

Analyse des données de puce à ADN

Houlgatte Rémi
TAGC INSERM ERM0206
houlgatte@tagc.univ-mrs.fr

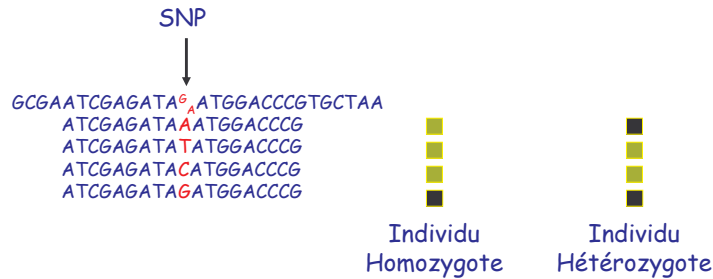
Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

Puces à ADN: applications

	Sondes	Cible	Référence
Criblage de banques	ADN	ADN	Gress TM et al. (1992) <i>Mam. Genome</i> 3: 609-619
Transcriptome	ADN	ARN	Schena et al. (1995) <i>Science</i> 270: 467-470
Oligo fingerprinting	ADN	Oligos	Drmanac S. et al. (1996) <i>Genomics</i> 37: 29-40
Reséquencage	Oligo	ADN	Kozal MJ et al. (1996) <i>Nat. Med.</i> 2: 753-759
Phénotypage	ADN	ADN	Gingeras et al. (1998) <i>Genome Res.</i> 8: 435-448
SNP	Oligo	ADN	Winzeler et al. (1998) <i>Science</i> 281: 1194-1197
CGH	ADN	ADN	Pollack JR. (1999) <i>Nature</i> 23: 41-46
DNA-BP	ADN	Protéines	Bulyk et al. (1999) <i>Nature Biotech.</i> 17: 573-577
Suivi de réplication	ADN	ADN	Khodursky et al. (2000) <i>PNAS</i> 97: 9419-9424
Détection de gènes	ADN	ARN	Penn et al. (2000) <i>Nature Genet.</i> 26: 315-318
ChIP-chip	ADN	ADN	Iyer et al. (2001) <i>Nature</i> 409: 533-538
Validation de gènes	ADN	ARN	Shoemaker et al. (2001) <i>Nature</i> 409: 922-927
Taille des ARN	ADN	ARN	Hurowitz et al. (2003) <i>GenBiol</i> 5: R2

Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

Single Nucleotide Polymorphism



Pour un SNP donné, on synthétise 4 oligonucléotides ayant la séquence connue et les 4 bases possibles en position centrale.

On hybride la puce constituée de ces oligonucléotides avec l'ADN d'un individu.

Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

ReSéquençage

```
GCGAATCGAGATAGATGGACCCGTGCTAA  
ATCGAGATAAATGGACCCG  
ATCGAGATATATGGACCCG  
ATCGAGATACATGGACCCG  
ATCGAGATAGATGGACCCG
```

```
GCGAATCGAGATAGATGGACCCGTGCTAA  
TCGAGATAGATGGACCCGT  
TCGAGATAGTTGGACCCGT  
TCGAGATAGCTGGACCCGT  
TCGAGATAGGTGGACCCGT
```

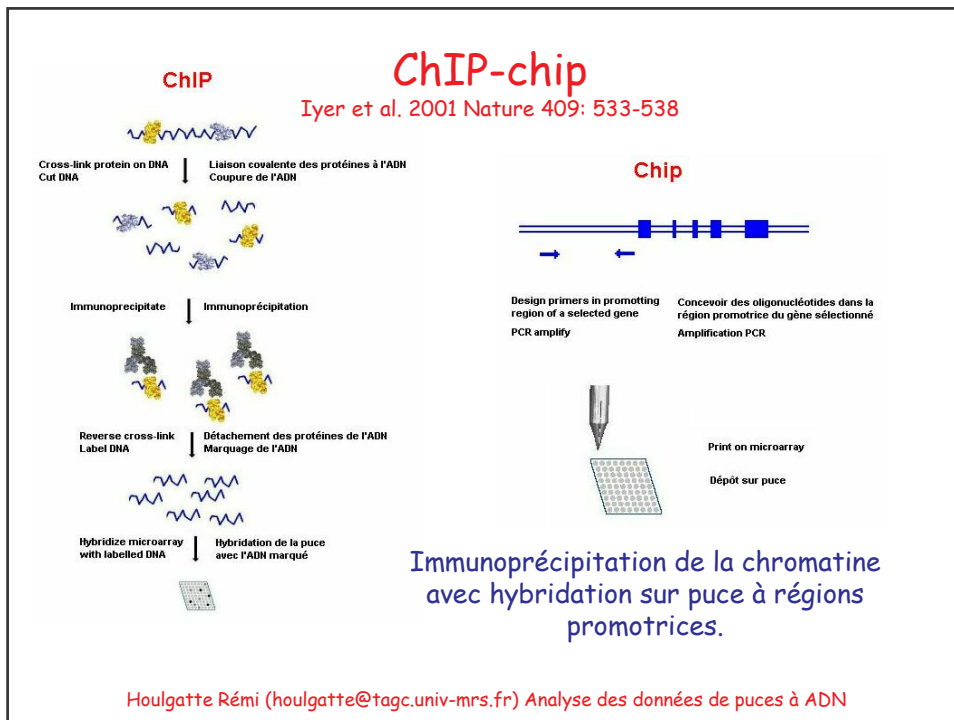
```
GCGAATCGAGATAGATGGACCCGTGCTAA  
CGAGATAGAAGGACCCGTG  
CGAGATAGATGGACCCGTG  
CGAGATAGACGGACCCGTG  
CGAGATAGAGGGACCCGTG
```

Pour toutes les positions à reséquencer, on synthétise 4 oligonucléotides ayant la séquence connue et les 4 bases possibles en position centrale.

On hybride la puce constituée de ces oligonucléotides avec l'ADN d'un individu.



Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN



Support solide

Intégration de sondes

Miniaturisation

Automatisation

Microfluidique

Electronique embarquée

Biopuces

	Sondes	Cible	Référence
Transcriptome	ADN	ARN	Schena et al. Science (1995) 270: 467-470
Tissue-array	Tissus	Anticorps	Kononen J. (1998) Nat. Med. 4: 844-847
Fonction	Protéines	Ligands	Zhu et al. Nature Genet (2000) 26: 283-289
Protéome	Anticorps	Protéines	De Wildt et al. Nature Biotech. (2000) 18: 989-994
Fonction	Cellules & Plasmides	Ligands	Ziauddin & Sabatini Nature (2001) 411: 107-110
Interactome	Protéines	Protéines	Zhu et al. Science (2001) 293: 2101-2105

Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

Tissue MicroArray

Kononen et al. 1998 Nat. Med. 4: 844-847

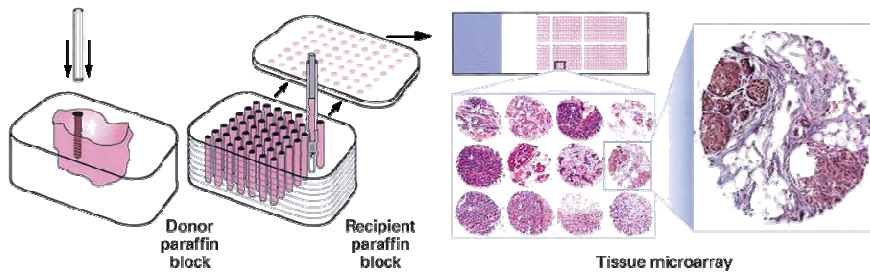
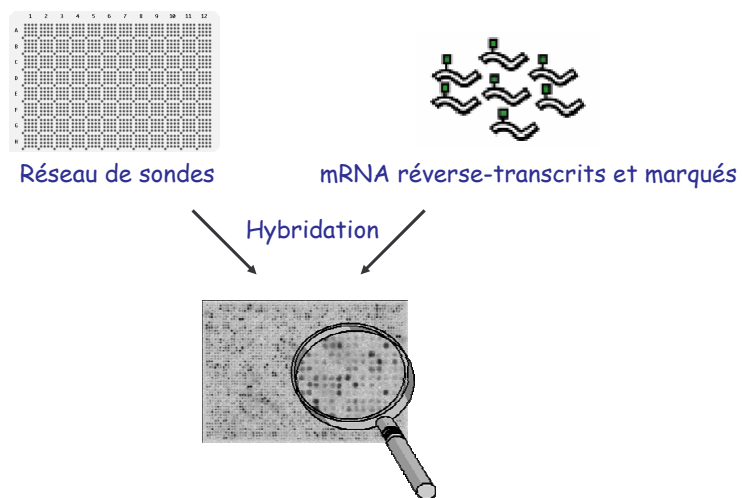


Figure tirée de Hedenfalk et al. 2001 NEJM 344: 539-548

Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

Puces à ADN: principe

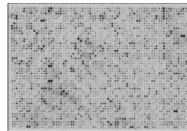


Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

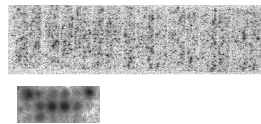
Puces à ADN: technologies

- Sondes:
 - cDNA (PCR; plasmides; ORF)
 - Oligonucléotides
 - PNA (Peptide Nucleic Acid)
- Fabrication:
 - Dépôt
 - Synthèse in-situ
- Marquage:
 - Radioactivité
 - Fluorescence
 - Colorimétrie
 - Chimiluminescence

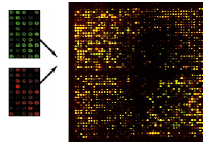
Nylon macroarray
~ 5.000 probes on 12 cm x 8 cm



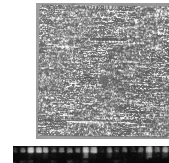
Nylon microarray
~ 10.000 probes on 7 cm x 2 cm



Glass-slide microarray
~ 10.000 probes on 2 cm x 2 cm



Oligochip
~ 100.000 probes on 1 cm x 1 cm



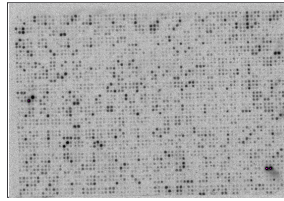
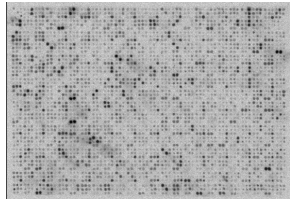
Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

Puces à ADN: propriété mesurée

Espèce hybridée	Propriété mesurée	Référence
mARN	Transcriptome	Schena et al. (1995) Science 270: 467-470
mARN marqués in situ (pulse) mARN marqués in situ (pulse-chase)	Taux de transcription Taux de dégradation	
mARN	Taux de dégradation	Wang et al. 2002 PNAS 99: 5860-5865
mRNA liés aux polysomes	Taux de traduction	Johannes et al. 1999 PNAS 96: 13118-13123
mRNA liés aux membranes	Localisation subcellulaire	Diehn et al. 2000 Nature Genetics 25: 58-62
mRNA liés aux polysomes mitochondriaux	Localisation subcellulaire	Marc et al. 2002 EMBO reports 31 (21)

Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

Approche différentielle



Gène différentiellement exprimé

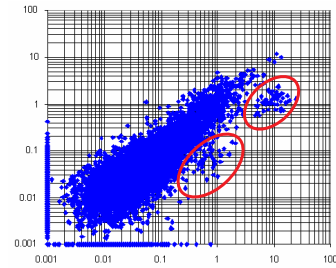
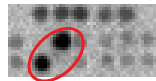
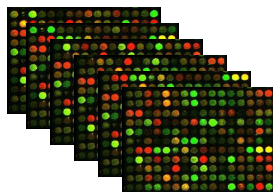


Table 2. List of the genes that show the most frequent differential expression between normal breast tissue and 34 breast carcinomas as measured by cDNA array analysis

Clone ID	Gene/protein identity	Gene symbol	Chromosome location	n	T/NB
Overexpressed genes					
154343	<i>Granzyme H</i>	GZMH	14q11.2	32	9.51
235947	<i>Stromelysin 3</i>	STMY3	22q11.2	31	15.92
207378	<i>MYB related protein B</i>	MYBL2	20q13.1	31	*
153275	<i>Cellular retinoic acid binding protein 2</i>	CRABP2	1q21.3	29	7.16
129757	<i>GATA-binding protein 3</i>	GATA3	10p15	28	17.76
120649	<i>T-lymphocyte surface CD2 antigen</i>	CD2	1p13.1	28	7.54
109677	<i>CREB binding protein</i>	CREBHP	16p13.3	28	5.08
172152	<i>EGFR-binding protein GRB2</i>	GRB2	17q24-q25	28	5.00
60969	<i>Transcription factor RELB</i>	RELB	19	28	3.61

Cause ?
Conséquence ?
On fait quoi ?

Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN



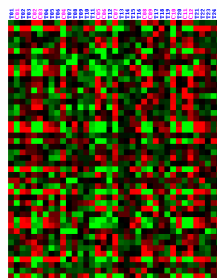
Hybridation

	E01	E02	E03	E04	E05	E06	E07	E08
G01	1,03	5,42	0,03	8,25	0,14	3,34	5,78	0,00
G02	6,18	1,45	6,89	0,25	5,06	7,25	0,00	8,22
G03	6,13	4,50	6,99	0,48	3,56	1,54	2,66	0,41
G04	4,76	0,33	0,01	0,96	1,22	5,87	0,00	3,85
G05	1,03	5,42	0,03	8,25	0,14	3,34	5,78	0,00
G06	6,18	1,45	6,89	0,25	5,06	7,25	0,00	8,22
G07	6,13	4,50	6,99	0,48	3,56	1,54	2,66	0,41
G08	4,76	0,33	0,01	0,96	1,22	5,87	0,00	3,85
G09	6,13	4,50	6,99	0,48	3,56	1,54	2,66	0,41
G10	4,76	0,33	0,01	0,96	1,22	5,87	0,00	3,85
G11	1,03	5,42	0,03	8,25	0,14	3,34	5,78	0,00
G12	6,18	1,45	6,89	0,25	5,06	7,25	0,00	8,22
G13	6,13	4,50	6,99	0,48	3,56	1,54	2,66	0,41
G14	4,76	0,33	0,01	0,96	1,22	5,87	0,00	3,85
G15	6,13	4,50	6,99	0,48	3,56	1,54	2,66	0,41

