



UNIL | Université de Lausanne

Département de biologie
moléculaire végétale

Introduction à la Génomique Fonctionnelle

Cours aux étudiants de BSc Biologie 3ème année

Philippe Reymond, MER

PLAN DU COURS

- Séquençage des génomes
- Fabrication de DNA microarrays
- Autres méthodes globales d'analyse de l'expression génétique
- Analyse des données, exemples de résultats

http://www.unil.ch/dbmv/page26947_en.html

GENOMIQUE

Analyse globale du génome d'un organisme



Tous les gènes et régions intergéniques

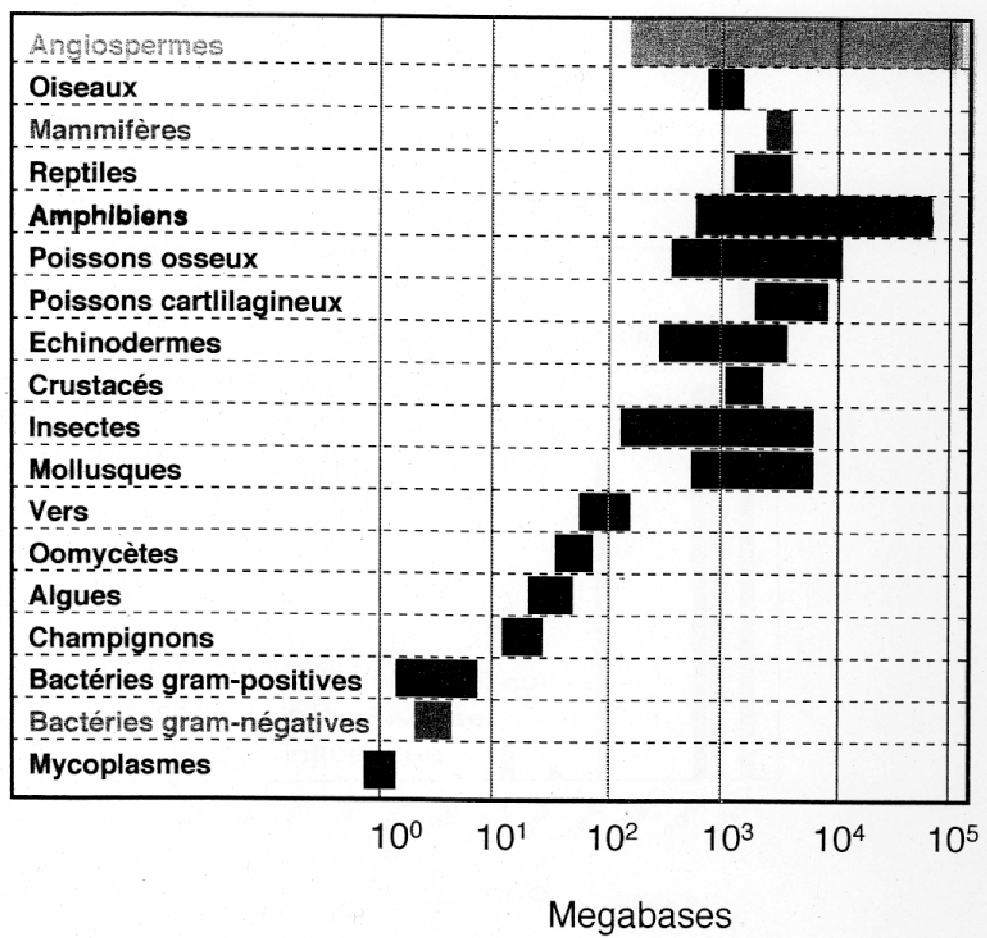
1. Génomique structurelle

- Organisation et position des gènes, taille du génome
- Comparaison des génomes entre différents organismes
- Séquençage de l'ADN et analyse des séquences

2. Génomique fonctionnelle

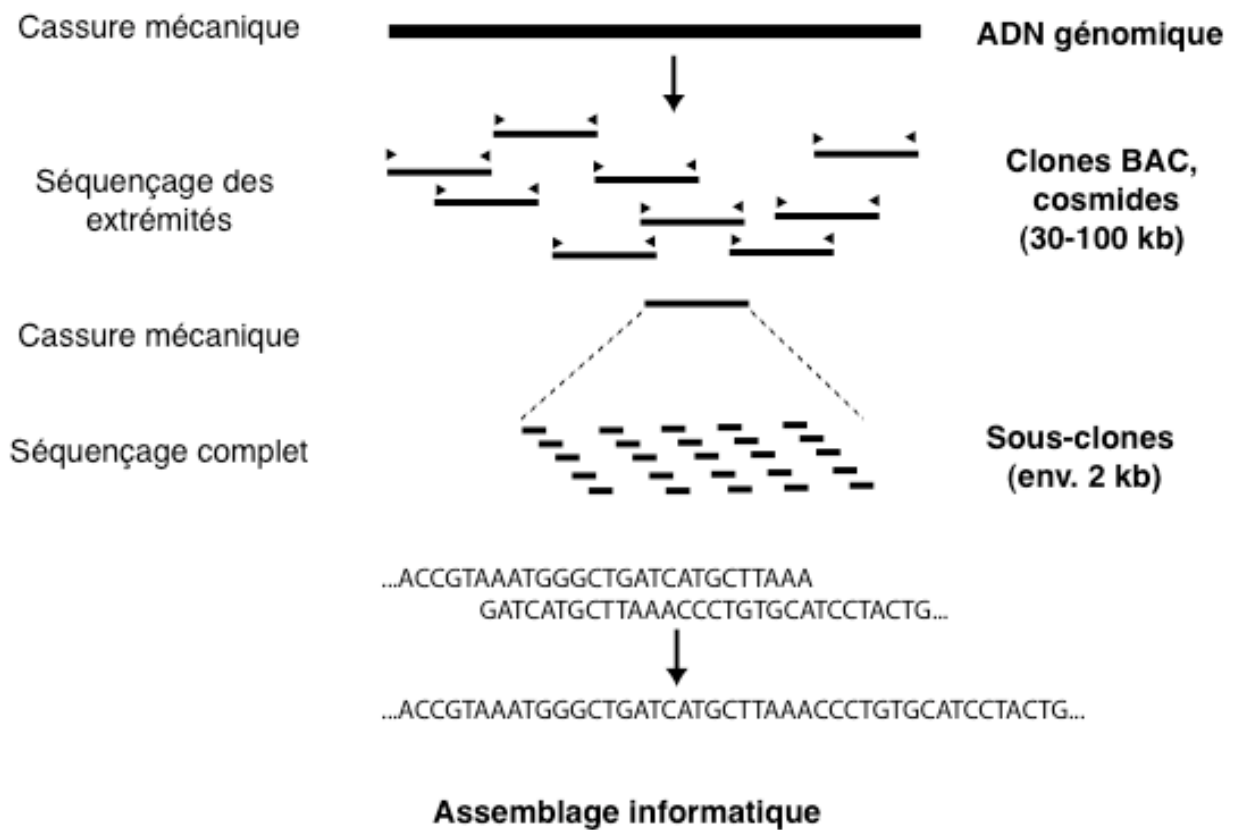
- Fonction des gènes
- Analyse globale de l'expression génétique
- Disruption systématique des gènes (Knock-outs, RNAi)

Taille des génomes



Méthode de séquençage de génomes

"Shotgun sequencing"



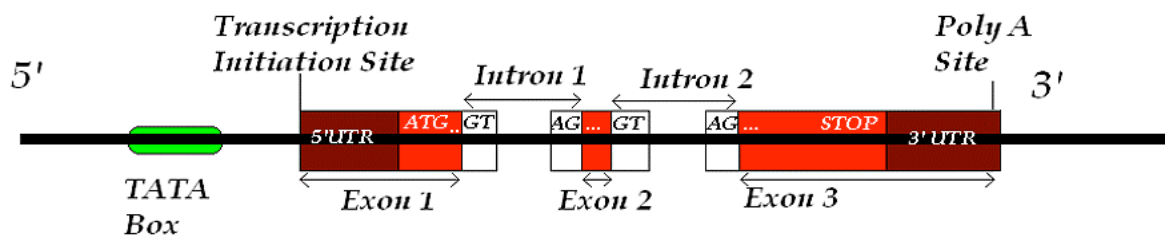
Séquençage du génome humain

HumanGenomeProject	Celera
<p>QUI SÉQUENCER?</p> <p>Plusieurs donneurs anonymes (5-10) Hommes et femmes</p> <p><u>ADN</u></p> <p>Sang et sperme</p> <p>COMMENT ?</p> <p><u>Séquençage</u></p> <p>"BAC to BAC"</p> <p>20 groupes gels et capillaires</p> <p>max 200'000 réactions/24h</p> <p>max 1000 b/sec</p> <p>23.1 Mia bases seq. (7.5 génomes)</p> <p><u>Assemblage</u></p> <p><u>Durée</u></p> <p><i>4 années</i></p>	<p>5 donneurs anonymes</p> <p>2 hommes, 3 femmes</p> <p>Noir américain, asiatique, mexicain 2 caucasiens</p> <p>Sang et sperme</p> <p>"Whole genome shotgun"</p> <p>65 techniciens 250 ABI Prism 3700 Analyzer (96 capillaires)</p> <p>175'000 réactions/24h</p> <p>1100 b/sec</p> <p>14.8 Mia bases séquencées (5.1 génomes)</p> <p>10 ordinateurs (4 processeurs, 4 Gb RAM) 1 ordinateur (16 processeurs, 64 Gb RAM) Calcul 20'000 heures</p> <p><i>9 mois</i></p>

Annotation du génome

TGACGCAATGAACGATGGCTAGGCTTAGCTTTTGGCCAGCATTACCCGGTACC
TGTTACTGCGATTGGCTAACCAGGCCGGATTTCCAAGGTTTCGCCAGGGGCTAA
TTGACACCACCACCTTACCTGGTATTCCCTAATGGTAAGCGCGATGCTA....

algorithmes de prédiction



Comment améliorer les prédictions?

- génomique comparative
- analyse de cDNAs complets
- entraînement des algorithmes sur des gènes connus
- sondes oligonucéotides scannant tout le génome
- vérification expérimentale

Comparison of mouse and human genomes followed by experimental verification yields and estimated 1'019 additional genes. Guido et al. (2004) PNAS 100:1140-1145

Nombre de gènes

<i>Organisme</i>	<i>Nb de gènes prédits</i>	<i>% gènes connus</i>
M. genitalium	470	69%
H. influenza	1'709	58%
E. coli	4'288	62%
S. cerevisiae	6'034	63%
Thalassiosira pseudonana	11'242	50%
D. melanogaster	13'601	46%
C. elegans	18'424	42%
A. thaliana	≈28'000	69%
M. musculus	≈30'000	50%
H. sapiens	≈30'000	50%

Quelle est la taille minimum du génome pour qu'une cellule fonctionne?

Mutations systématiques par transposons du génome de *Mycoplasma genitalium*:

265-350 gènes sont essentiels pour une croissance en laboratoire (100 n'ont pas de fonction connue!)

