

---

# Structure de corrélations entre les gènes et impact sur la pertinence des signatures génomiques

F. Valet, E. Gravier, Y. De Rycke et B. Asselain  
Institut Curie, Ecole des Mines de Paris, Inserm U900

*Journées des Statisticiens des Centres de Lutte contre le Cancer  
Conférence Francophone d'Epidémiologie Clinique*

---

# Contexte

---

- Données de transcriptome

	$P_1$	$P_2$	$P_3$	$P_4$	...	$P_i$	...	$P_{100}$
$V_1$	-4.031	-0.520	0.559	0.112	...	0.677	...	-0.207
$V_2$	-1.467	-0.440	1.153	0.086	...	-0.487	...	-0.423
$V_3$	1.376	0.735	-1.344	-0.367	...	0.433	...	-0.051
$V_4$	1.398	0.169	-0.817	-1.233	...	-1.967	...	-0.402
...	...	...	...	...	...	...	...	...
$V_{10000}$	0.123	-0.569	7.236	5.001	...	-1.799	...	9.788
...	...	...	...	...	...	...	...	...
$V_{55000}$	2.297	-0.001	0.378	-0.810	...	-2.151	...	-0.458

- Détection des gènes différentiellement exprimés
  - entre deux **conditions** (ex. Groupe **A** vs. Groupe **B**)
  - associés à un **événement** (ex. **Survie**, **Apparition de métastases**)

# Contexte

- Données de transcriptome

	P <sub>1</sub>	P <sub>2</sub>	P <sub>3</sub>	P <sub>4</sub>	...	P <sub>i</sub>	...	P <sub>100</sub>
V <sub>1</sub>	-4.031	-0.520	0.559	0.112	...	0.677	...	-0.207
V <sub>2</sub>	-1.467	-0.440	1.153	0.086	...	-0.487	...	-0.423
V <sub>3</sub>	1.376	0.735	-1.344	-0.367	...	0.433	...	-0.051
V <sub>4</sub>	1.398	0.169	-0.817	-1.233	...	-1.967	...	-0.402
...	...	...	...	...	...	...	...	...
V <sub>10000</sub>	0.123	-0.569	7.236	5.001	...	-1.799	...	9.788
...	...	...	...	...	...	...	...	...
V <sub>55000</sub>	2.297	-0.001	0.378	-0.810	...	-2.151	...	-0.458

- Signature génomique

- kesako ?
- littérature galopante
- traitements «assistés» par la génomique

« toute combinaison prédictive ou pronostique de gènes »



# Plan

---

- Structures de corrélations
- Deux exemples de signatures
- Corrélations et limites des modèles (« signatures »)
- Contrôler la corrélation dans les modèles
- Stabilité et multiplicité des modèles
- Applicabilité des « signatures »

# Données : deux exemples

---

## Cancer du Sein → Réponse complète histologique (RCH)

- Essai REMAGUS02 (chimiothérapie néo-adjuvante)
- Facteurs génomiques prédictifs de RCH ?

TRAIN → N=81

TEST → N=72

**SIGNATURE → modèle logistique multivarié**

## Mélanome Uvéal → Apparition de métastases

- Facteurs génomiques associés à un mauvais pronostic ?