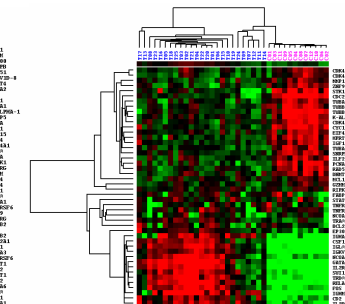
Quantification

Classification hiérarchique

Eisen et al. 1998 PNAS 95: 14863-14868



Colorisation
Rouge = sur-exprimé
Vert = sous-exprimé

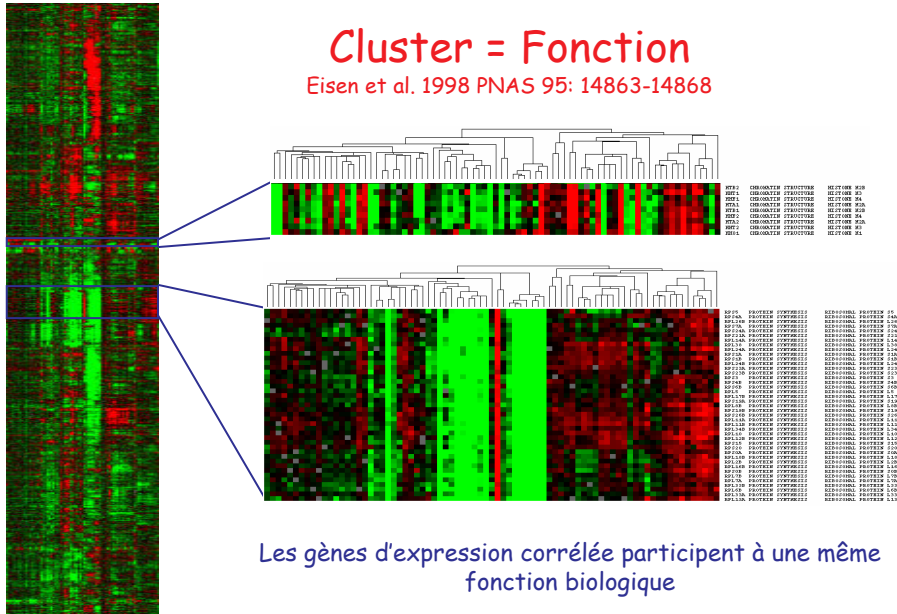


Classification

Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

Cluster = Fonction

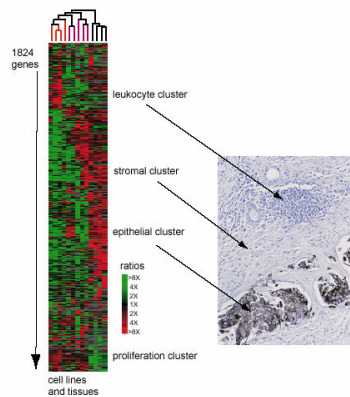
Eisen et al. 1998 PNAS 95: 14863-14868



Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

Dissection virtuelle

Ross et al. 2000 Nat. Genet. 24: 227-235

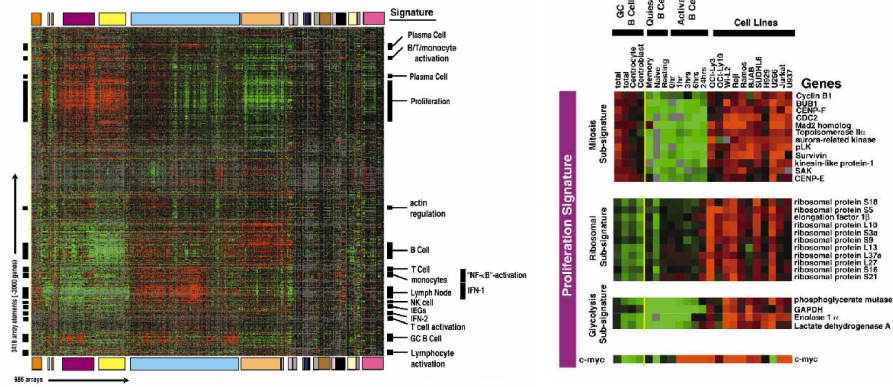


Dans un tissu biologique, on observe des signatures correspondant aux différentes populations cellulaires

Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

Microdissection virtuelle

Shaffer et al. 2001 Immunity 15: 375-385

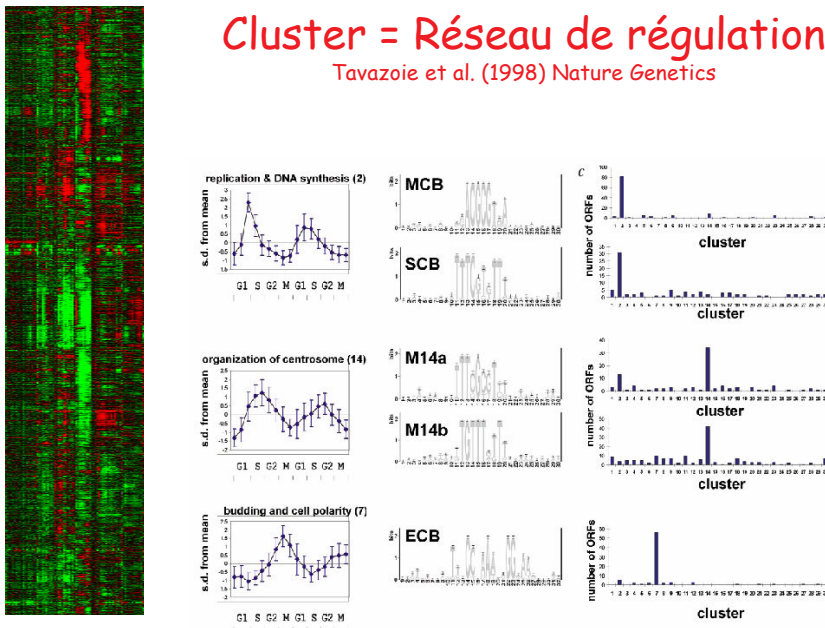


Plus on a d'échantillons en rapport avec une question biologique, plus les signatures (fonctions) deviennent précises.

Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

Cluster = Réseau de régulation

Tavazoie et al. (1998) Nature Genetics



Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

Bioinformatique

1) Conception de la puce

Sélectionner les gènes et concevoir les sondes.
Recherche et analyse des gènes, Conception des sondes et des contrôles ...

2) Plan d'expérience

Prévoir les expériences nécessaires en fonction de la question biologique et des méthodes d'analyse.
Nombres de réplicats, Choix de la puce ...