Hutchinson et al. (1999) Science 286: 2165-2169

Séquençage des génomes

Organismes modèles

Escherichia coli

Blattner et al. (1997) Science 277: 1453-1462

Saccharomyces cerevisiae

Goffeau et al. (1996) Science 274: 546-567

Caenorhabditis elegans

The C. elegans consortium (1998) Science 282: 2012-2046

Drosophila melanogaster

Adams et al. (2000) Science 287: 2185-2195

Arabidopsis thaliana

The Arabidopsis genome initiative (2000) Nature 408: 796-815

Homo sapiens

International Human Genome Sequencing Consortium (2001) Nature 409: 860-921

Venter et al. (2001) Science 291: 1304-1351

Mus musculus

Mouse Genome Sequencing Consortium (2002) Nature 420: 520-562

Takifugu rubripes

Aparicio et al. (2002) Science 297: 1301-1310

Rattus norvegicus

Rat Genome Sequencing project Consortium (2004) Nature 428: 493-521

Gallus gallus

International Chicken genome Sequencing Consortium (2004) Nature 432: 695-716

Pan troglodytes (chimpanze)

The Chimpanzee Sequencing and Analysis consortium (2005) Nature 437: 69-87

Autres (en cours)

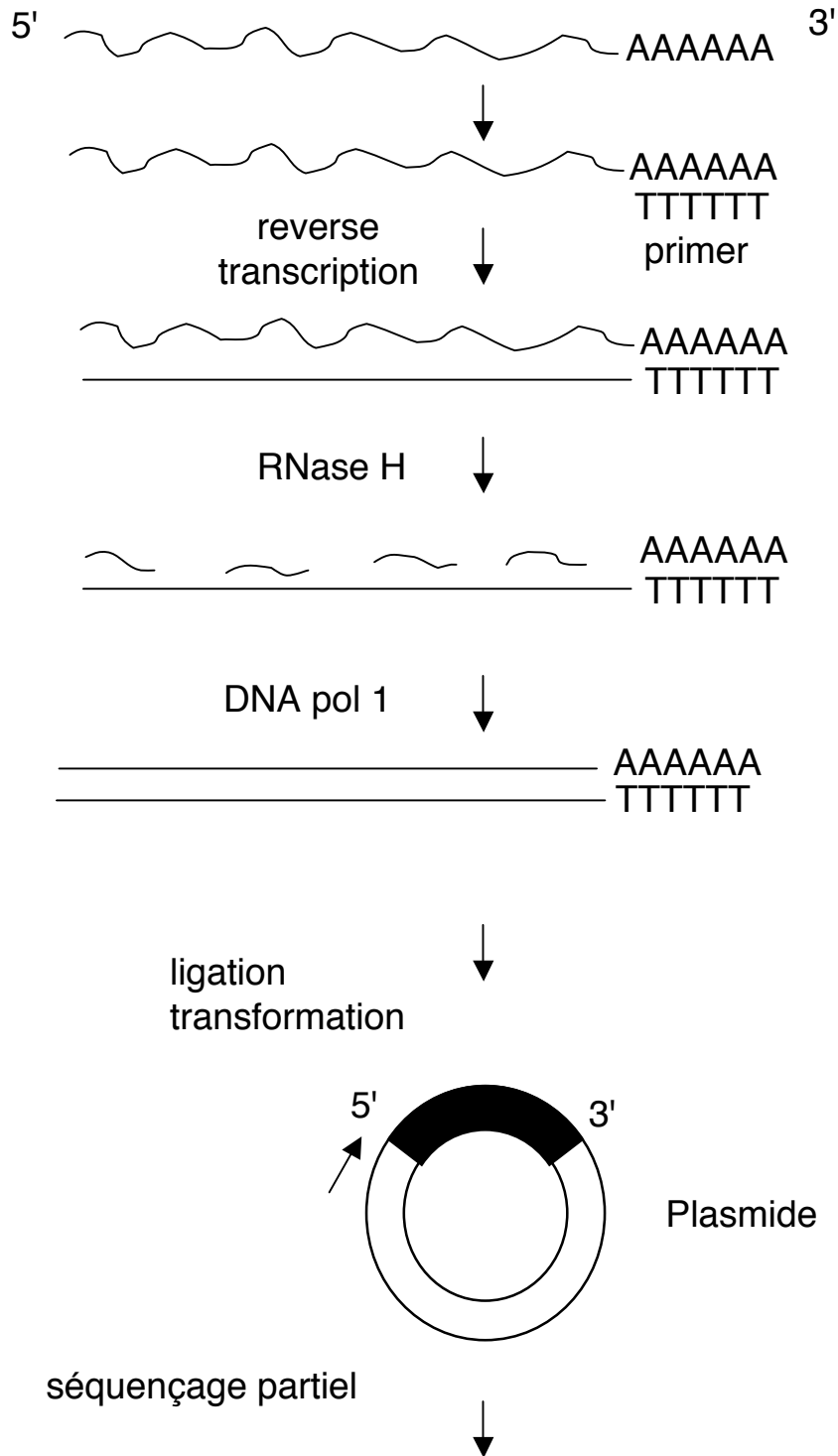
http://www.genomenetwork.org/resources/sequenced_genomes/genome_guide_p1.shtml

<http://www.ncbi.nlm.nih.gov/Genomes/index.html>

<http://hgsc.bcm.tmc.edu/projects/>

<http://www.sanger.ac.uk/Projects/>

EST (Expressed Sequence Tag)

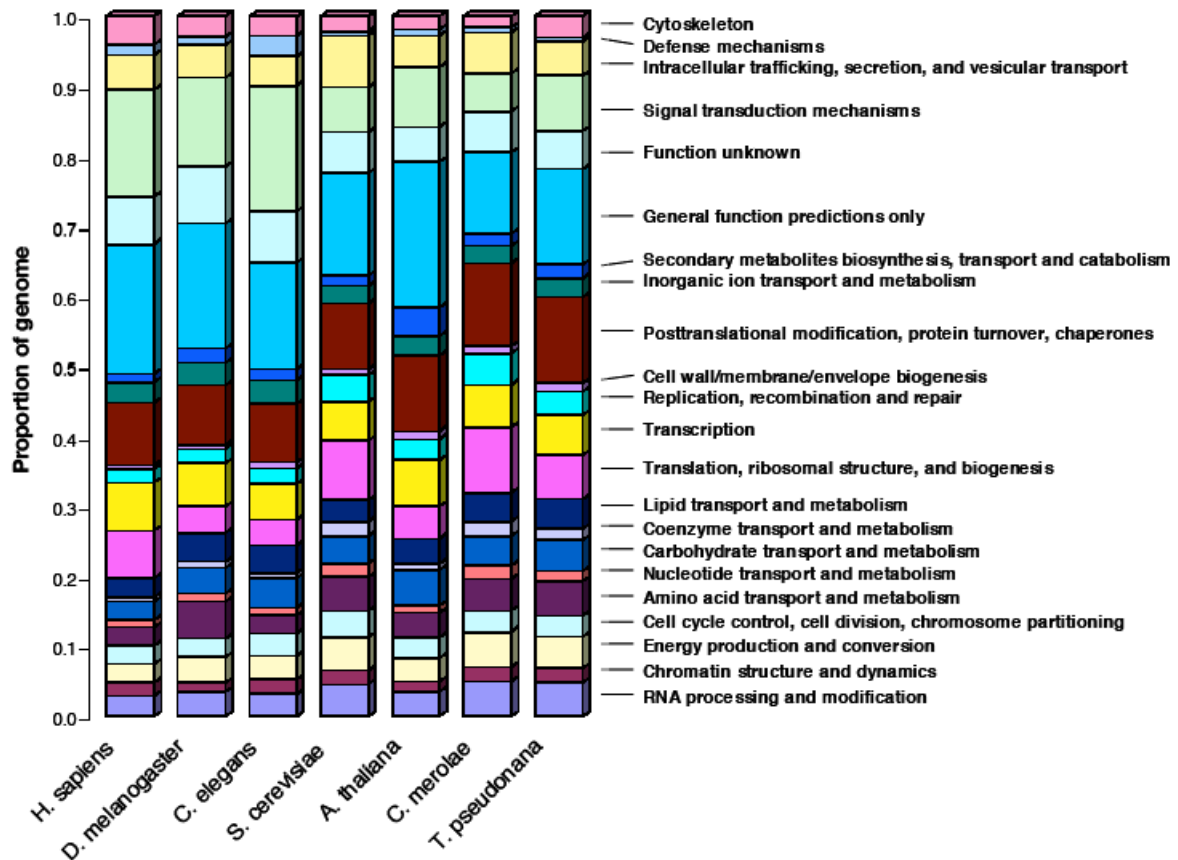


vecteur

cDNA

TGACGCAATGAACGGCTAGGCTTAGCTTTTGGCCAGCATTACCCGGTACC
 TGT TACTGCGATTGGCTAACCAGGCCGGATTTC CAAGGTT CGCCAGG
 TTGACACCACCACCTTACCTGGTATTCCC

Comparaisons de génomes (homme-mouche-ver-levure-plante-algue)



Les fonctions cellulaires élémentaires – métabolisme de base, réplication, transcription de l'ADN, traduction- sont présentes partout, sont apparues une fois dans l'évolution et sont restées fixées depuis.

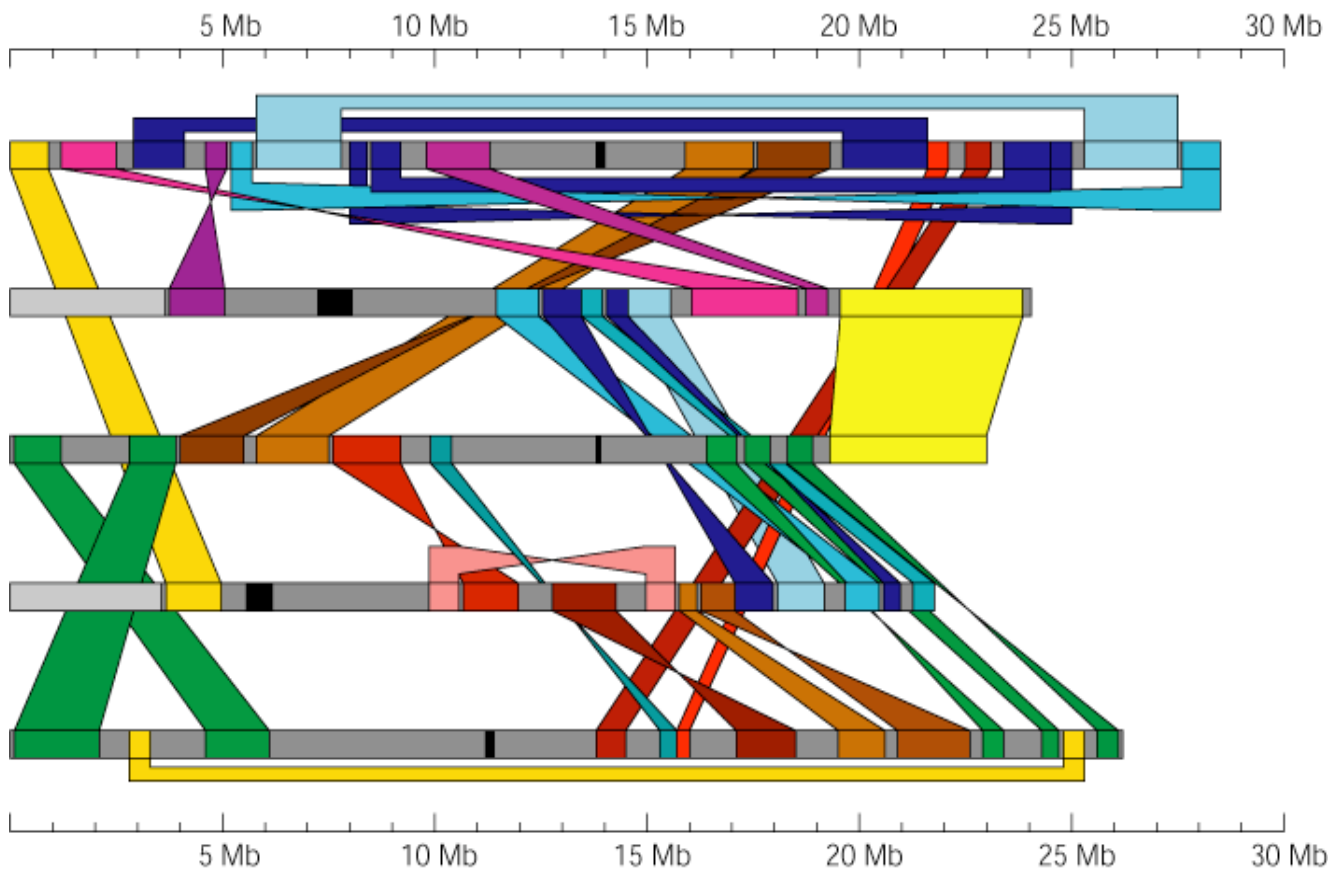
Seulement 94 familles de protéines (sur 1278) sont uniques aux vertébrés.

La plus grande différence entre l'être humain et le ver ou la mouche réside dans la complexité des protéines : il y a plus de domaines par protéine et beaucoup plus de combinaisons de domaines.