TRAIN → N=63

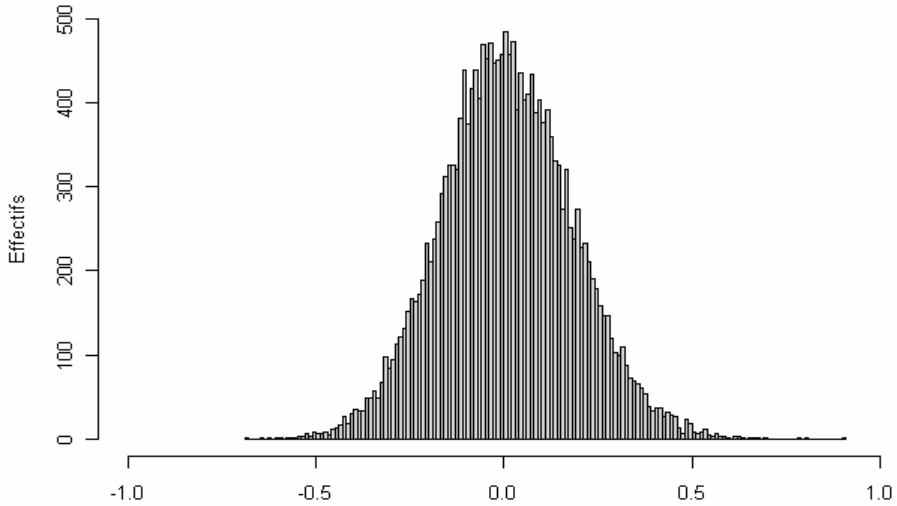
TEST → N=22 (Onken)

**SIGNATURE → modèle de Cox multivarié**

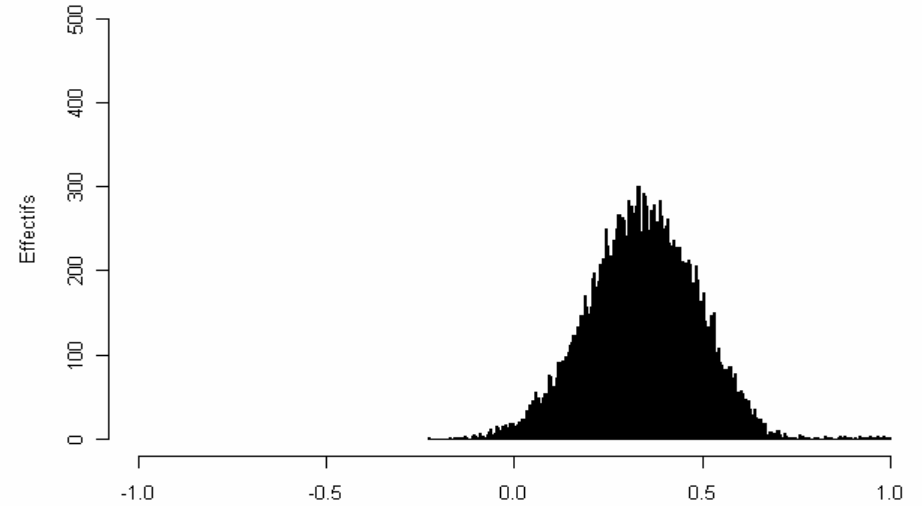
variables de départ ~**200 gènes différentiels** (FDR <5%)

# Les corrélations entre les gènes...

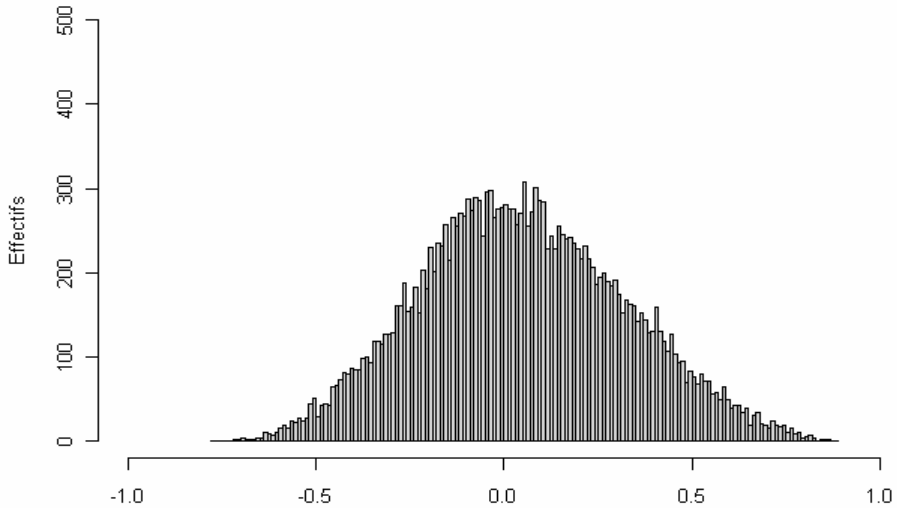
200 gènes pris au hasard (Cancer du Sein, N=81)



