

Les Biomarqueurs émergents : intérêts et limites des études disponibles

Pr Pascal Roy MD PhD

Dr Delphine Maucort-Boulch MD PhD

Equipe Biostatistique-Santé
UMR 5558, CNRS, Université de Lyon, Hospices Civils de Lyon
Pôle Rhône Alpin de Bio-Informatique

Société Française de Sénologie et Pathologie Mammaire
31èmes journées, Lyon, 11-13 novembre 2009



Hôpitaux de Lyon



INFERENCE STATISTIQUE → ESTIMATIONS

Epidémiologie Descriptive
Risque de cancer

Recherche Clinique
**Probabilités de guérison ≠
Modèle pronostique**

Intervalles de confiance à 95%

ETUDES PRONOSTIQUES

Identifier des patients à risque (récidive, décès)

Traitement adapté

Inclusion dans des essais

Facteurs Clinico-Biologiques Classiques (**FCBC**)

Identifier de Nouveaux Biomarqueurs (**NB**)

Hypothèse unique

Analyse à haut débit

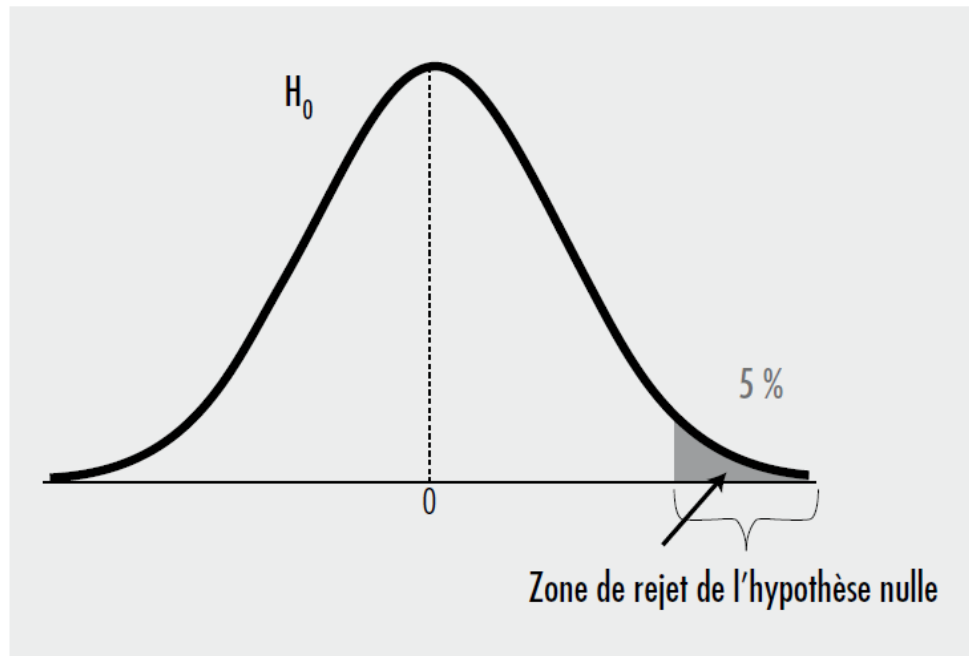
TEST D'UNE HYPOTHESE UNIQUE

Construire une grandeur test

$$\frac{M_{\text{Récidive}} - M_{\text{Indemne}}}{\text{Erreur - type}(\Delta)}$$

Sous H_0

Figure 1. Distribution d'échantillonnage de la grandeur test sous H_0 (zone de rejet définie dans le cadre d'un test unilatéral).