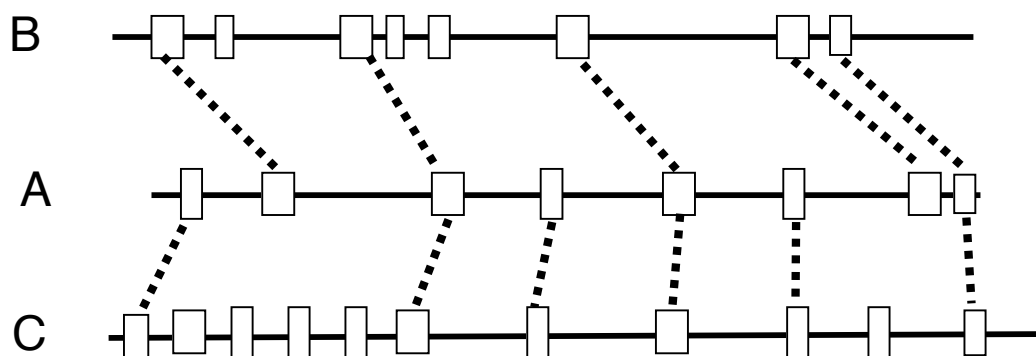
Nouvelles familles de protéines ou expansion de familles chez l'être humain:

- réponse immunitaire
- développement , structure et fonction du système nerveux
- signalisation intra- et intercellulaire dans le développement et l'homeostasie
- système sanguin
- apoptose

Duplications du génome d'Arabidopsis



Nature (2001), 408: 796-815



3 duplications: 221 Mio années

162 Mio années

75 Mio années

(calcul: 6.1 mutations synonymes/Mia années/site)

PNAS (2002), 99:13627-13632

Les puces à ADN (Microarrays, DNA chips)

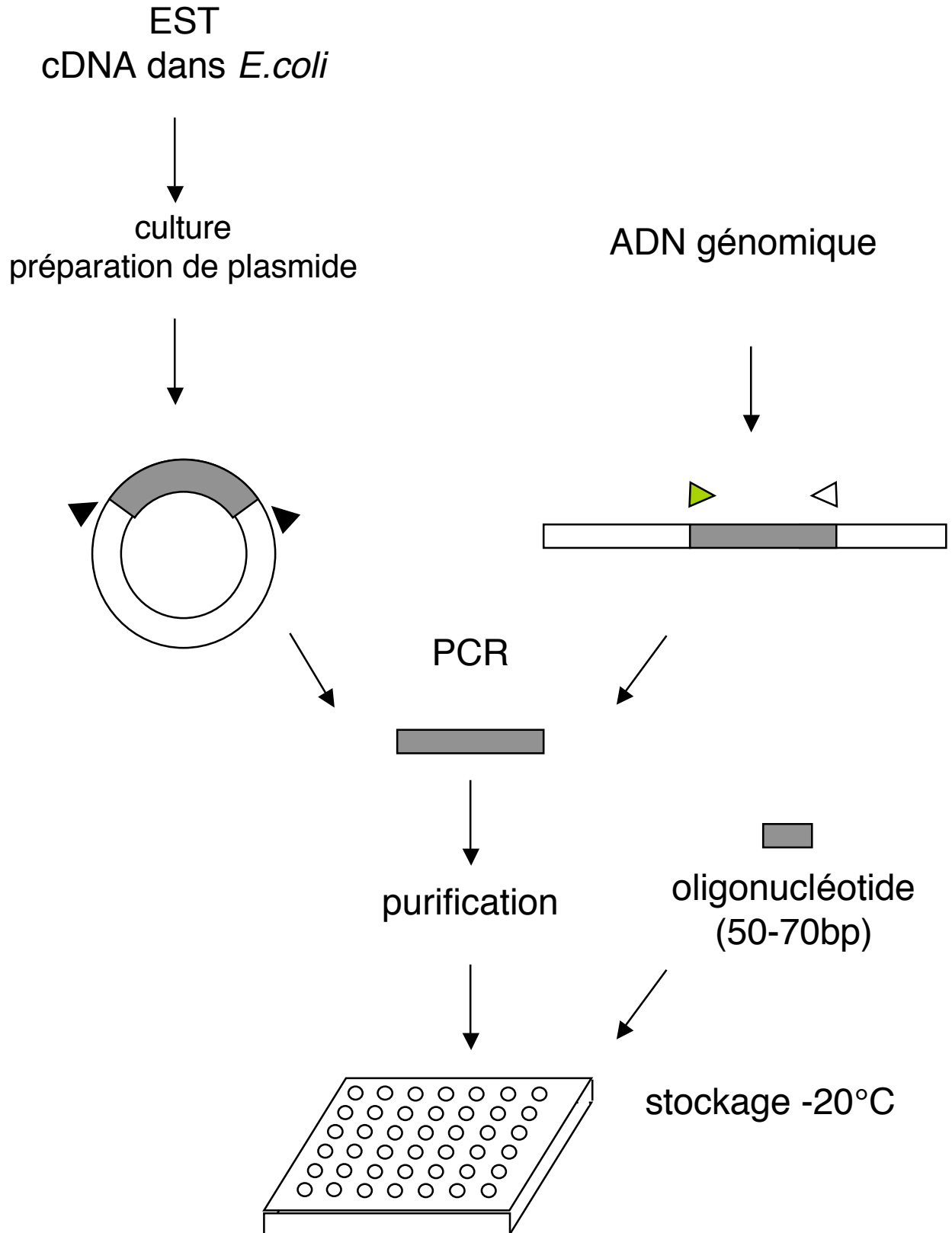
M. Schena, D. Shalon, R. Davis, P. Brown (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270: 467-470

Lipshutz, R. J.; Morris, D.; Chee, M.; Hubbell, E.; Kozal, M. J.; Shah, N.; Shen, N.; Yang, R.; Fodor, S. P. (1995) Using oligonucleotide probe arrays to access genetic diversity. BioTechniques 19(3), 442-7

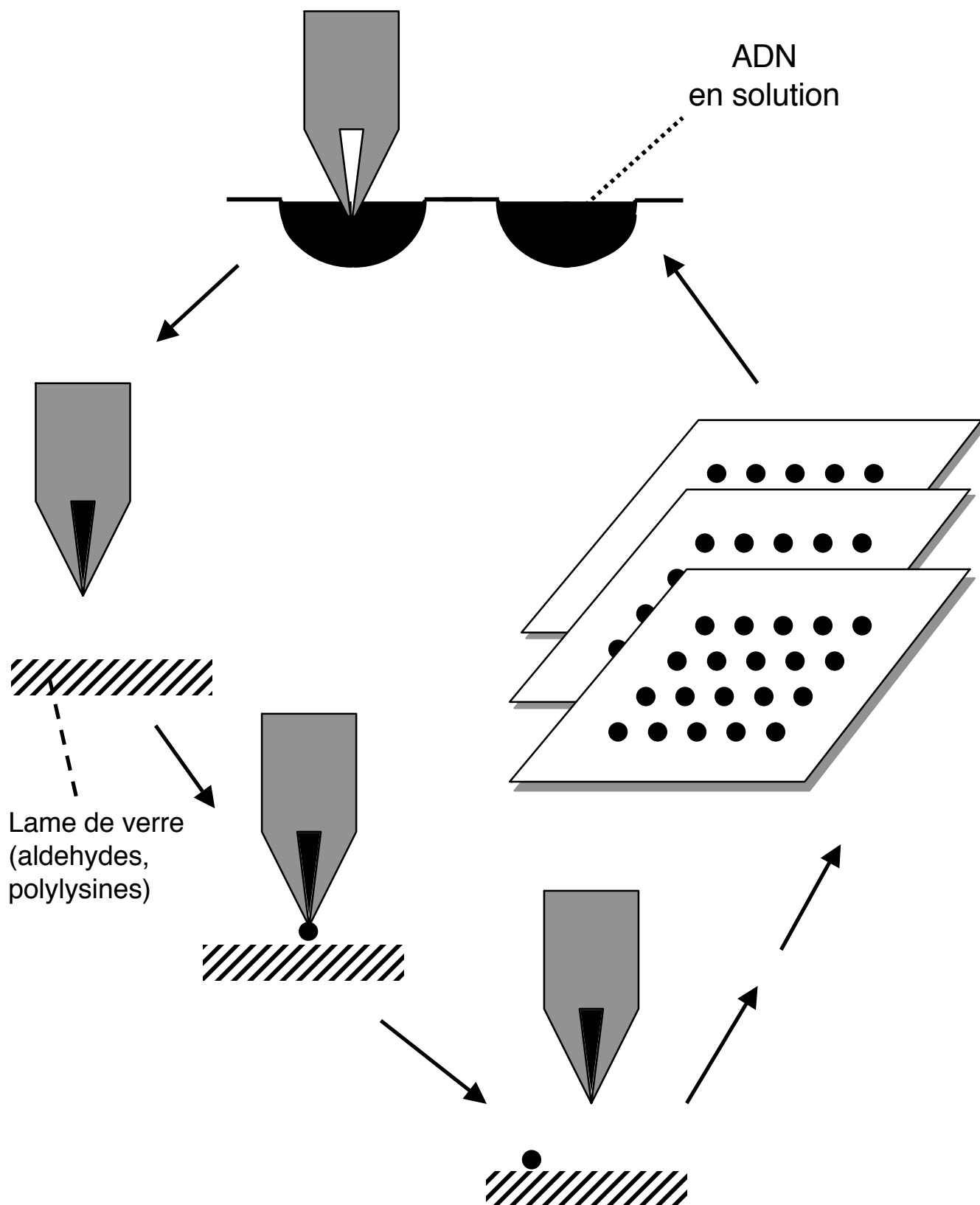
Applications

- Mesure globale de l'expression génétique
Recherche fondamentale et appliquée dans tous les domaines
Diagnostic médical, nouveaux médicaments
- Analyse d'amplifications et de délétions de gènes
Caractéristique de certains cancers
- Analyses du génome
Single Nucleotide Polymorphisms (SNP), mutations, séquences répétées
- *Etc...*