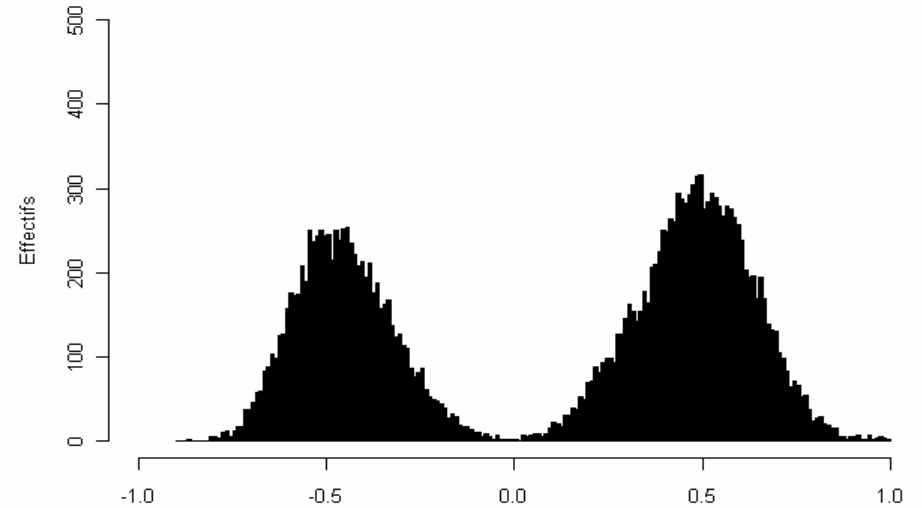
200 gènes associés à la RCH (Cancer du Sein, N=81)



200 gènes pris au hasard (Mélanome, N=63)



200 gènes associés à la survie (Mélanome, N=63)



# ...induisent des corrélations entre les paramètres

---

Sein

Gènes	Logistique Multivarié			Logistique Univarié		
	$\hat{\beta}$	se	pv	$\hat{\beta}$	se	pv
V <sub>1</sub>	-124.28	>10 <sup>4</sup>	~1	-2.23	0.62	<0.001
V <sub>2</sub>	-139.03	>10 <sup>4</sup>	~1	-2.23	0.67	<0.001
V <sub>3</sub>	-550.68	>10 <sup>4</sup>	~1	-6.86	2.87	0.017
V <sub>4</sub>	-45.35	>10 <sup>4</sup>	~1	-1.56	0.51	0.002

# ...induisent des corrélations entre les paramètres

Sein

Gènes	Logistique Multivarié			Logistique Univarié		
	$\hat{\beta}$	se	pv	$\hat{\beta}$	se	pv
V <sub>1</sub>	-124.28	>10 <sup>4</sup>	~1	-2.23	0.62	<0.001
V <sub>2</sub>	-139.03	>10 <sup>4</sup>	~1	-2.23	0.67	<0.001
V <sub>3</sub>	-550.68	>10 <sup>4</sup>	~1	-6.86	2.87	0.017
V <sub>4</sub>	-45.35	>10 <sup>4</sup>	~1	-1.56	0.51	0.002

Corrélation entre les paramètres

	$\hat{\beta}_{V_1}$	$\hat{\beta}_{V_2}$	$\hat{\beta}_{V_3}$	$\hat{\beta}_{V_4}$
$\hat{\beta}_{V_1}$	1.00	0.62	0.73	0.78
$\hat{\beta}_{V_2}$		1.00	0.15	0.55
$\hat{\beta}_{V_3}$			1.00	0.67
$\hat{\beta}_{V_4}$				1.00



