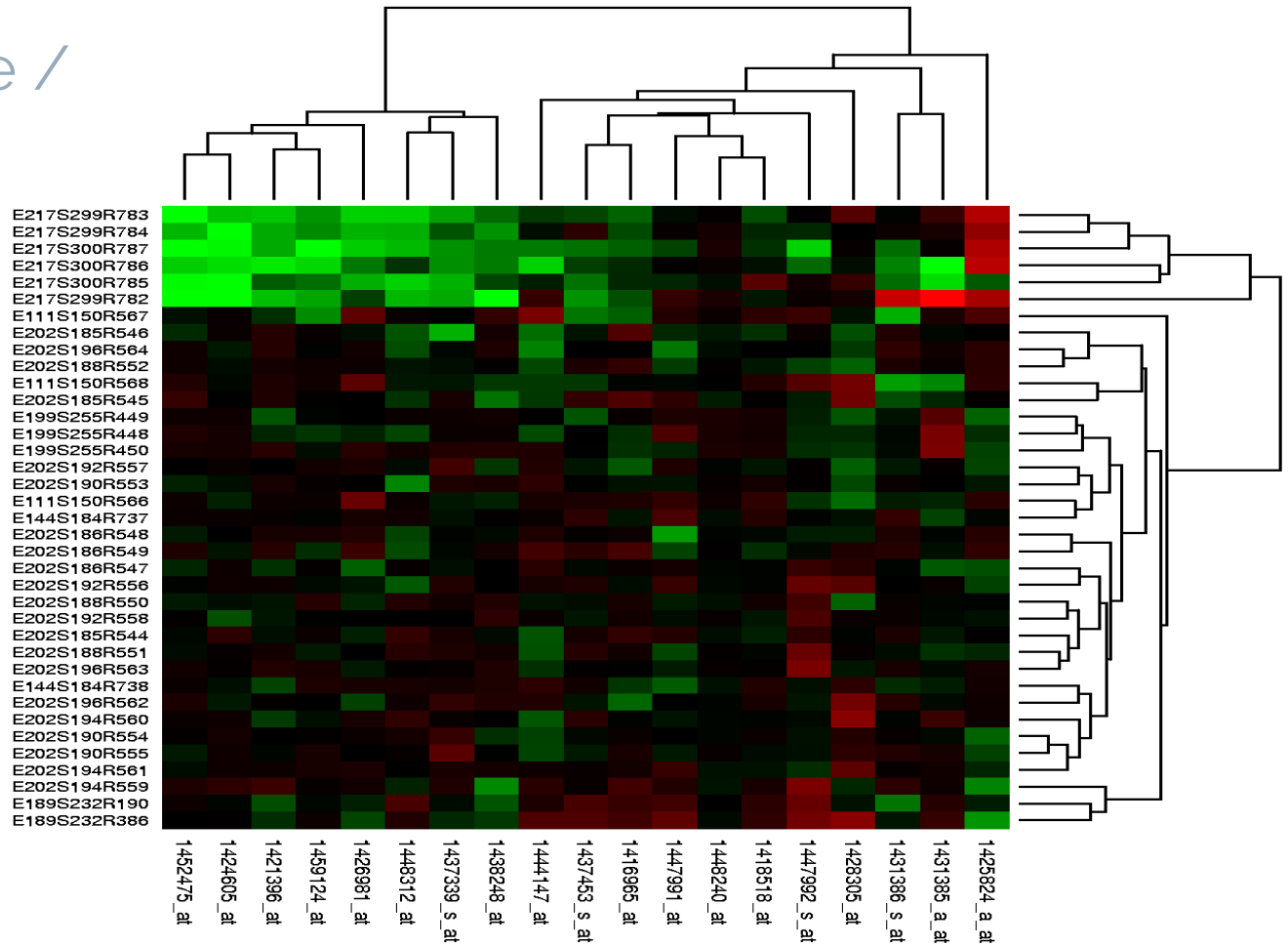


a)