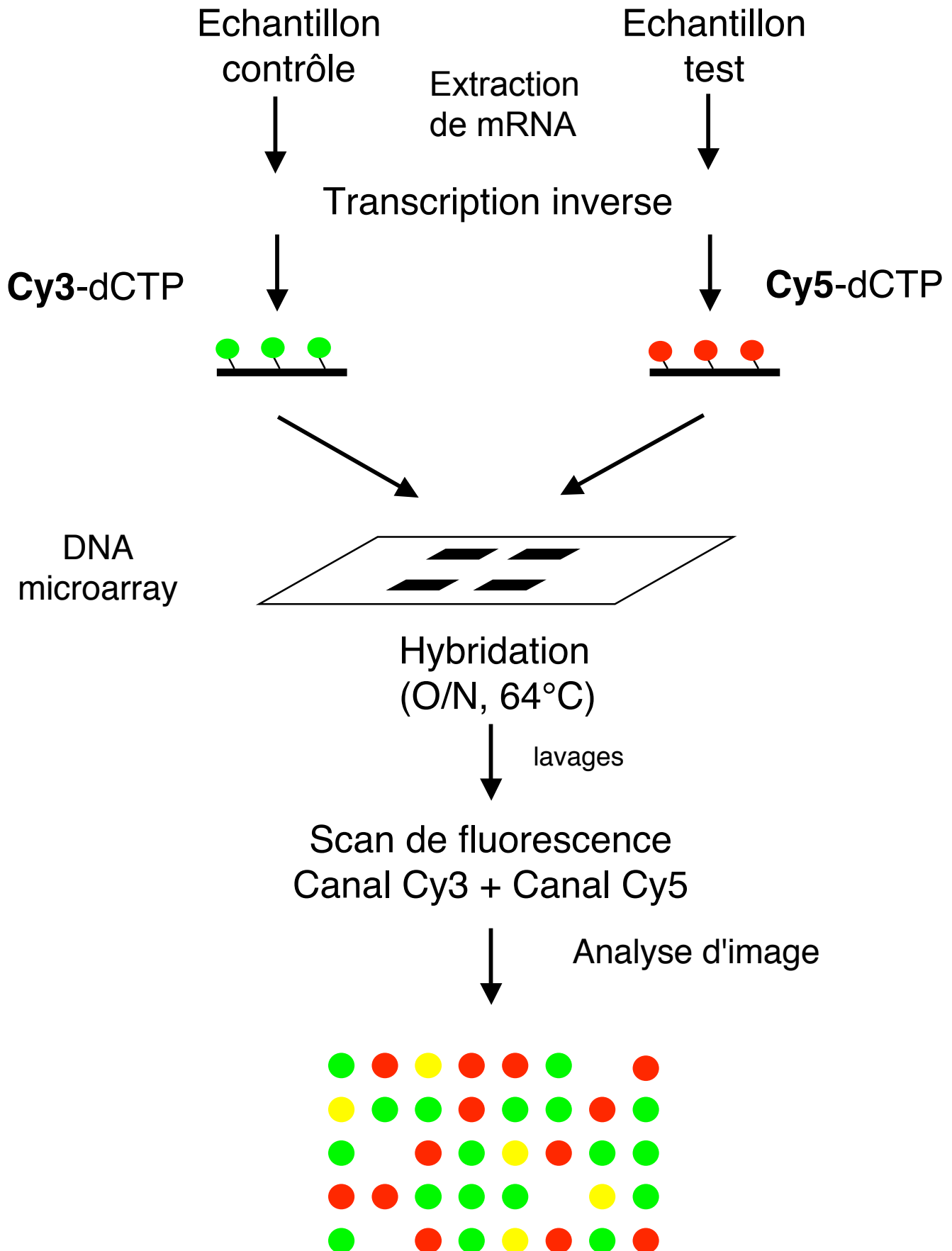
Préparation d'ADN pour microarrays



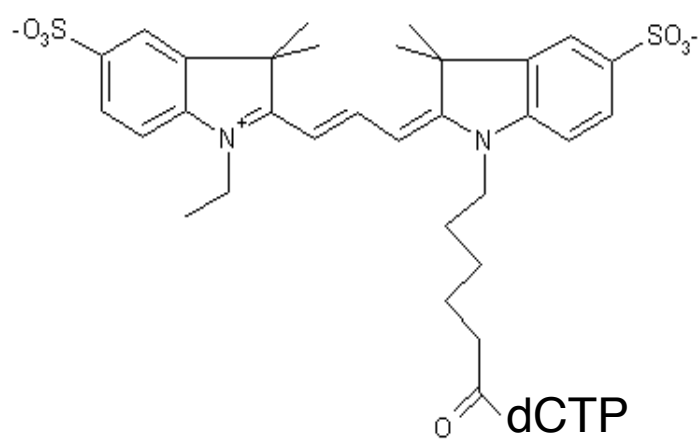
Impression de puces à ADN



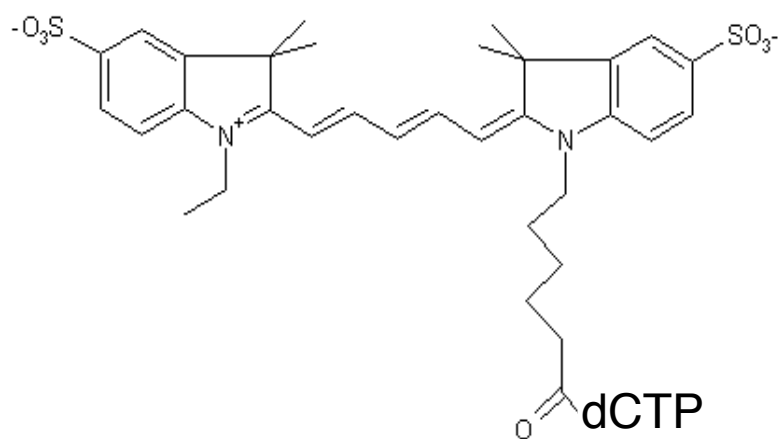
Marquage et hybridation



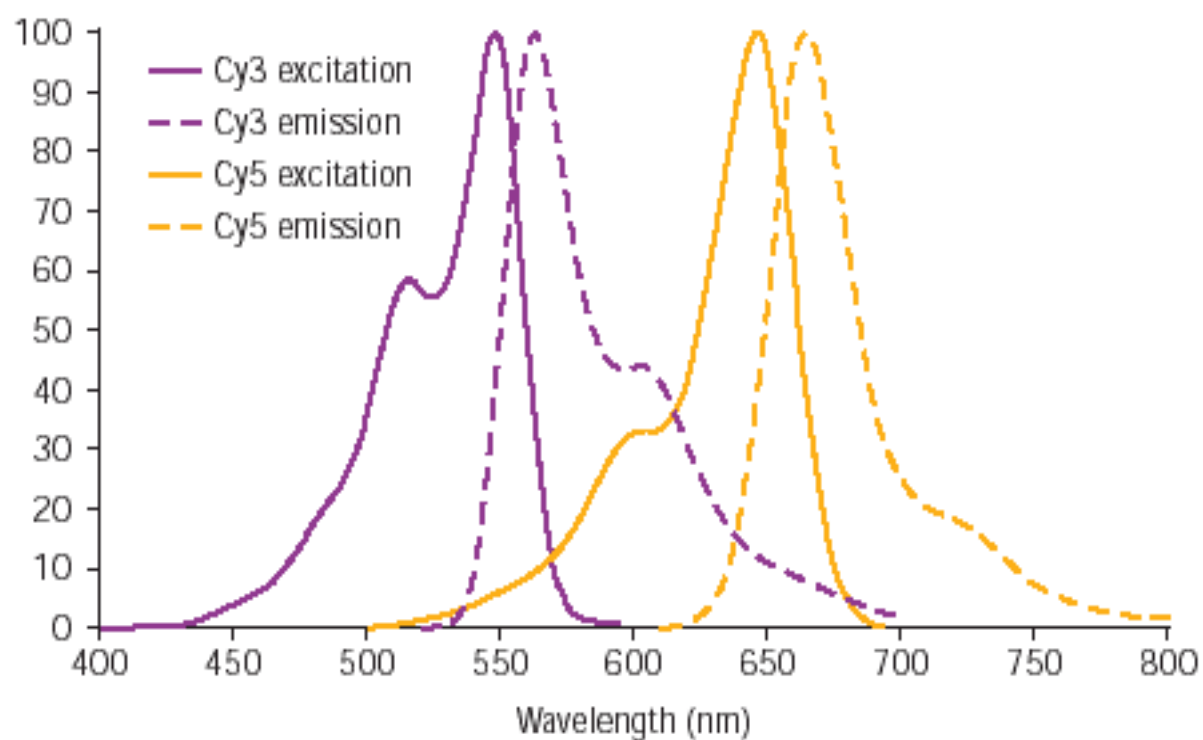
Cy3-dCTP



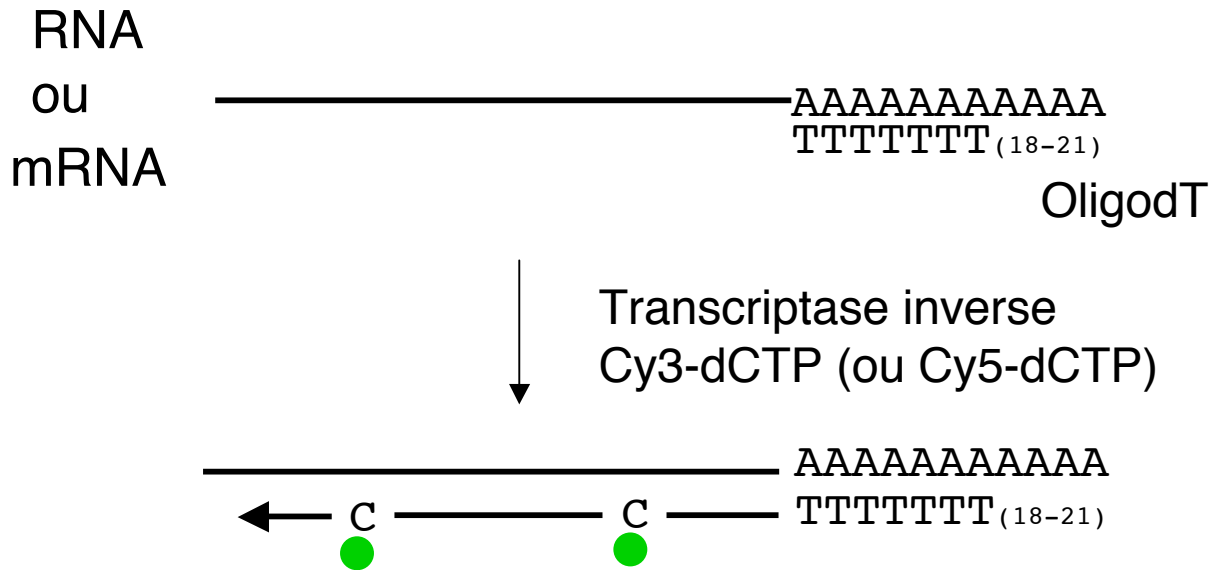
Cy5-dCTP



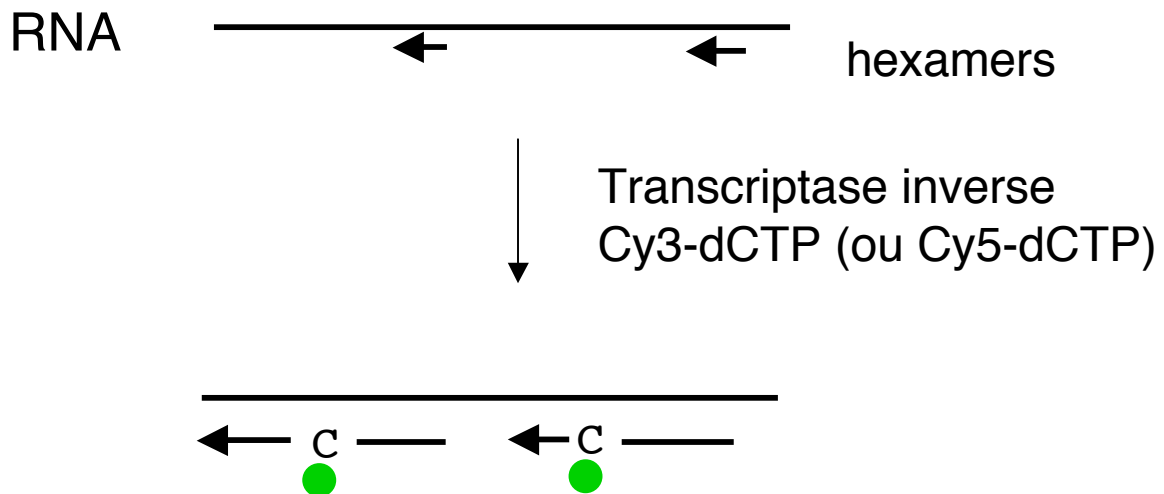