# ...induisent des corrélations entre les paramètres

Mélanome

Gènes	Cox Multivarié			Cox Univarié		
	$\hat{\beta}$	se	pv	$\hat{\beta}$	se	pv
V <sub>1</sub>	0.58	0.31	0.06	1.16	0.24	<0.001
V <sub>2</sub>	1.03	0.23	<0.001	0.60	0.18	<0.001
V <sub>3</sub>	0.87	0.24	<0.001	0.62	0.17	<0.001
V <sub>4</sub>	1.11	0.30	<0.001	0.78	0.16	<0.001
V <sub>5</sub>	-0.81	0.31	0.010	0.57	0.17	<0.001
V <sub>6</sub>	0.63	0.27	0.019	1.04	0.25	<0.001

# ...induisent des corrélations entre les paramètres

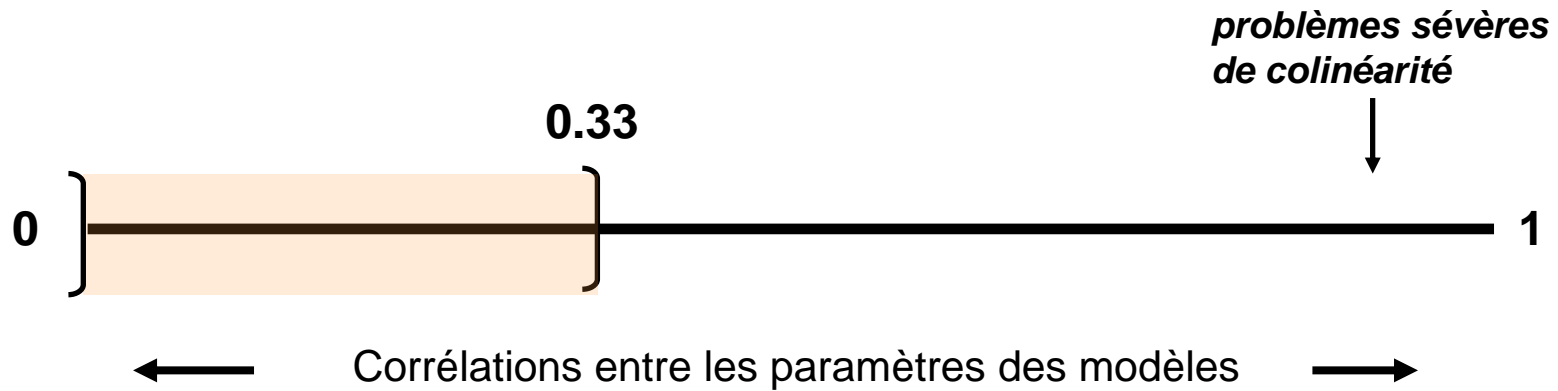
Mélanome

Gènes	Cox Multivarié			Cox Univarié		
	$\hat{\beta}$	se	pv	$\hat{\beta}$	se	pv
V <sub>1</sub>	0.58	0.31	0.06	1.16	0.24	<0.001
V <sub>2</sub>	1.03	0.23	<0.001	0.60	0.18	<0.001
V <sub>3</sub>	0.87	0.24	<0.001	0.62	0.17	<0.001
V <sub>4</sub>	1.11	0.30	<0.001	0.78	0.16	<0.001
V <sub>5</sub>	-0.81	0.31	0.010	0.57	0.17	<0.001
V <sub>6</sub>	0.63	0.27	0.019	1.04	0.25	<0.001