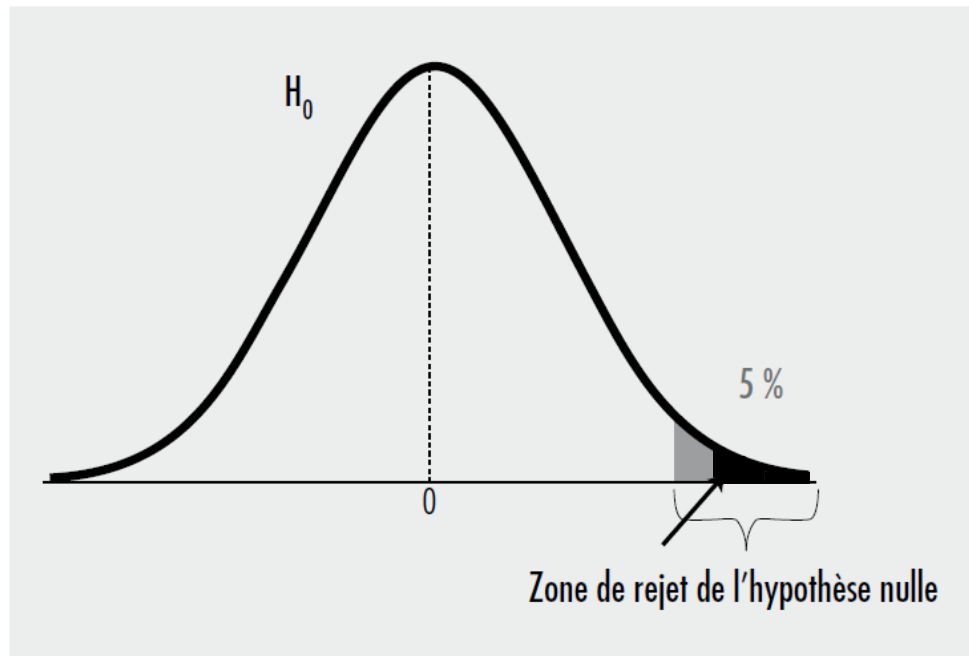
TEST D'UNE HYPOTHESE UNIQUE

Construire une grandeur test

$$\frac{M_{\text{Récidive}} - M_{\text{Indemne}}}{\text{Erreur - type}(\Delta)}$$