Fluorescence



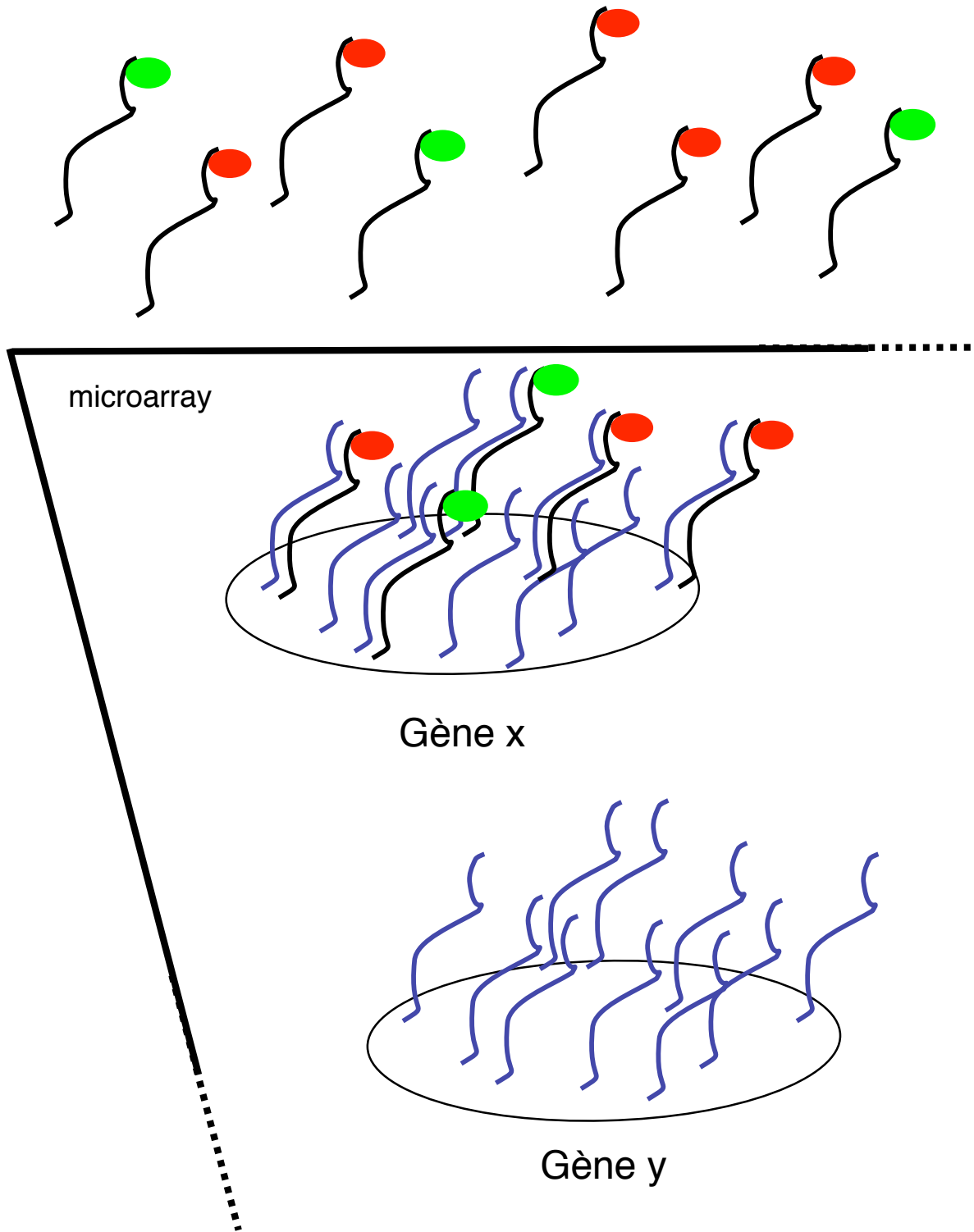
Eucaryotes



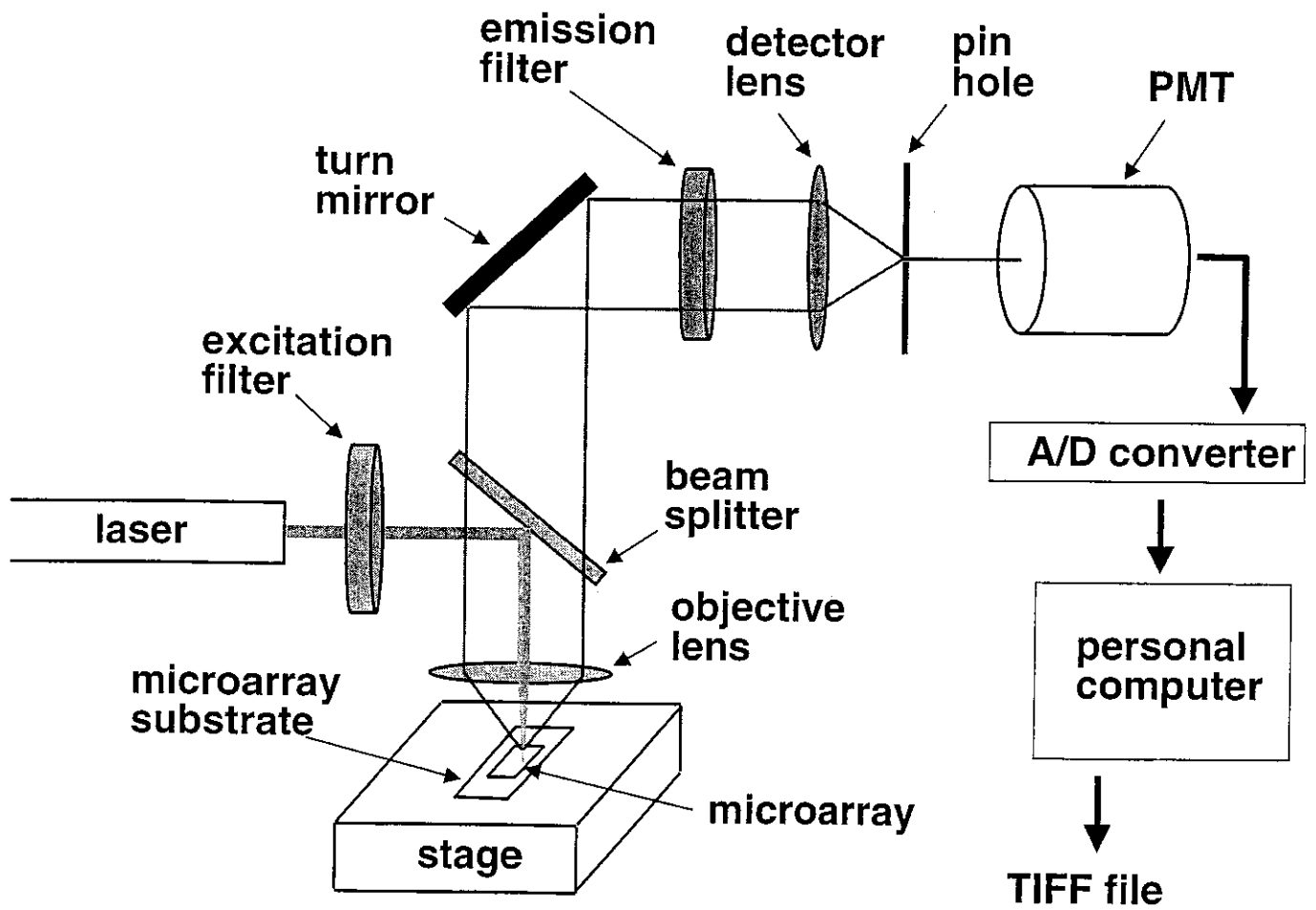
Procaryotes



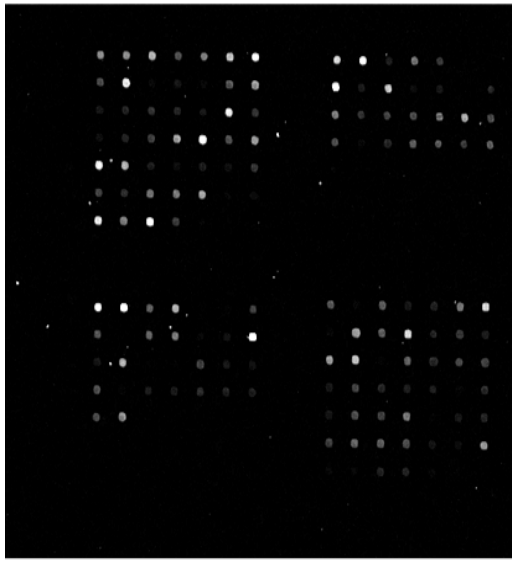
Hybridation



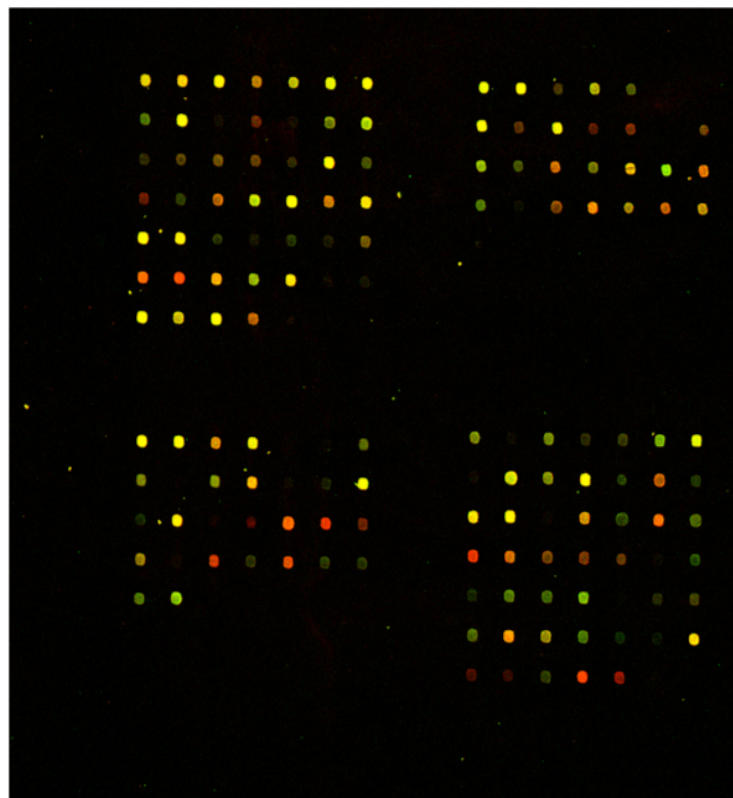
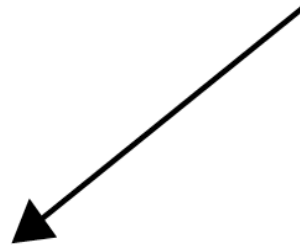
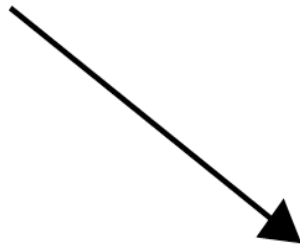
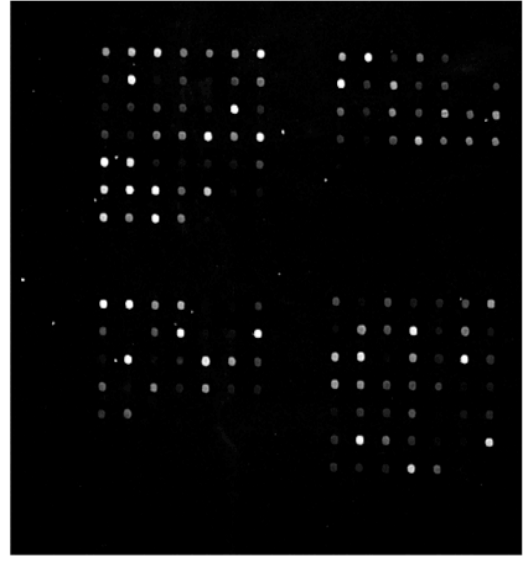
Scanner de fluorescence