3) Stockage des données

Assurer la pérennité et l'échange des données.
Base de données, Structure et formats ...

4) Traitement des données

Transformer les données brutes en données validées.
Il existe plutôt un consensus par technologie et avec des subtilités par plate-forme.

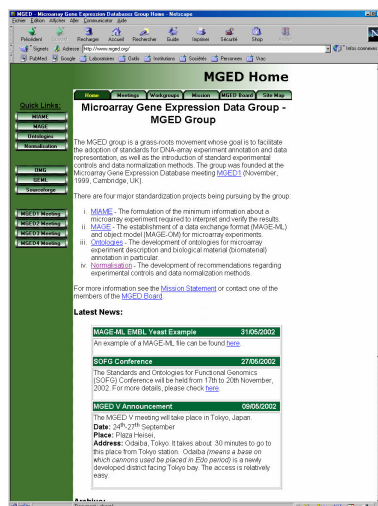
5) Analyse des données

Donner du sens aux résultats obtenus.
Il existe des méthodes avec des forces et des faiblesses, Le choix dépend de la question que l'on veut poser.

Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

Microarray Gene Expression Data

www.mged.org



- i. [MIAME](#) - The formulation of the minimum information about a microarray experiment required to interpret and verify the results.
- ii. [MAGE](#) - The establishment of a data exchange format (MAGE-ML) and object model (MAGE-OM) for microarray experiments.
- iii. [Ontologies](#) - The development of ontologies for microarray experiment description and biological material (biomaterial) annotation in particular.
- iv. [Normalisation](#) - The development of recommendations regarding experimental controls and data normalization methods.

Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

Traitement des données

Traitements visant à transformer les valeurs brutes en mesures d'expression. Ces traitements sont dépendant de la technologie et de la plate-forme.

Bruit de fond: local, global, estimé ...

Valeurs aberrantes: doughnuts, overshining ...

Correction des canaux: différence d'incorporation et de spectre, correction vecteur ...

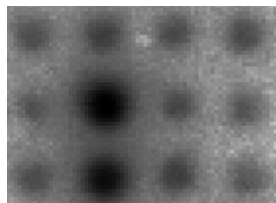
Normalisation: ubiquitaires, référence interne (spike), en masse ...

Effets non-linéaires: Lowess, print-tip Lowess ...

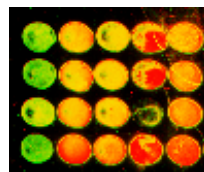
Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

Valeurs aberrantes

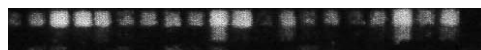
Radioactivité: Overshining



Fluorescence: Problèmes d'homogénéité

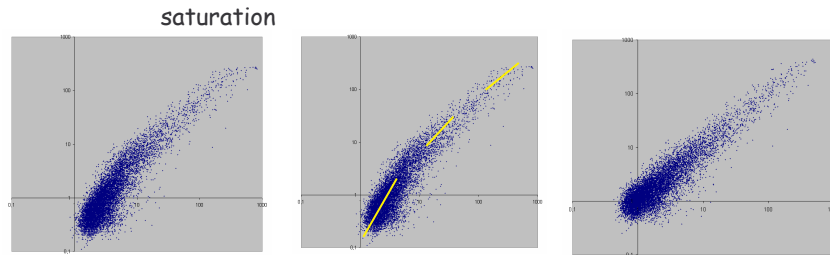


Puces à Oligonucléotides: Mesures très différentes



Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

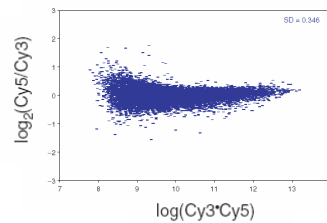
Correction des effets non-linéaires



bruit de fond

Correction des canaux (Rouge / Vert)
Correction du bruit de fond local
Correction de la saturation
Correction des biais d'aiguille

Méthodes utilisant une régression locale
ex: Lowess (Yang et al. 2002 NAR 30:e15)



Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

Normalisation

Correction des effets du marquage,
de l'exposition, de la lecture ...

Gènes ubiquitaires → L'expression de ces gènes ne varie pas → Tous les gènes sont régulés

Référence externe (spike) → ARN de référence ajouté en quantité connue → Problème de dosage

Normalisation en masse → La distribution des rapports est centrée sur 1 → Problème si variation globale

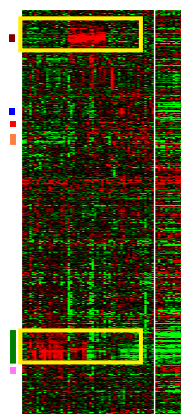
Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

Analyse de données

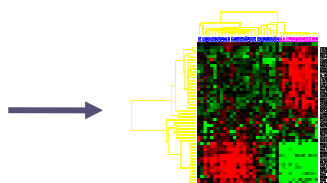
Sélection des données	Méthodes Non Supervisées
Poser des questions	Exploration des données
Choix des échantillons de départ	Classifications hiérarchiques
Utilisation d'échantillons de référence	Nuées dynamiques (K-means)
Sous-Sélection d'échantillons	Cartes de Kohonen (SOM)
Filtrage	Analyse en composantes principales
	Méthodes Supervisées
Transformations	Découverte de classes
Changer le poids relatifs de certaines valeurs	Tests statistiques « classiques »
Log	Utilisation des expressions corrélées
Centrage médian gènes	Score discriminant
Ratio au contrôle	Réseaux de neurones
...	Méthodes de Validation
Annotations	Significativité, Puissance ...
Mettre en évidence les fonctions	Rééchantillonnage
Ontologies	Utilisation du bruit
Echantillons de référence	Leave-one out
...	Apprentissage-Validation (Prédicteur)

Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

Sélection des données



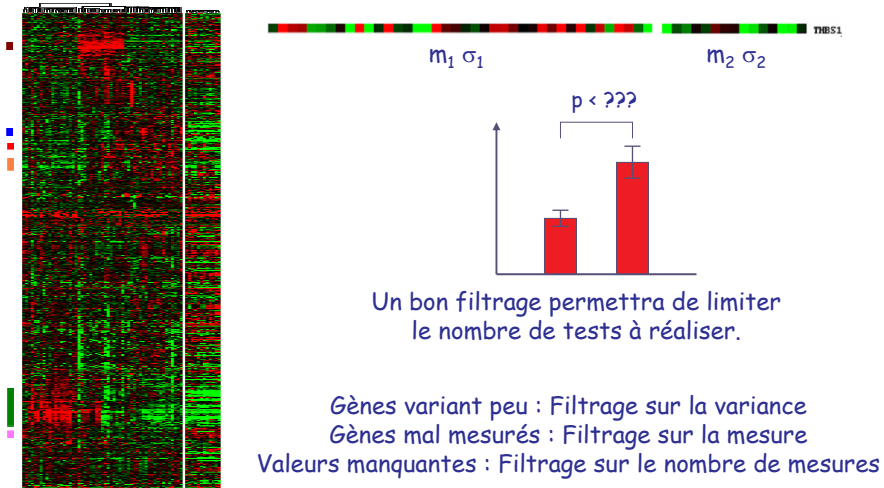
On peut à tout moment poser des questions à un jeu limité de données