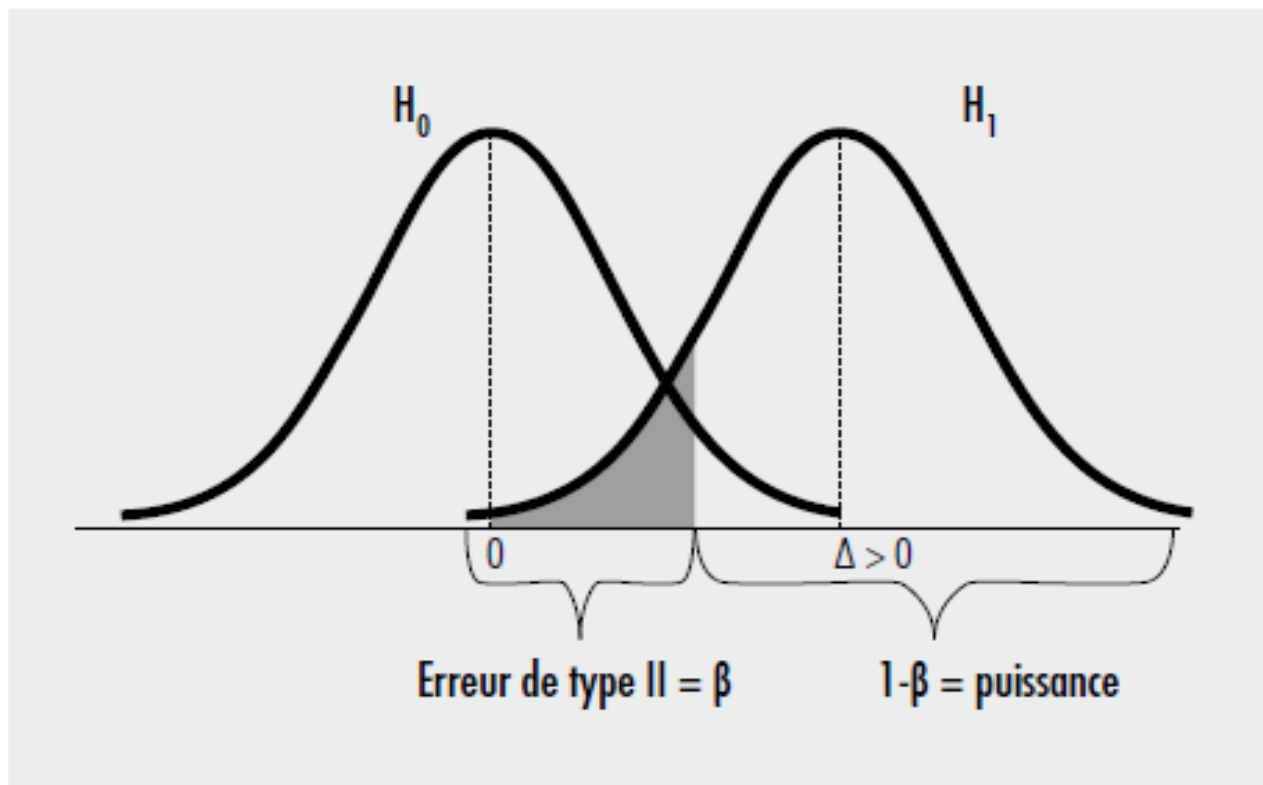
Sous H_0

Figure 1. Distribution d'échantillonnage de la grandeur test sous H_0 (zone de rejet définie dans le cadre d'un test unilatéral).