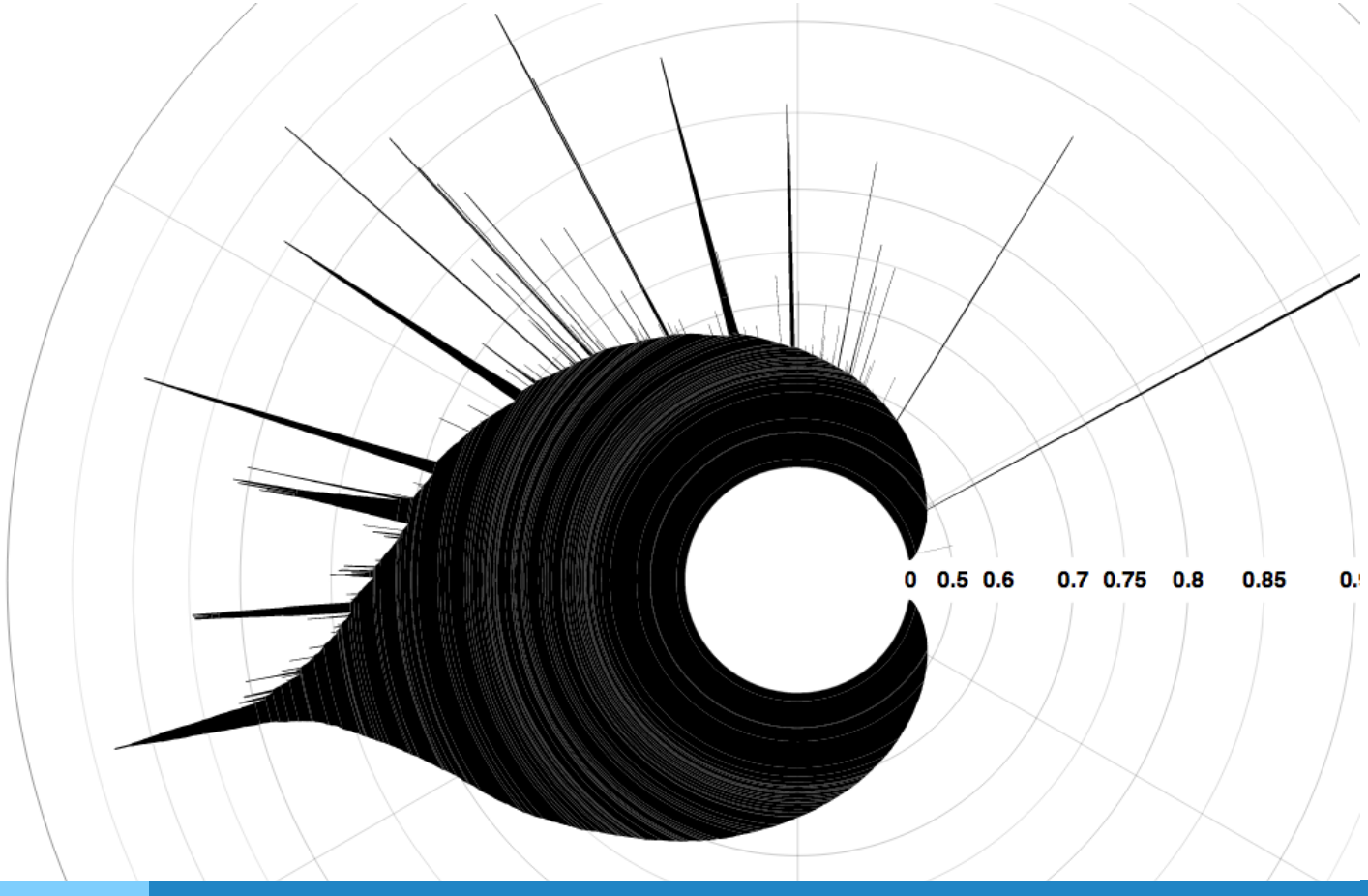


b) ELMAP2D c) PCA2D

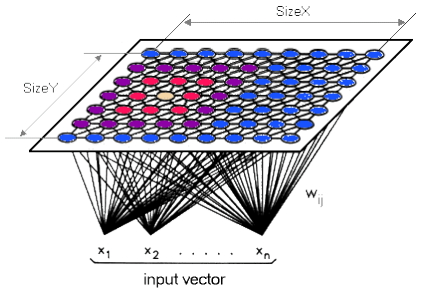
Clustering Hiéarchique / Heat Map



Icicle



...Idée!



- **Self Organizing Maps (SOM)**
  Approche d'apprentissage machine non supervisé. Tout comme Heat Map et PCA, permet la visualisation des données multidimensionnelles dans un espace dimensionnel plus restreint (e.g. 2D).

Dataset : **TCGA**

Données RNA-Seq provenant du *Cancer Genome Atlas*.

Fichier d'expressions en format CSV transformé $\log_2$ (RSEM+1) pour l'ensemble AML.

Data

	Sample_A	Sample_B	Sample_C	...
MTVR2	0.7372	0.6584	0.6615	...
HIF3A	0.2804	0.3382	0.2943	...
RNF10	1.113	1.080	1.1006	...
...	...	...	...	...

Voilà!

Questions?

Projet

URL : www-ens.iro.umontreal.ca/~seguinjo/biohack15