On peut décider d'utiliser un jeu d'échantillons pour faire apparaître certaines fonctions. Puis regarder l'état de ces fonctions dans un deuxième groupe d'échantillons

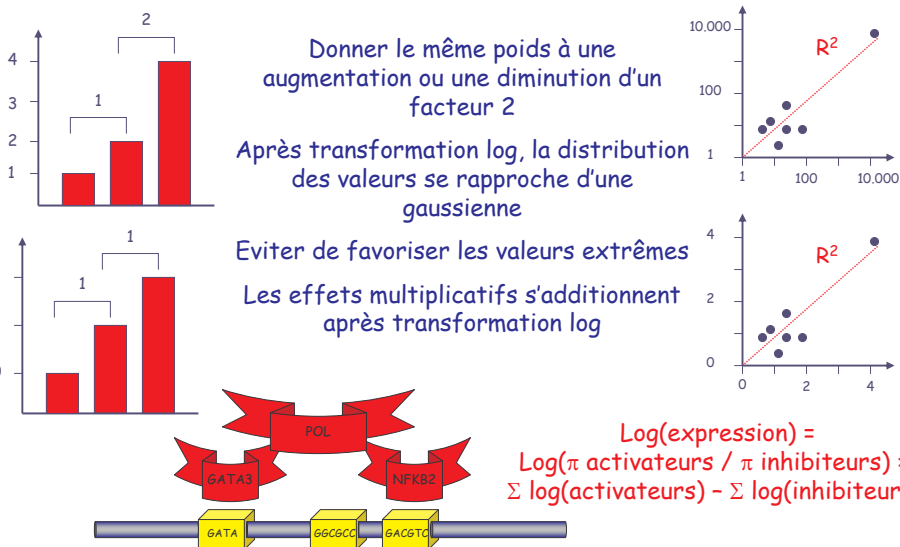
Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

Sélection des données: Filtrage



Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

Transformation log

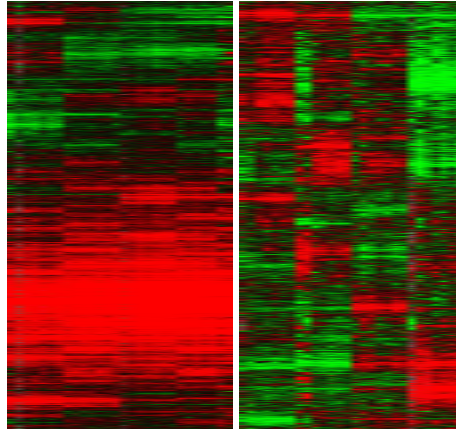


Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

Centrage médian des gènes

Après transformation log: $\text{Log}(\text{mesure}) - \text{Log}(\text{médiane}) = \text{Log}(\text{mesure}/\text{médiane})$

Si l'on classe les données telles quelles, on observe que certains gènes sont faiblement exprimés, d'autres fortement.



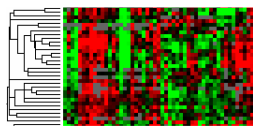
En fait, ce qui nous intéresse c'est de savoir si un gène s'exprime différemment dans certains échantillons.

Le centrage médian des gènes permet de visualiser de tels changements.

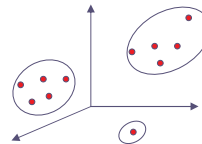
C'est aussi un constat d'ignorance: le niveau « normal » d'un gène, c'est la médiane

Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

Méthodes non supervisées

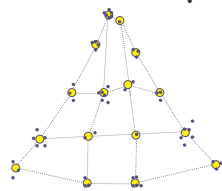


Classifications hiérarchiques

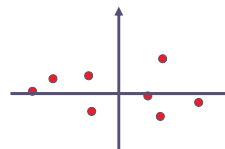


Nuées dynamiques (K-means)

Exploration des données



Cartes de Kohonen (SOM)



Analyse en composantes principales

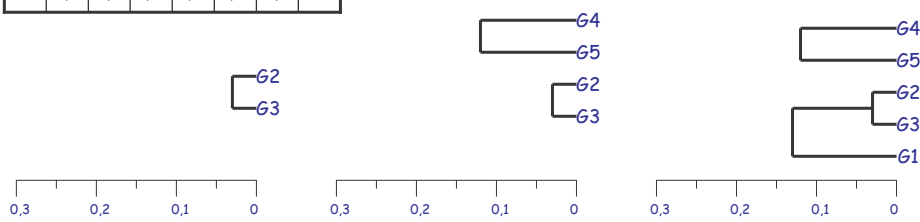
Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

Classification hiérarchique

	G1	G2	G3	G4	G5	G6	G7
G1	0						
G2	0,1	0					
G3	0,16	0,03	0				
G4	0,9	0,8	0,5	0			
G5	0,25	0,4	0,77	0,12	0		
G6	1,3	1,5	0,95	1,0	1,1	0	
G7	1,2	1,7	1,02	0,85	0,92	0,3	0

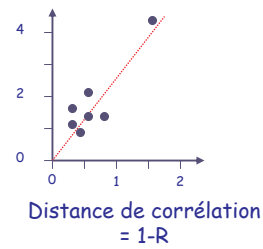
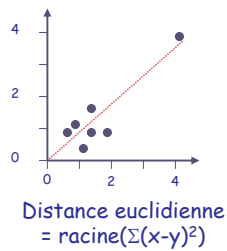
	G1	G2-G3	G4	G5	G6	G7
G1	0					
G2-G3	0,13	0				
G4	0,9	0,65	0			
G5	0,25	0,58	0,12	0		
G6	1,3	1,22	1,0	1,1	0	
G7	1,2	1,36	0,85	0,92	0,3	0

	G1	G2-G3	G4-G5	G6	G7
G1	0				
G2-G3	0,13	0			
G4-G5	0,57	0,61	0		
G6	1,3	1,2	1,0	0	
G7	1,2	1,3	0,87	0,3	0



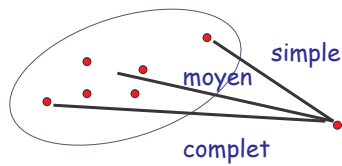
Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

Classification hiérarchique: distance



Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

Classification hiérarchique: lien



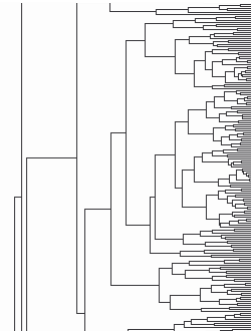
Quelle est la distance d'un gène à un groupe de gènes déjà constitué ?