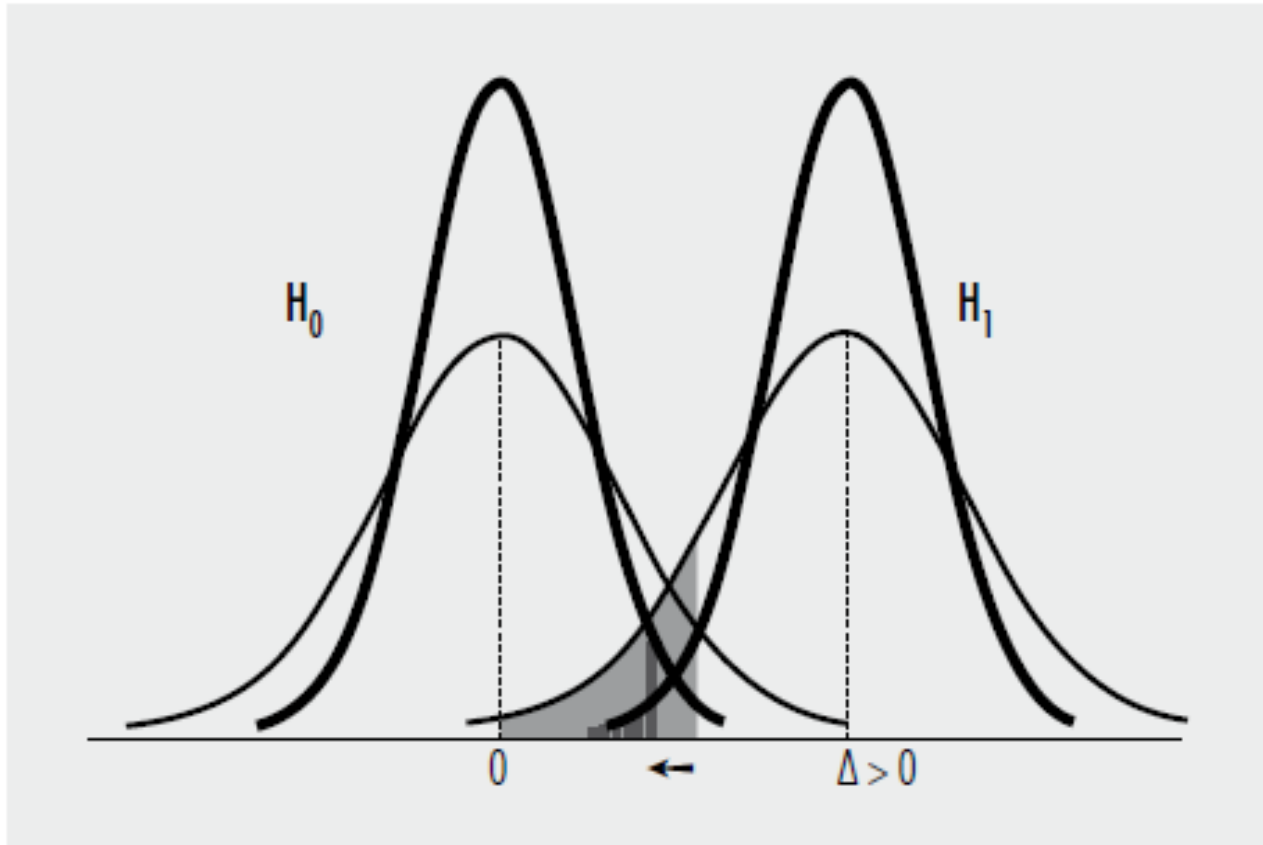
PUISSANCE D'UN TEST UNIQUE

Figure 2. *Fluctuation d'échantillonnage de la grandeur test sous H_0 et H_1 .*



PUISSANCE ET TAILLE DES ETUDES

Figure 3. *Influence de la taille de l'échantillon sur la puissance du test.*



AJUSTEMENT ET PREDICTION

Echantillon de travail → adéquation

Nouvelles données → prédiction

Sélection de variables → optimisme

Echantillon travail ↑ → optimisme ↓

PARTICULARITES HAUT-DEBIT

		Décision		
		H ₀ non rejetée	H ₀ rejetée	
Réalité	H ₀ vraie	U	V	m ₀
	H ₀ fausse	T	S	m ₁
		m-R	R	m

Table 1. Répartition des résultats des différents tests effectués

$$\Pr(V \geq 1 | H_0^C) = 1 - (1 - \alpha)^{10} = 1 - (0,95)^{10} = 0,401.$$

PARTICULARITES HAUT-DEBIT

		Décision		
		H ₀ non rejetée	H ₀ rejetée	
Réalité	H ₀ vraie	U	V	m₀
	H ₀ fausse	T	S	m₁
		m-R	R	m

Table 1. Répartition des résultats des différents tests effectués

$$FDR = E(Q)$$

$$FWER = p(V \geq 1)$$

$$Q = \begin{cases} V/R, & \text{if } R > 0 \\ 0, & \text{if } R = 0 \end{cases}$$

PARTICULARITES HAUT-DEBIT

		Décision		
		H ₀ non rejetée	H ₀ rejetée	
Réalité	H ₀ vraie	U	V	m₀
	H ₀ fausse	T	S	m₁
		m-R	R	m

Table 1. Répartition des résultats des différents tests effectués

$$\beta_F = p(T \geq 1) = 1 - p(T = 0)$$

$$\text{Puissance} = \frac{E(S)}{m_1} = 1 - \beta_1$$

PUISSANCE ET HAUT DEBIT

Opposition α et β

puissance $\uparrow \Rightarrow$ FRD $\uparrow$

$\uparrow$ Ecart H_0 et H_1

puissance $\uparrow$

m et FDR fixés

$m_0 \uparrow \Rightarrow$ puissance $\downarrow$

m_1 et FDR fixés

$m \uparrow \Rightarrow$ puissance $\downarrow$

IDENTIFICATION DES SIGNATURES

7 études

Signatures pronostiques de cancers

Validation interne par re-échantillonnage

Propriétés prédictives des signatures

Michiels S, Koscielny S, Hill C. Prediction of cancer outcome with microarrays: a multiple random validation strategy. *Lancet* 2005;**365**:488–92.