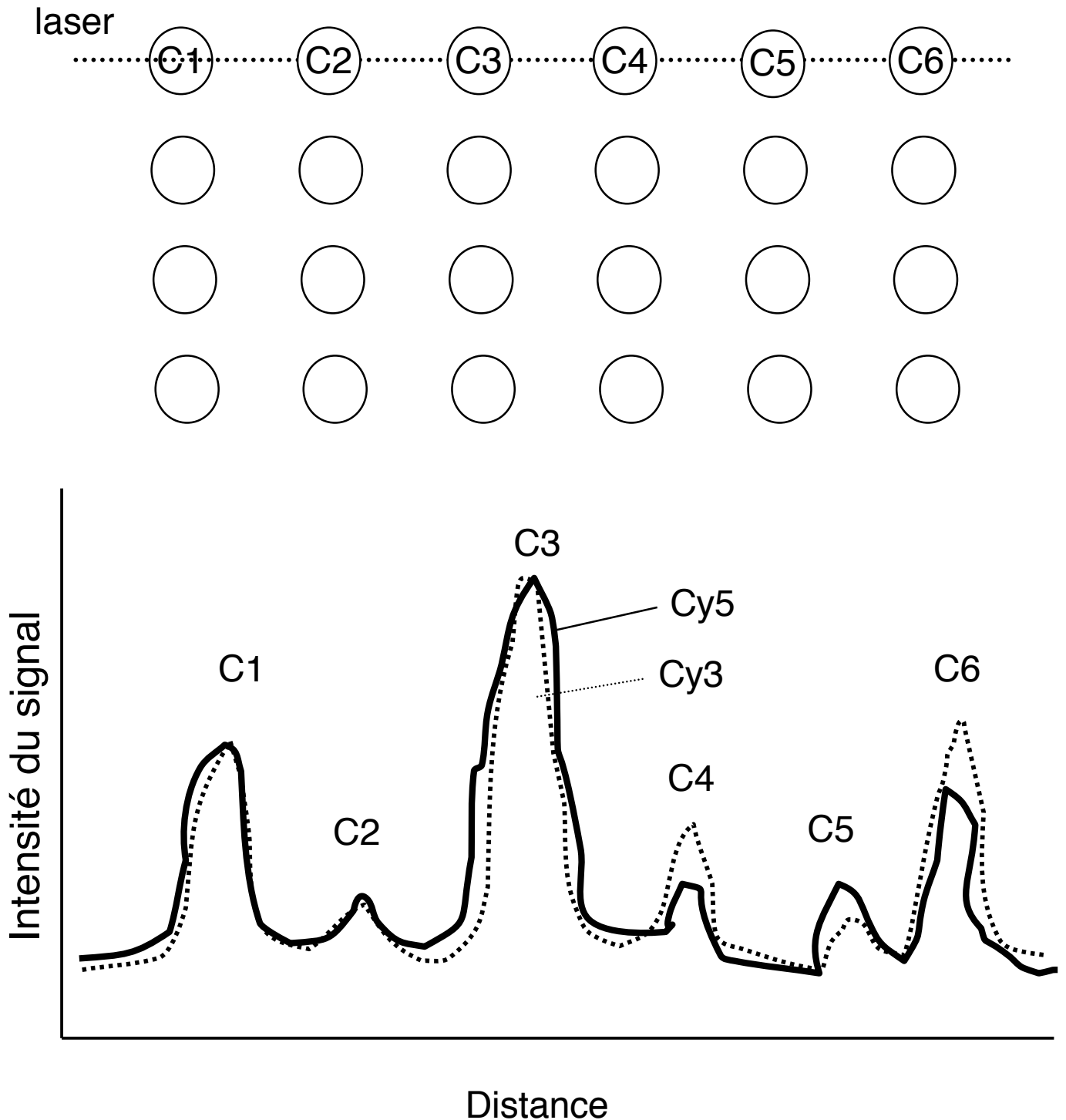
Cy3



Cy5

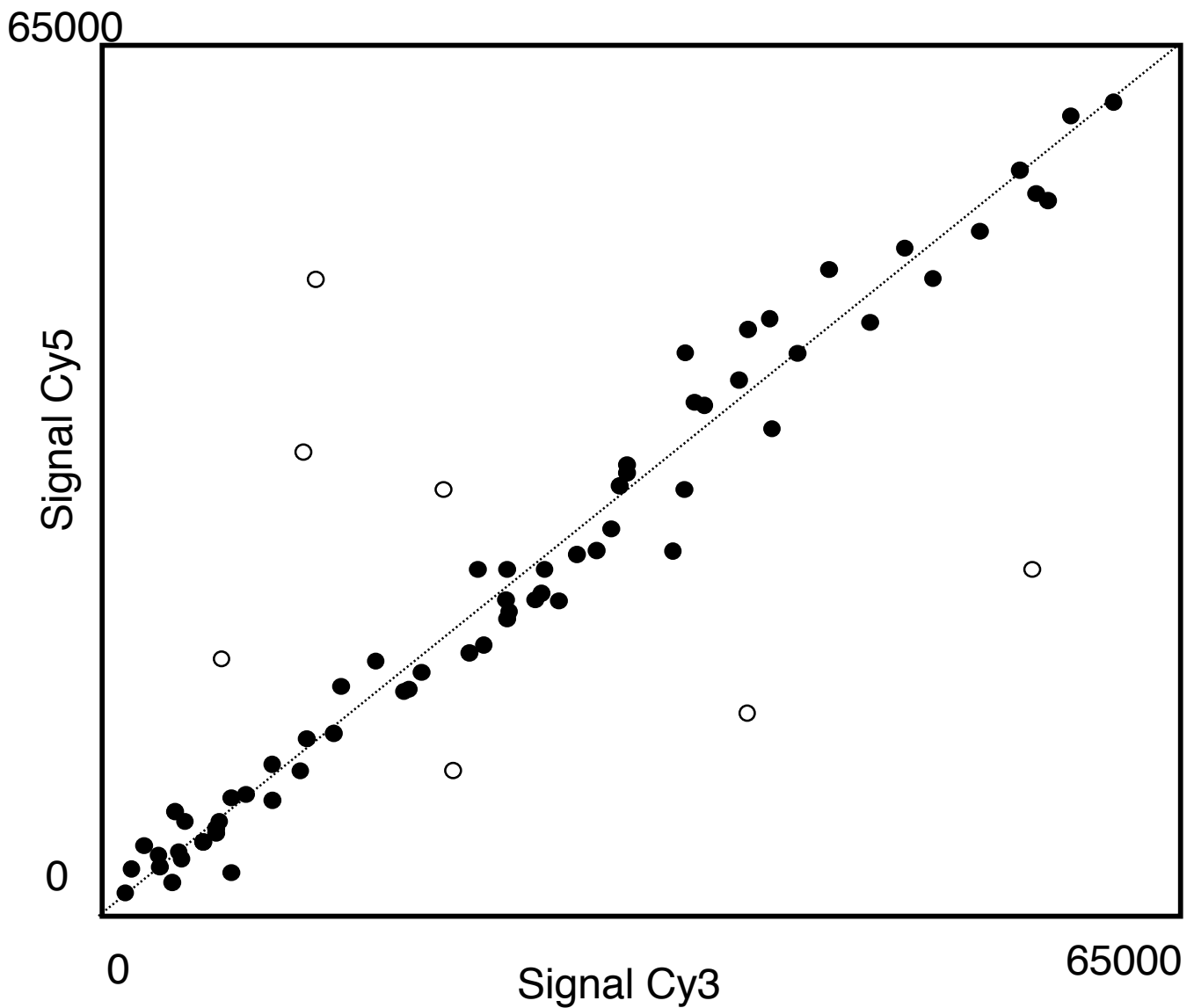


Normalisation des canaux (petit nombre de gènes)



Ajustement du laser et du PMT pour que les gènes constitutifs
aient le même signal dans chaque canal

Normalisation des canaux (grand nombre de gènes)