simple



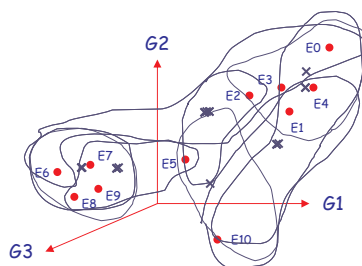
moyen



complet

Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

Nuées Dynamiques: K-means



Attribution aléatoire de chaque échantillon à un groupe

Calcul du centre de gravité des groupes

Attribution des échantillons au centre de gravité le plus proche

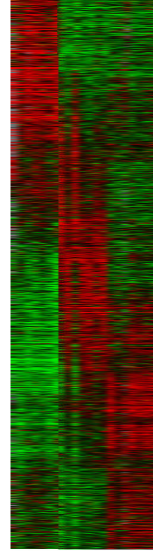
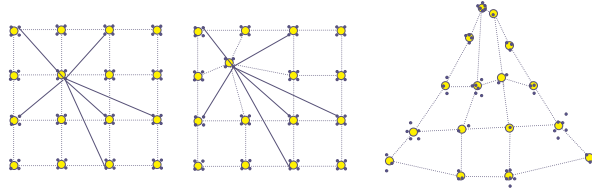
Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

Cartes de Kohonen

Self-Organizing Maps (SOM)

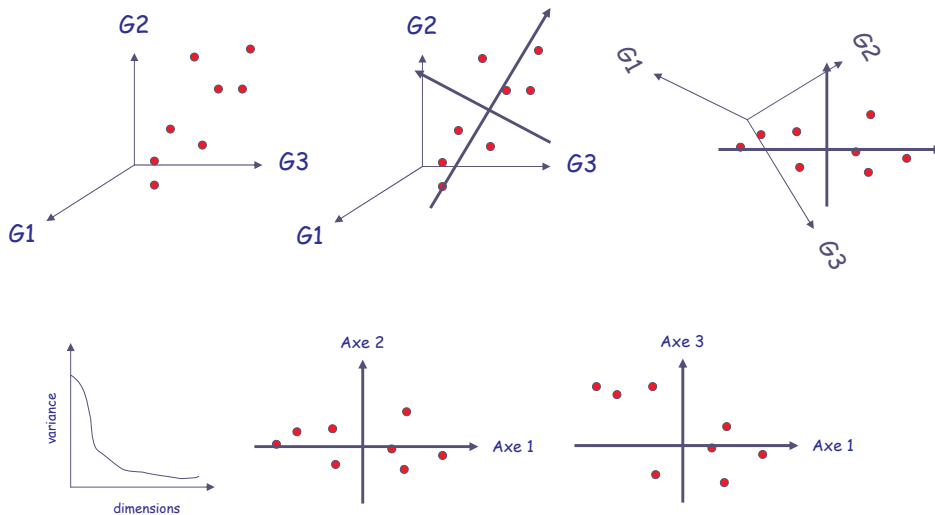
```

E0  E1  E2  E3  E4
E0  0
E1  0.3  0
E2  0.1  1.2  0
E3  2.4  8.6  5.4  0
E4  0.3  0.5  1.4  6.5  0
    
```



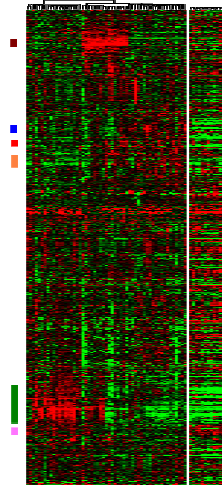
Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

Analyse en composantes principales



Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

Méthodes supervisées

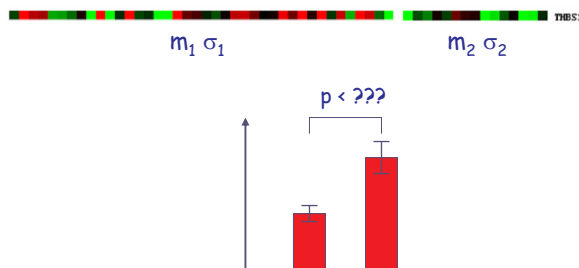
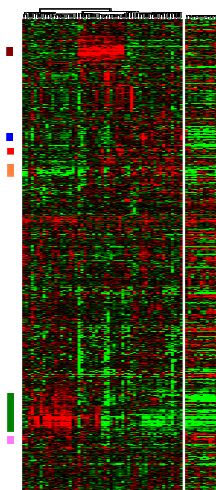


Méthodes Supervisées
Découverte de classes
Tests statistiques « classiques »
Utilisation des expressions corrélées
Score discriminant
Analyse linéaire discriminante LDA
Plus proches voisins KNN
Centroïdes
Réseaux de neurones ANN
Arbres de décision CART
Machines à support vecteur SVM

Problème des tests multiples !!!

Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

Problème des tests multiples

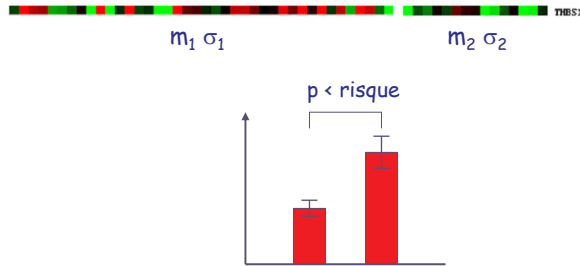
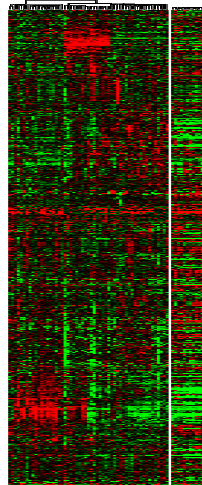


Approche laxiste:
- Test à 5 % de risque
- 10.000 paramètres
500 faux-positifs

Approche drastique:
Correction de Bonferoni:
Test à $5 \cdot 10^{-6}$ (5%/10.000)
Aucun vrai-positif (?)

Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

Mesure du taux de faux-positifs (FPR)

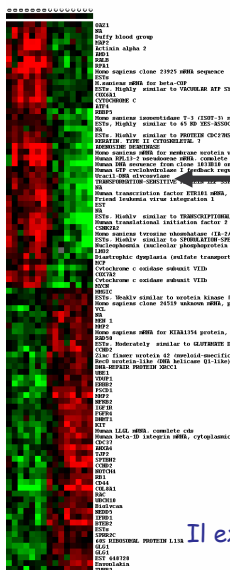


Faux-Positifs = Risque * nb tests

Il faut démontrer que Positifs \gg Faux-Positifs
Tous les tests inutiles génèrent des faux-positifs :
importance du filtrage

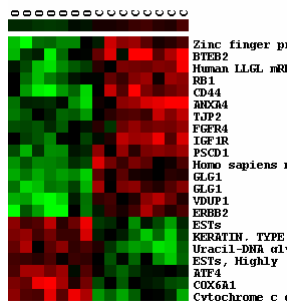
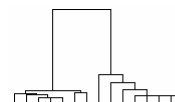
Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

Taux de Faux-Positifs



DS et seuil à 1/1000
cad 5 gènes faux positifs

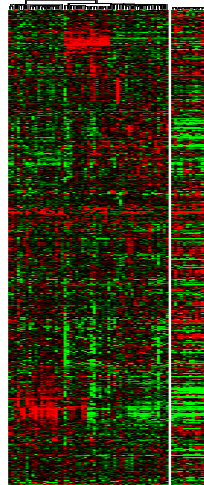
DS et seuil à 1/10000
cad 0,5 gènes faux positifs