IDENTIFICATION DES SIGNATURES

Composition du prédicteur (gènes) instable

Echantillon travail $\uparrow$ $\rightarrow$ prédiction $\uparrow$

Publications initiales $\rightarrow$ optimisme

5 / 7 études, $\rightarrow$ Prédications $\approx$ hasard

Manque de puissance

Michiels S, Koscielny S, Hill C. Prediction of cancer outcome with microarrays: a multiple random validation strategy. *Lancet* 2005;**365**:488–92.

IDENTIFICATION DES SIGNATURES

Signature prédictive

Validation externe → optimisme

Plusieurs signatures

Signatures robustes → grande étude

1. Van't Veer LJ *et al.* *Nature* 2002;**415**:530–6.
2. Van de Vijver MJ *et al.* *N Engl J Med* 2002;**347**:1999-2009.
3. Buyse M *et al* *J Natl Cancer Inst* 2006;**98**:1183–92.
4. Ein-Dor L *et al.* *Bioinformatics* 2005;**21**:171-178.
5. Ein-Dor L *et al.* *PNAS* 2006;**103**:5923-5928.

DISCUSSION

Technologie haut débit

Plans expérimentaux adaptés

Variabilités technique et biologique

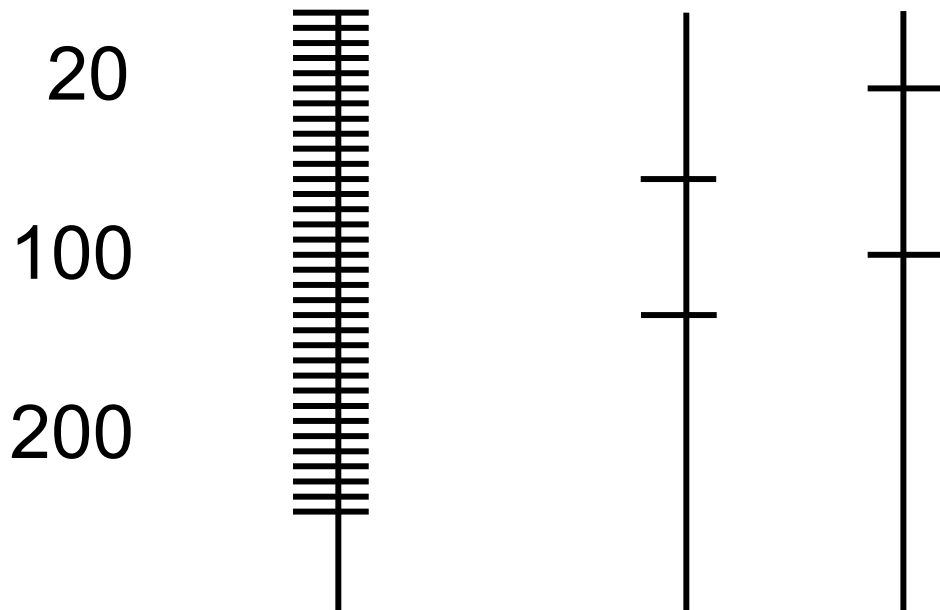
Dimension (gènes différentiels et signatures)