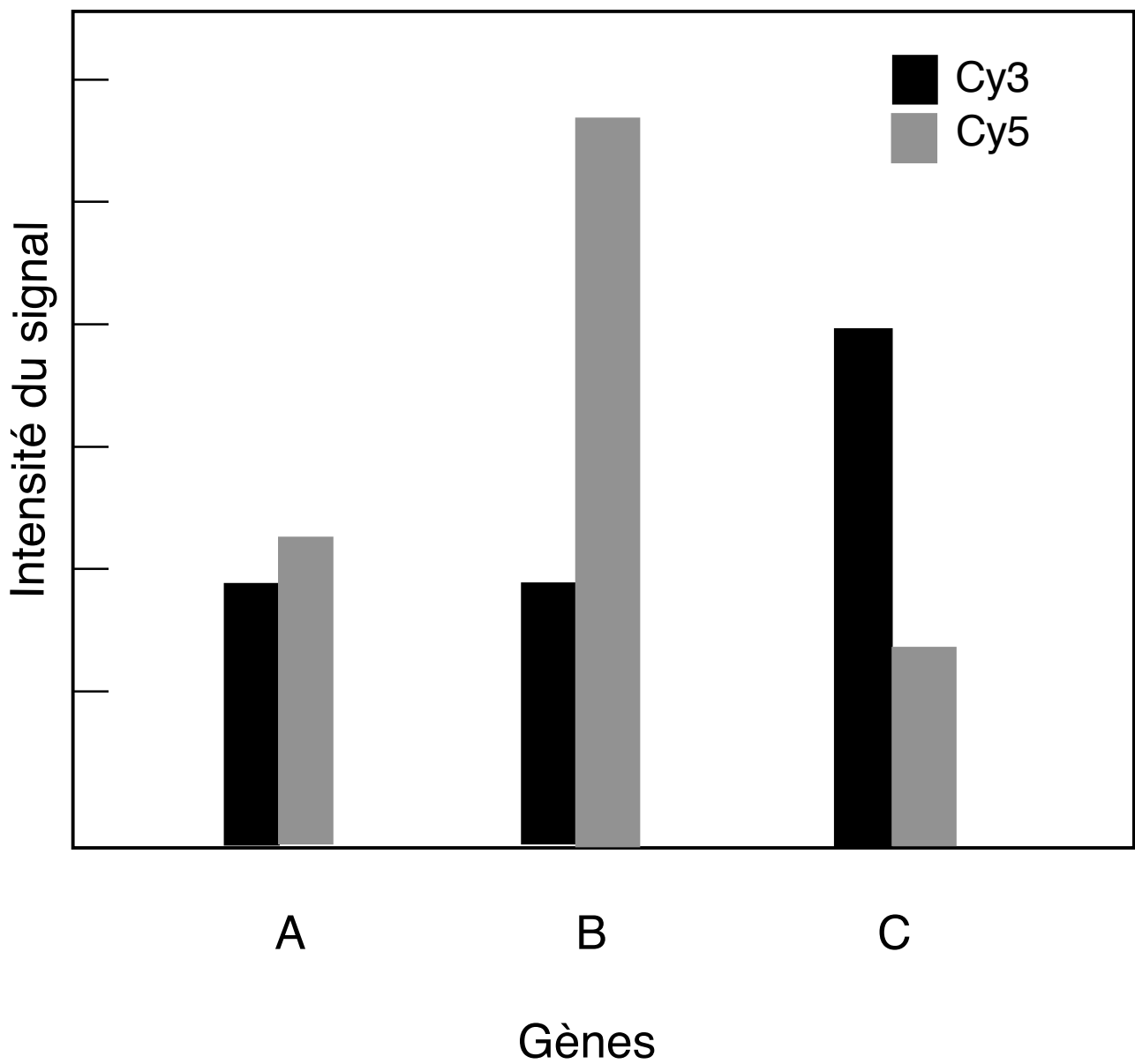


Itérations mathématiques pour que la
médiane des rapports Cy3/Cy5 soient de 1.0

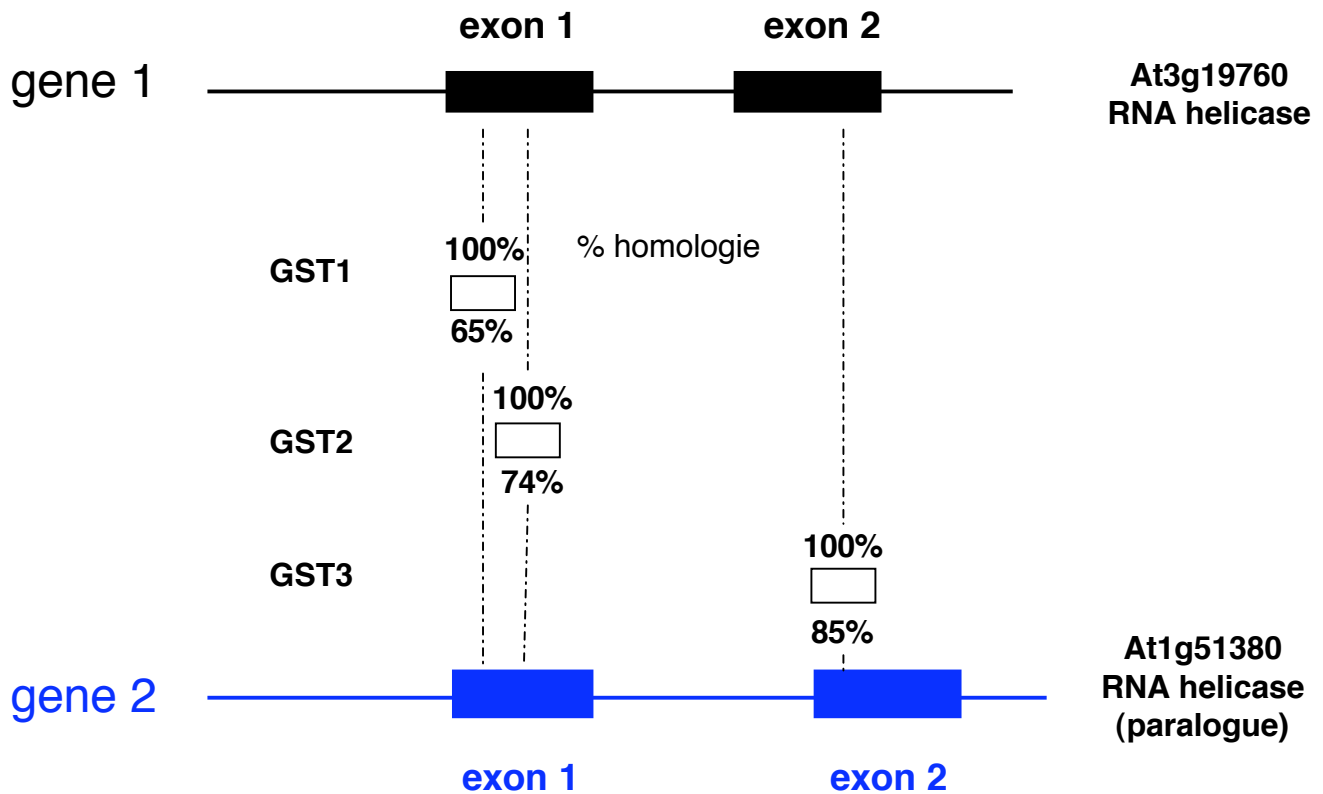
Mesure du rapport d'expression

$$R = (\text{Cy5-Bgd}) * F / (\text{Cy3-Bgd})$$

F=facteur de normalisation

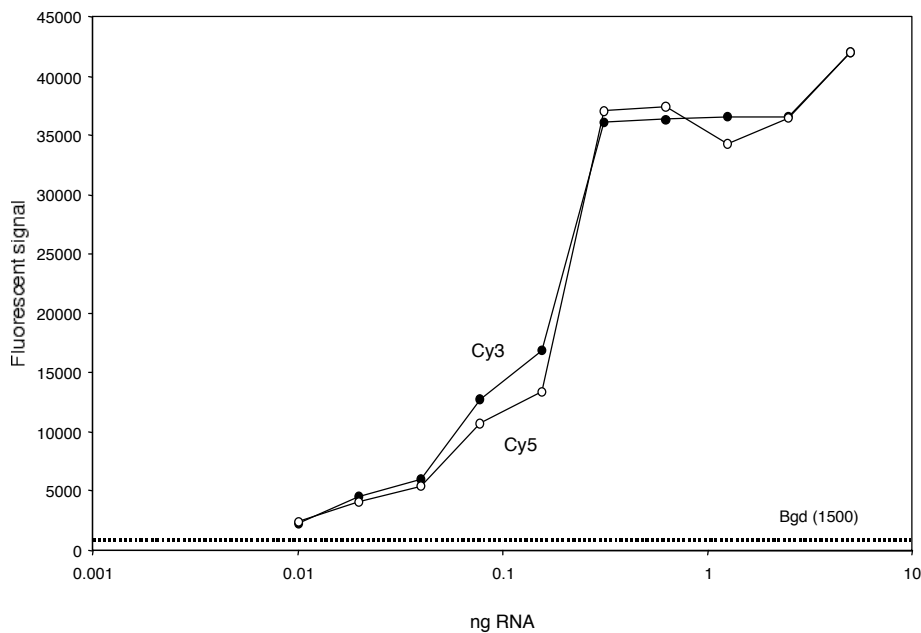


Test de cross-hybridisation



1. Hybridation avec le gène1 marqué au Cy3 -> 100% signal avec les trois GSTs
 2. Hybridation avec le gène2 marqué au Cy3 -> seulement 15% signal avec GST3
- Cross-hybridation si >70-75% homologie**

Test du seuil de sensibilité

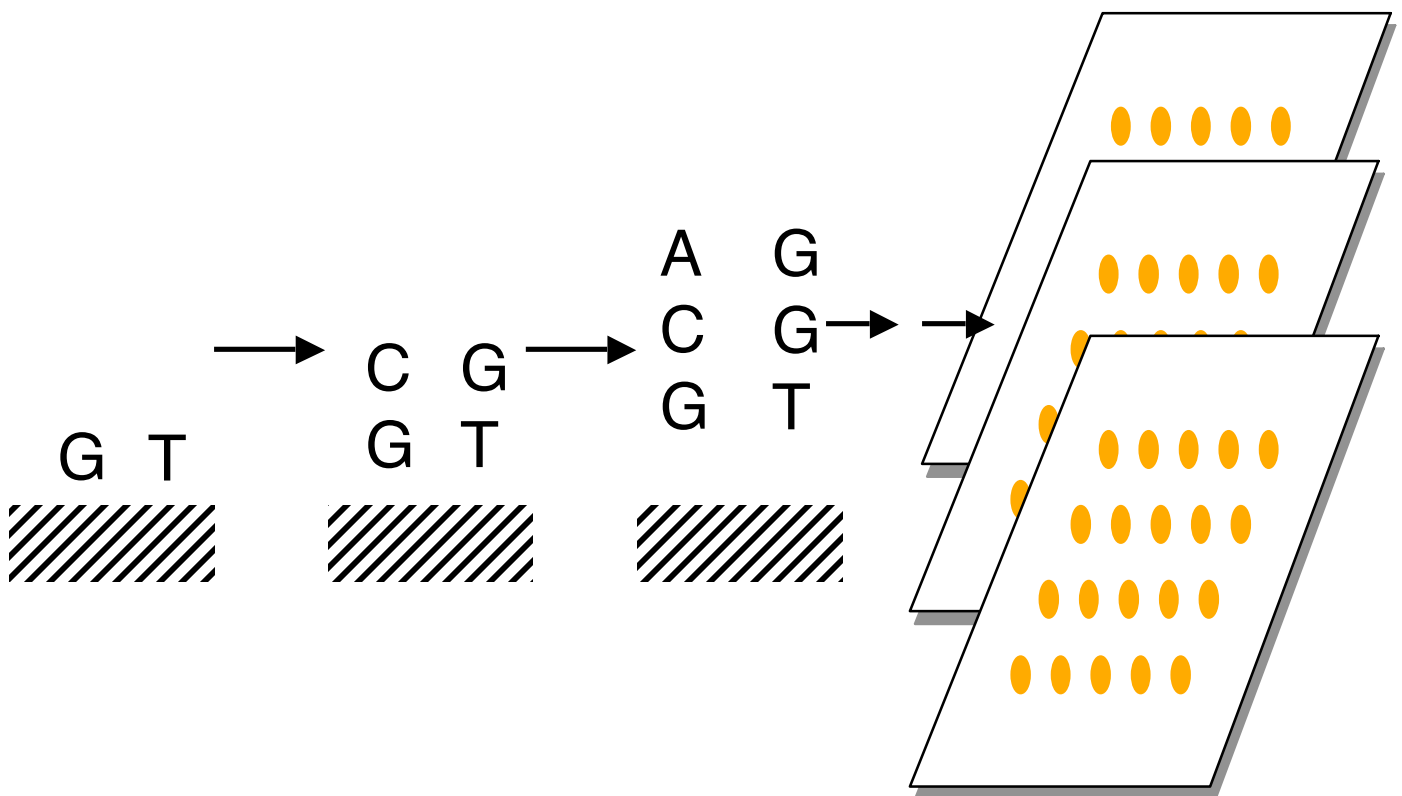


1. Il y a 100'000-500'000 mRNAs par cellule
 2. On détecte 0.01 ng de mRNA sur un total de 2 µg marqués (rapport 1/200'000)
- > on peut mesurer la présence de env. 1 copie par cellule

Fabrication de puces à oligonucléotides (10-70 nt)

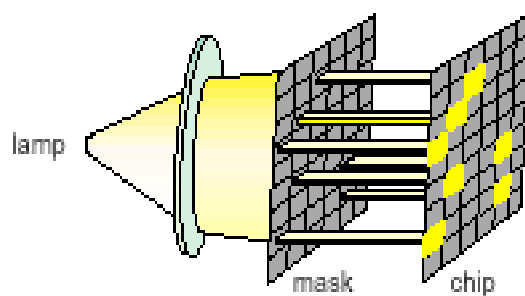
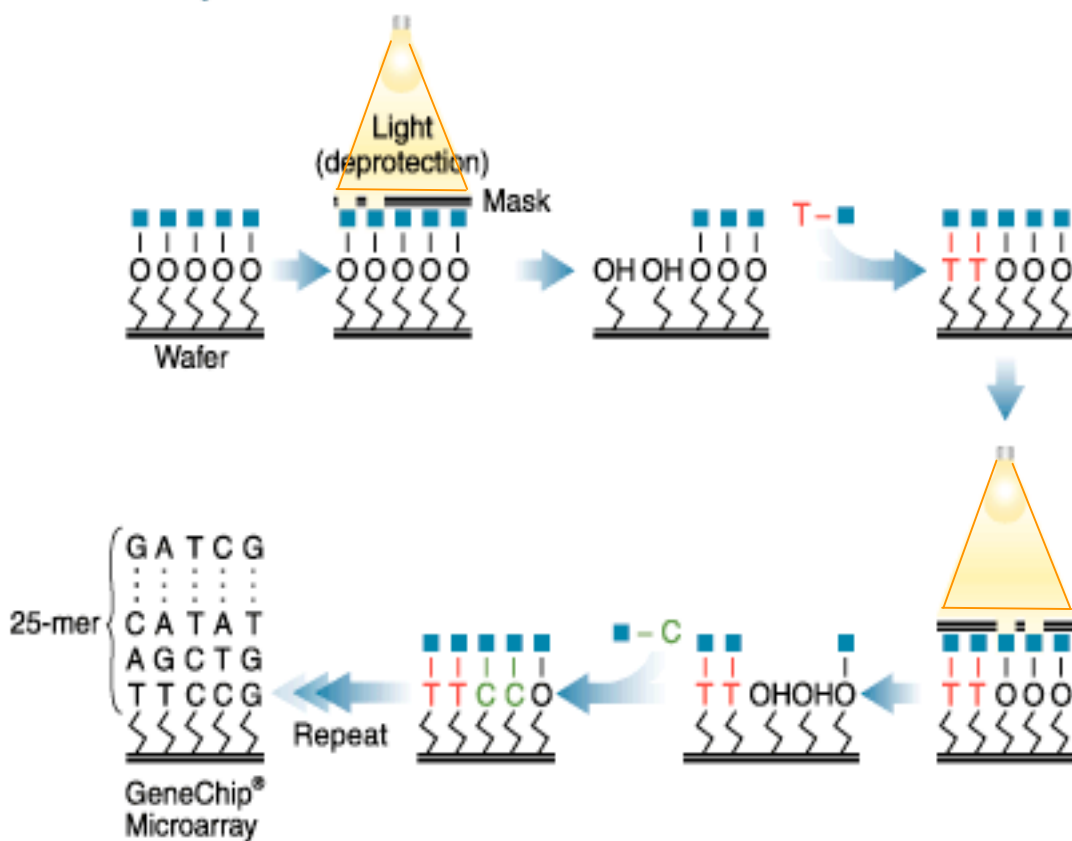
Synthèse chimique sur une surface de verre

- Chips Affymetrix: photolithogravure
- Chips NimbleGen: photolithogravure
- Autres: imprimantes à jet d'encre