Il existe un net surplus de gènes. Ce n'est pas un effet du hasard

Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

Solutions pour ne pas faire trop de tests



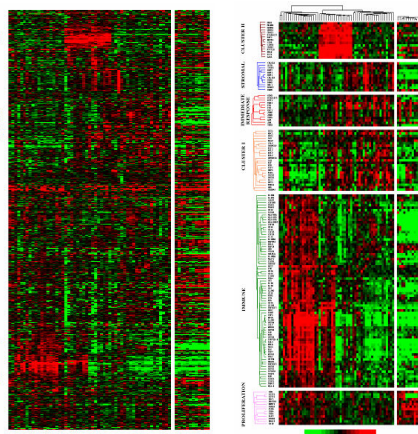
Ne pas faire de test	Aucun test
Filtrage des données	Moins de tests
Exploration visuelle	Aucun test
Connaissance préalable	Aucun test
Utilisation des expressions corrélées (clusters)	Quelques tests
Score discriminant	Aucun test

$$DS = \frac{m_1 - m_2}{s_1 + s_2}$$



Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

Tests sur les clusters

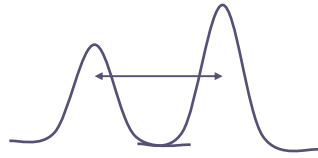


Il vaut mieux faire un seul test par groupe de gènes co-exprimés

10.000 gènes mais 6 groupes de coexpression (clusters) → 6 tests

Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

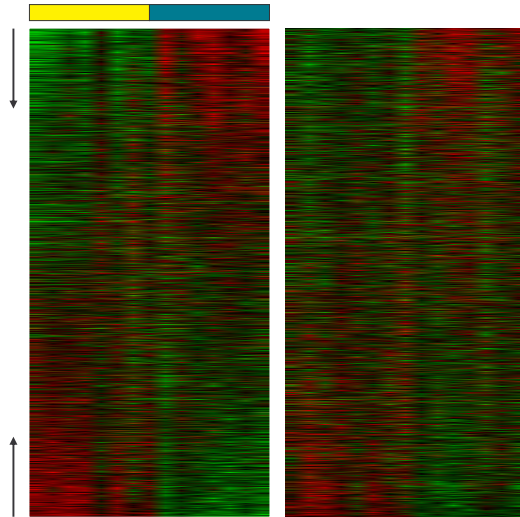
Score Discriminant; SAM ...



$$SD = \frac{m_1 - m_2}{s_1 + s_2}$$

$$SAM = \frac{m_1 - m_2}{s_1 + s_2 + \epsilon}$$

On sélectionne ensuite les meilleurs gènes.
La significativité statistique est repoussée à plus tard



Classes réelles

Classes aléatoires

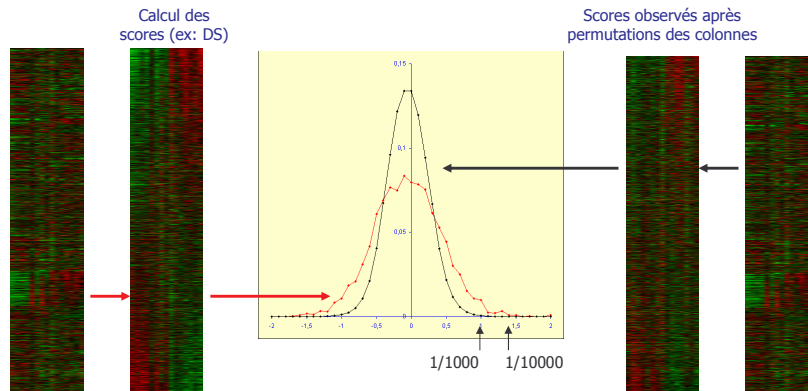
Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

Méthodes de validation

Méthodes de Validation
Significativité, Puissance ...
Rééchantillonnage
(bootstrap, permutations de colonnes)
Utilisation du bruit
Leave-one out
Apprentissage-Validation (Prédicteur)

Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

Rééchantillonnage: permutation de colonnes

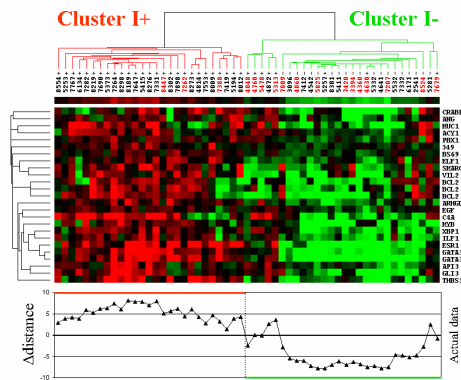


Intéressant quand la distribution n'est pas gaussienne (elle l'est très souvent).
On cherche un surplus de positifs // faux-positifs

Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

Construction d'un prédicteur

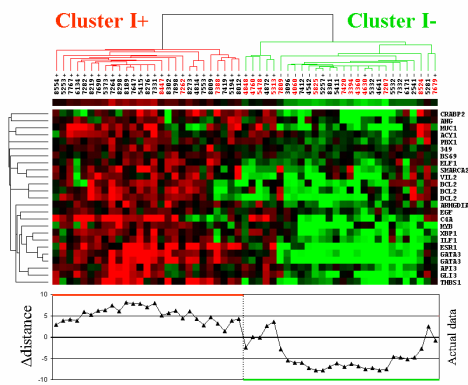
Bertucci et al. (2002) Hum. Mol. Genet. 11: 863-872



On construit un profil moyen rouge et un profil moyen vert.
On mesure la distance de chaque échantillon à ces 2 profils.

Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

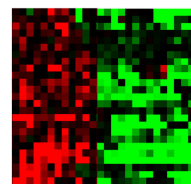
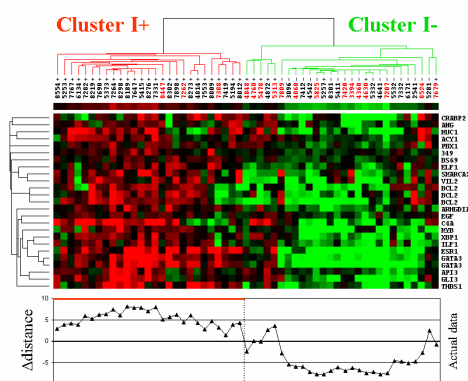
Leave-one out



Pour chacun des échantillons:
On isole cet échantillon.
On construit un prédicteur avec le reste des données.
On prédit l'échantillon isolé.
On estime ainsi la qualité (cohérence) du prédicteur

Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

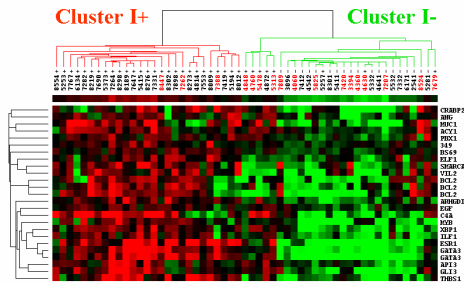
Prédiction-Validation



On isole une partie des données (1/3 est un bon compromis).
On construit un prédicteur avec le reste des données.
On mesure la qualité du prédicteur avec les échantillons isolés.

Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

Validation indépendante



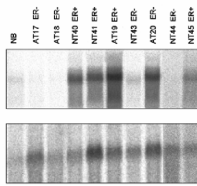
On prédit quelque chose au moyen d'un jeu de mesure, puis on le revalide au moyen de:

Une seconde étude indépendante

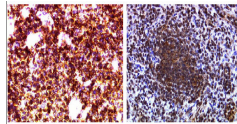
Une autre méthode mesurant la même propriété (Northern Blot, PCR quantitative ...)

Une autre méthode l'expression des gènes à un autre niveau (Western Blot, Immunohistochimie ...)

...



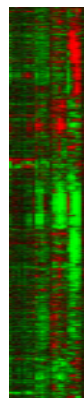
Northern Blot



Immunohistochimie

Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

Annotation



- Cyclin A
- BUB1 mitotic kinase
- Cyclin B1
- SOCS-1
- Ki67
- p59CDC
- PLK=polo-like kinase
- CIP2/Cdi1/KAP1
- aurora-related kinase 1
- p16
- Thymidine kinase
- CDC21 homologue
- RAD54
- Dihydrofolate reductase

1 cluster = 1 fonction
Oui mais laquelle ?

2 solutions:

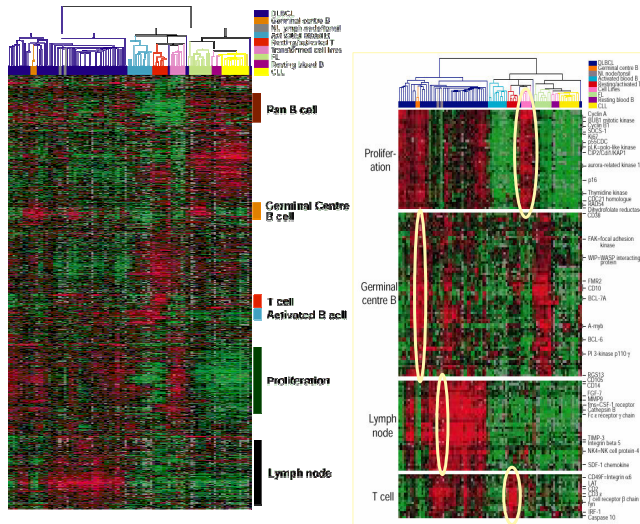
Expérimentale: utiliser des modèles et des cellules purifiées

Bioinformatique: utiliser des ontologies

Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

Annotation expérimentale

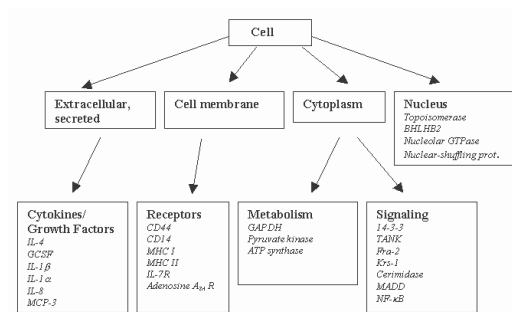
Alizadeh et al. (2000) Nature 403: 503-511



L'utilisation de modèles simples (lignées ou cellules en culture, stimulées ou non) permet d'établir les fonctions des signatures transcriptionnelles.

Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

Ontologie



Vocabulaire contrôlé et structuré associé à des objets

Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

Gene Ontology (www.geneontology.org)

Ashburner et al. 2000 Nat. Genet. 25: 25-29.

The screenshot shows the Gene Ontology Consortium website with the following content:

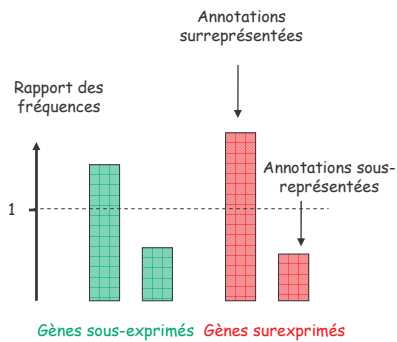
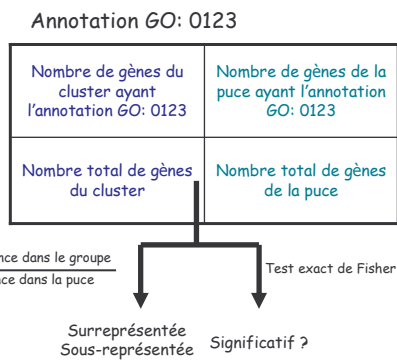
- GENE ONTOLOGY CONSORTIUM**
- OBO** (Open Biomedical Ontologies)
- What is the Gene Ontology?** / **Download the Ontologies**
- Goal: "The goal of the Gene Ontology™ (GO) Consortium is to produce a controlled vocabulary that can be applied to all organisms even as knowledge of gene and protein roles in cells is accumulating and changing. GO provides three structured vocabularies of defined terms to describe gene product attributes. GO is one of the controlled vocabularies of the [Open Biological Ontologies](#)."
 - Submit new GO term suggestions via the [Curator Request Tracker](#) at submit@geneontology.org
 - Send comments and questions to go@geneontology.org

The hierarchical tree on the right shows the following structure:

- Gene_Ontology
 - cellular_component
 - cell
 - bud
 - cell body
 - cell fraction
 - intracellular
 - RNA polymerase complex
 - UDP-N-acetylglucosamine-peptide N-acetylglucosaminyltransferase complex
 - apical part of cell
 - cytoplasm
 - DNA polymerase complex
 - DNA topoisomerase IV complex
 - pyruvate dehydrogenase complex
 - eukaryotic 48S initiation complex
 - eukaryotic 43S preinitiation complex
 - cytosolic small ribosomal subunit (sensu Eukarya) *
 - eukaryotic translation initiation factor 2 complex
 - eukaryotic translation initiation factor 3 complex
 - ribosome
 - cytosolic ribosome (sensu Archaea)
 - cytosolic large ribosomal subunit (sensu Archaea)
 - cytosolic small ribosomal subunit (sensu Archaea)
 - cytosolic ribosome (sensu Bacteria)
 - cytosolic ribosome (sensu Eukarya)
 - cytosolic large ribosomal subunit (sensu Eukarya)
 - cytosolic small ribosomal subunit (sensu Eukarya) *
 - large ribosomal subunit
 - small ribosomal subunit
 - cytosolic small ribosomal subunit (sensu Archaea)
 - cytosolic small ribosomal subunit (sensu Bacteria)
 - cytosolic small ribosomal subunit (sensu Eukarya) *

Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN

Statistiques GO



GO Miner
Fatigo

Houlgatte Rémi (houlgatte@tagc.univ-mrs.fr) Analyse des données de puces à ADN