Entités diagnostiques et pronostiques

DISCUSSION

Faible puissance des études réalisées

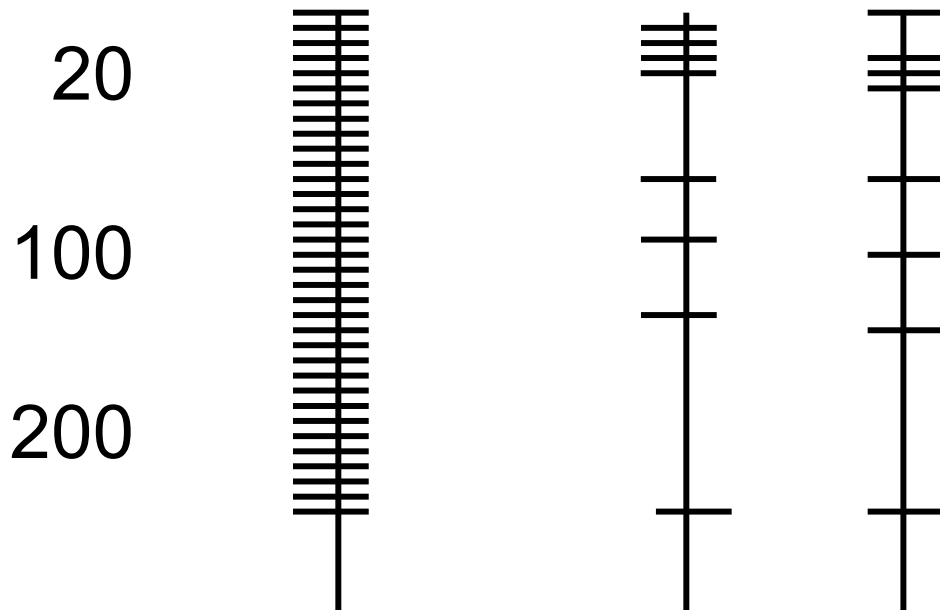
Taille des études initiales (identification)



DISCUSSION

Faible puissance des études réalisées

Taille des études initiales (identification)



DISCUSSION

Taille des études d'identification et de validation

Etudes d'identification et optimisme

Analyse conjointe FCBC et NB

Truntzer C, Maucourt-Boulch D, Roy P. Comparative optimism in models involving both classical clinical and gene expression information. BMC Bioinformatics 2008 Oct 15;9:434.

1. Benjamini, Y. and Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* 1995;**5**:289-300.
2. Yekutieli, D. and Benjamini, Y. A resampling based false discovery rate controlling multiple test procedure. *J. Statist. Plann. Inference* 1999;**82**:171-196.
3. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. *Ann Stat* 2002;**29**:1165-1188.
4. Mei-Ling Ting Lee and G.A. Whitmore. Power and sample size for DNA microarray studies. *Statistics in Medicine* 2002;**21**:3543-3570.
5. Pawitan Y, Michiels St, Koscielny S, Gusnanto A and Ploner A. False discovery rate, sensitivity and sample size for microarray studies. *Bioinformatics* 2006;**21**:3017-3024.
6. Michiels S, Koscielny S, Hill C. Prediction of cancer outcome with microarrays: a multiple random validation strategy. *Lancet* 2005;**365**:488–92.
7. Van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002;**415**:530–6.
8. Van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 2002;**347**:1999-2009.
9. Buyse M, Loi S, van't Veer L, Viale G, Delorenzi M, Glas AM, et al. Validation and Clinical Utility of a 70-Gene Prognostic Signature for Women With Node-Negative Breast Cancer *J Natl Cancer Inst* 2006;**98**:1183–92
10. Ein-Dor L, Kela I, Getz G, Givol D and Domany E. Outcome signature genes in breast cancer: is there a unique set ? *Bioinformatics* 2005;**21**:171-178.
11. Ein-Dor L, Zuk O, Domany E. Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer? *PNAS* 2006;**103**:5923-5928.
12. Golub T, Slonim D, Tamayo P: Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999;**286**:531-537.
13. Truntzer C, Mercier C, Estève J, Gautier C, Roy P. Importance of data structure in comparing two dimension reduction methods for classification of microarray gene expression data. *BMC Bioinformatics* 2007 Mars 13;**8**:90
14. Truntzer C, Maucort-Boulch D, Roy P. Comparative optimism in models involving both classical clinical and gene expression information. *BMC Bioinformatics* 2008 Oct 15;**9**:434.