Corrélation entre les paramètres

	$\hat{\beta}_{v_1}$	$\hat{\beta}_{v_2}$	$\hat{\beta}_{v_3}$	$\hat{\beta}_{v_4}$	$\hat{\beta}_{v_5}$	$\hat{\beta}_{v_6}$
$\hat{\beta}_{v_1}$	1.00	-0.06	-0.01	-0.26	-0.27	-0.11
$\hat{\beta}_{v_2}$		1.00	0.42	0.47	-0.18	-0.01
$\hat{\beta}_{v_3}$			1.00	0.48	-0.58	-0.08
$\hat{\beta}_{v_4}$				1.00	-0.57	0.09
$\hat{\beta}_{v_5}$					1.00	-0.11
$\hat{\beta}_{v_6}$						1.00

# Niveaux de corrélations entre les paramètres



- Contrôle de la corrélation dans les modèles

- méthodes complexes (lasso, ridge, ...)
- méthode simple

**contrôle de corrélation entre les paramètres  
(ex. seuil fixé à 0.33)**

- Evaluation de l'impact des structures de corrélations sur

- la stabilité des modèles
- la multiplicité des modèles

# Stabilité des modèles (sur le jeu TRAIN)

---

## Evaluation

- modèle d'origine sur les N patients ( $M_o$ )
- N modèles à partir des (N-1) patients

**Stabilité = proportion de modèles identiques au modèle de départ**

## Application

	Cancer Sein (N=81)	Mélanome Uvéal (N=63)
$M_o$ (corrélation < 0.33)	<b>Stabilité (4 gènes) 64 %</b>	<b>Stabilité (6 gènes) 25 %</b>

# Multiplicité des modèles & validation sur le jeu TEST

## Evaluation

- modèle  $M_O$  (d'origine) construit à partir des 200 gènes on retire le premier gène du modèle
- modèle  $M_{LGO}$  à partir des 199 gènes...puis des 198...100
- performances (AUC et C-Index)

**multiplicité = proportion de  $M_{LGO}$  avec des performances  $\geq M_O$**

## Application

	<b>Cancer Sein (N=81)</b>	<b>Mélanome Uvéal (N=63)</b>
	Modèle Logistique (AUC)	Modèle de Cox (C-index)
$M_O$ (corrélation < 0.33)	<b>0.80</b>	<b>0.76</b>
$M_{LGO}$ (corrélation < 0.33)	<b>Multiplicité 15 %</b>	<b>Multiplicité 58 %</b>

# Conclusion

---

- Impact des structures de corrélations
  - méthodologique (interprétabilité)
  - utilisation de méthodes pour contrôler les corrélations
  - instable et multiple → quelle(s) signature(s) choisir ?
- Méthode leave-patient/gene-out conseillées
  - permet d'évaluer la robustesse d'une méthode
  - évaluations des méthodes de régression pénalisées

# Remerciements

---

## Groupe REMAGUS02

(E. Rolland, P. De Cremoux, F. Spyrtatos, coordinateur M. Marty)

## Groupe Mélanome Uvéal

(C. Plancher, S. Piperno-Neuman, coordinateur S. Roman Roman)

## Service de Bioinformatique, U900

(N. Servant, C. Laurent)

# Discussion

---

## Signatures génomiques ajustées ?

- littérature : signatures brutes ou confrontation *a posteriori* !
- nécessité de prendre en compte les facteurs classiques
- travail en cours : évaluation de l'apport du transcriptome

Cancer du sein : **RP + Taille + G<sub>1</sub> + G<sub>2</sub> + G<sub>3</sub> + G<sub>4</sub>...** ?

**Modèle avec un seul gène**

contrôle de corrélation

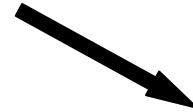
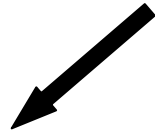
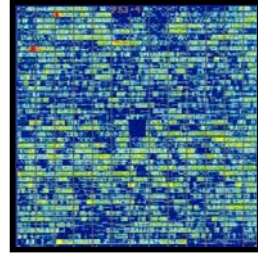
stable / se valide mieux que les modèles avec plusieurs



Prédire

vs.

Interpréter



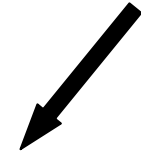
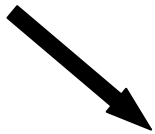
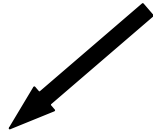
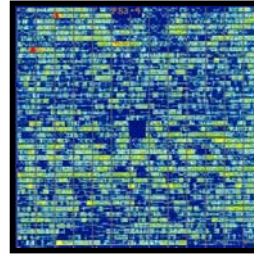
- Signatures**
- sensibilité à la chimiothérapie (6)
  - prédictive de récidence (8)
  - pronostique (4)

- Mécanismes**
- processus métastatique
  - interactions entre gènes
  - gènes prédictifs / pronostiques

Prédire

vs.

Interpréter



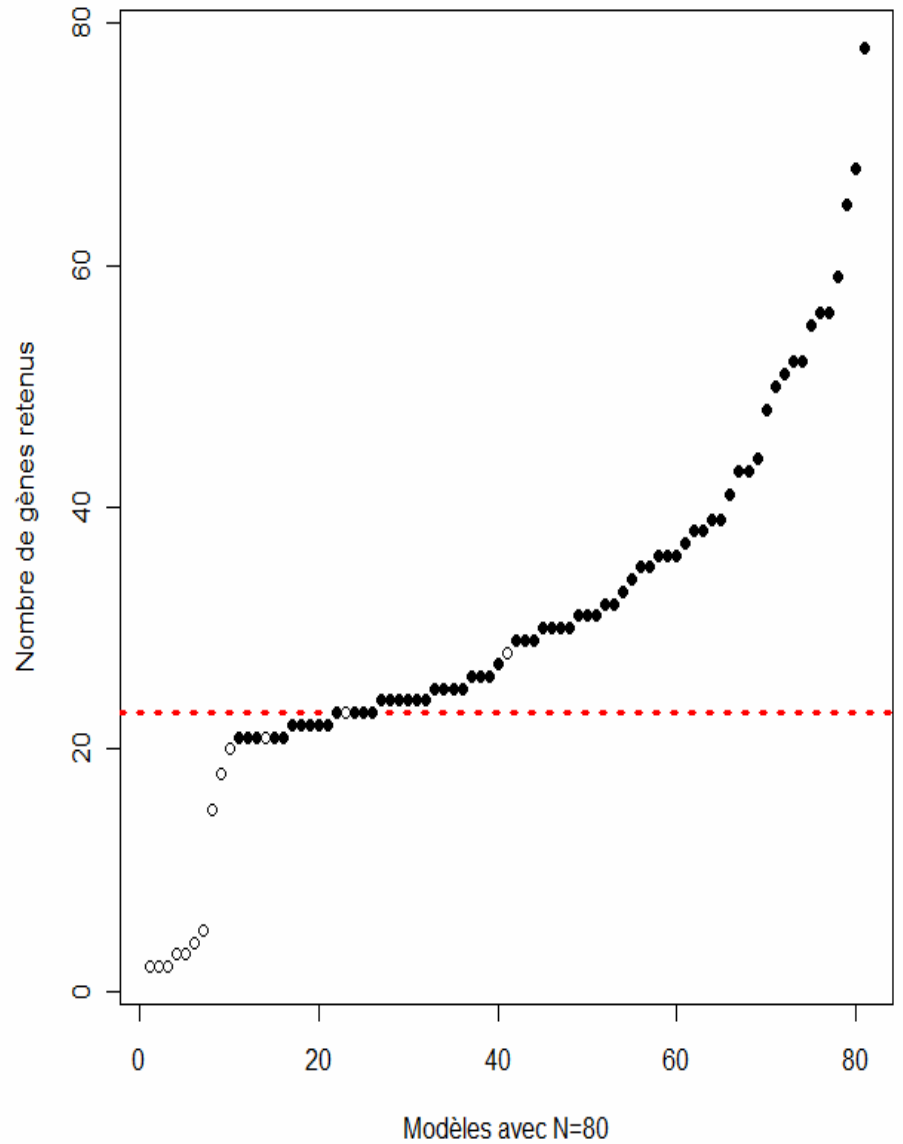
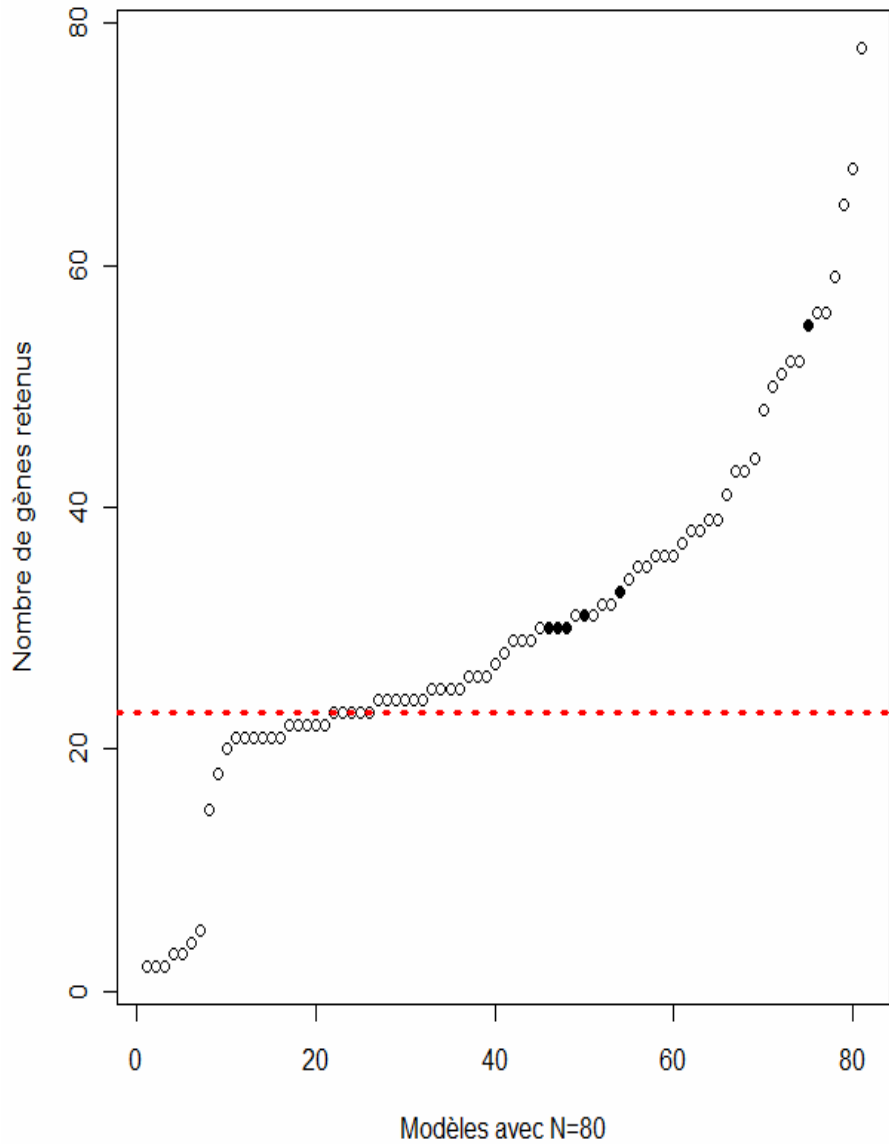
### Signatures

- sensibilité à la chimiothérapie (6)
- prédictive de récidence (8)
- pronostique (4)

### Mécanismes

- processus métastatique
- interactions entre gènes
- gènes prédictifs / pronostiques

# Stabilité avec Lasso



# Variabilité dans le traitement et l'analyse des données génomiques