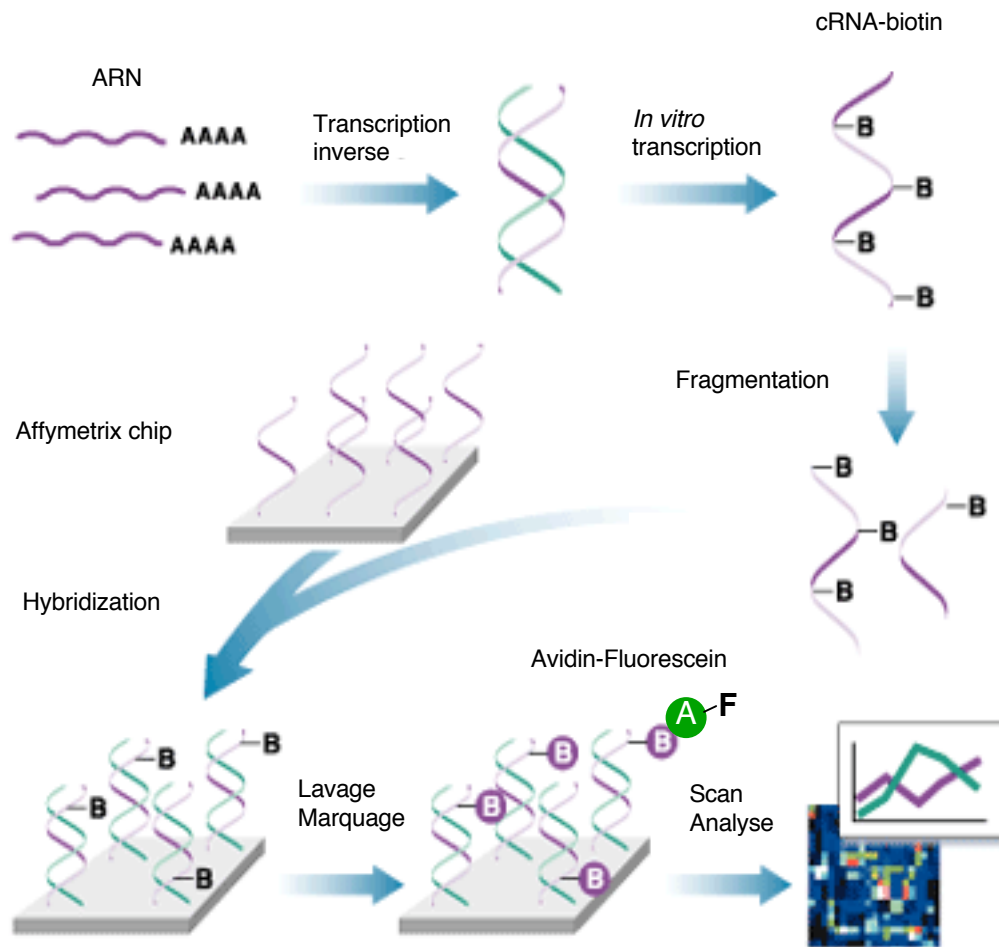
Affymetrix chips

Fabrication par photolithogravure



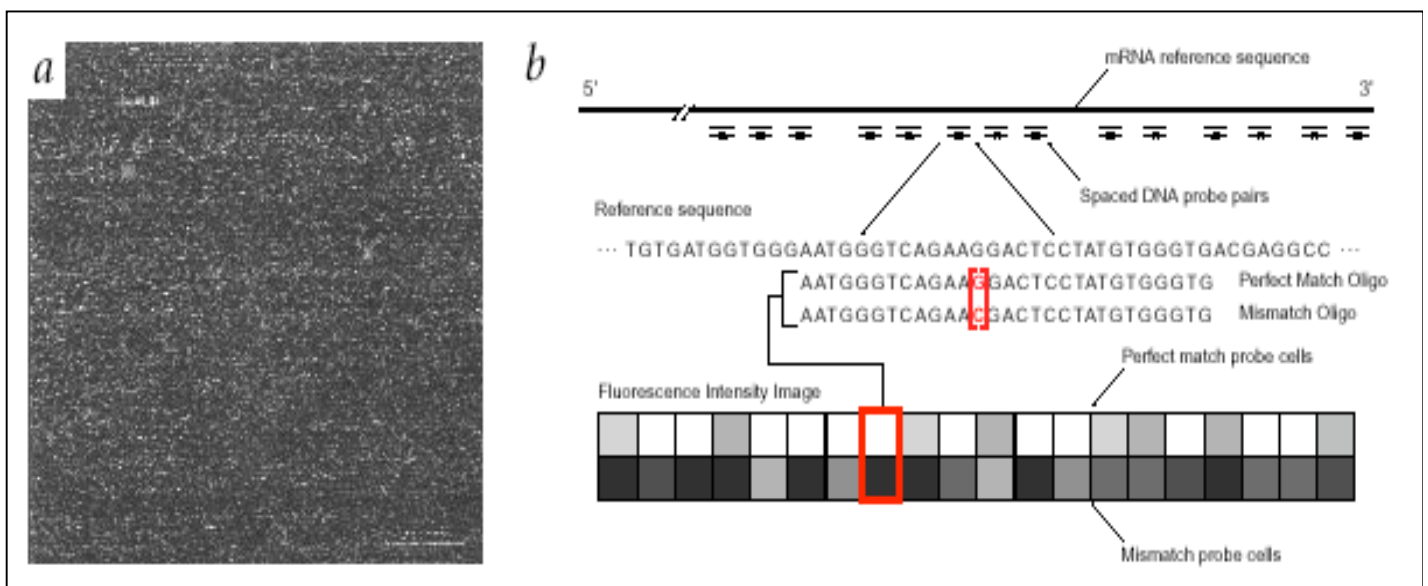
Affymetrix chips

Protocole d'analyse



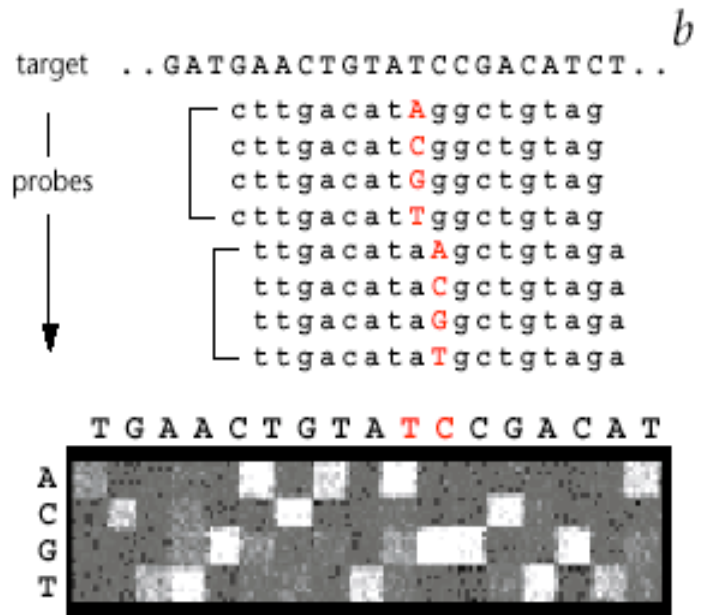
Affymetrix chips

Mesure d'expression

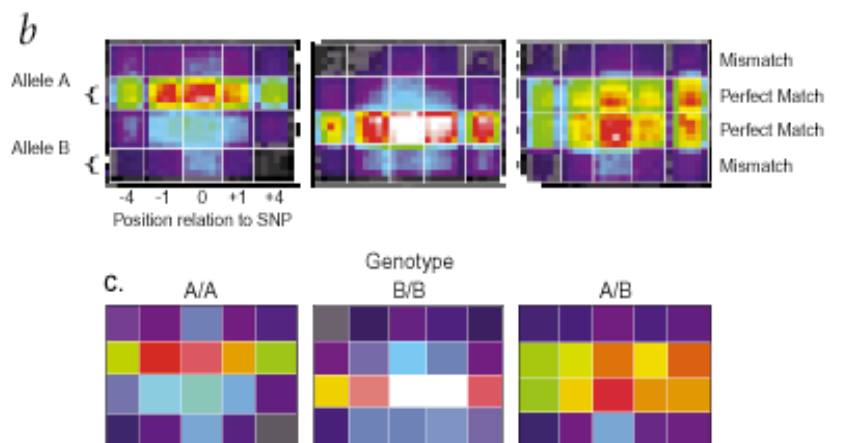
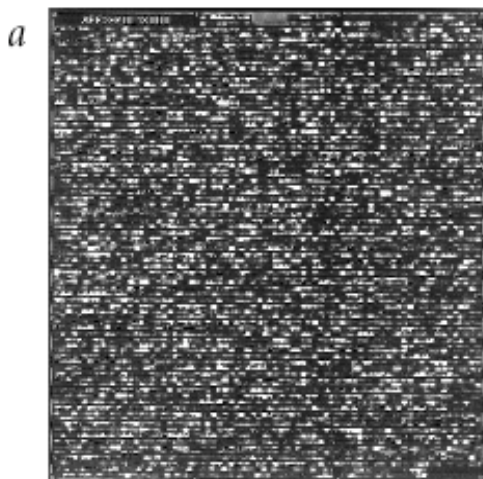


Affymetrix chips

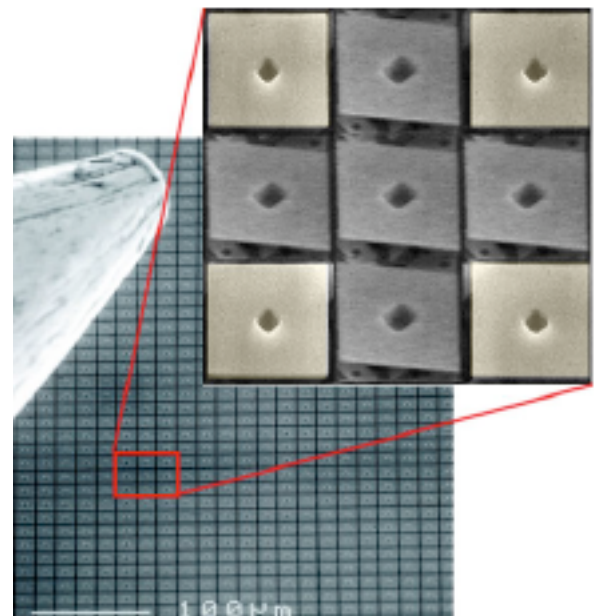
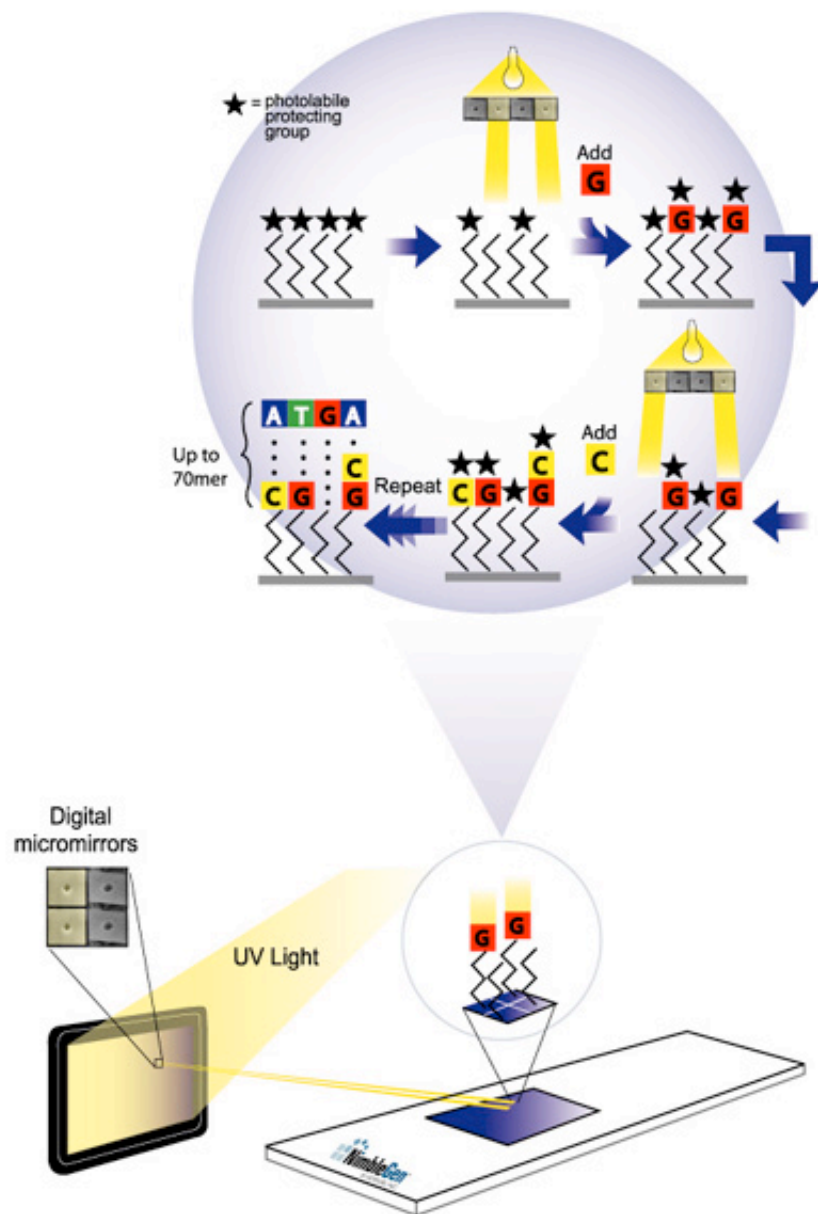
Séquençage



Analyse de SNP



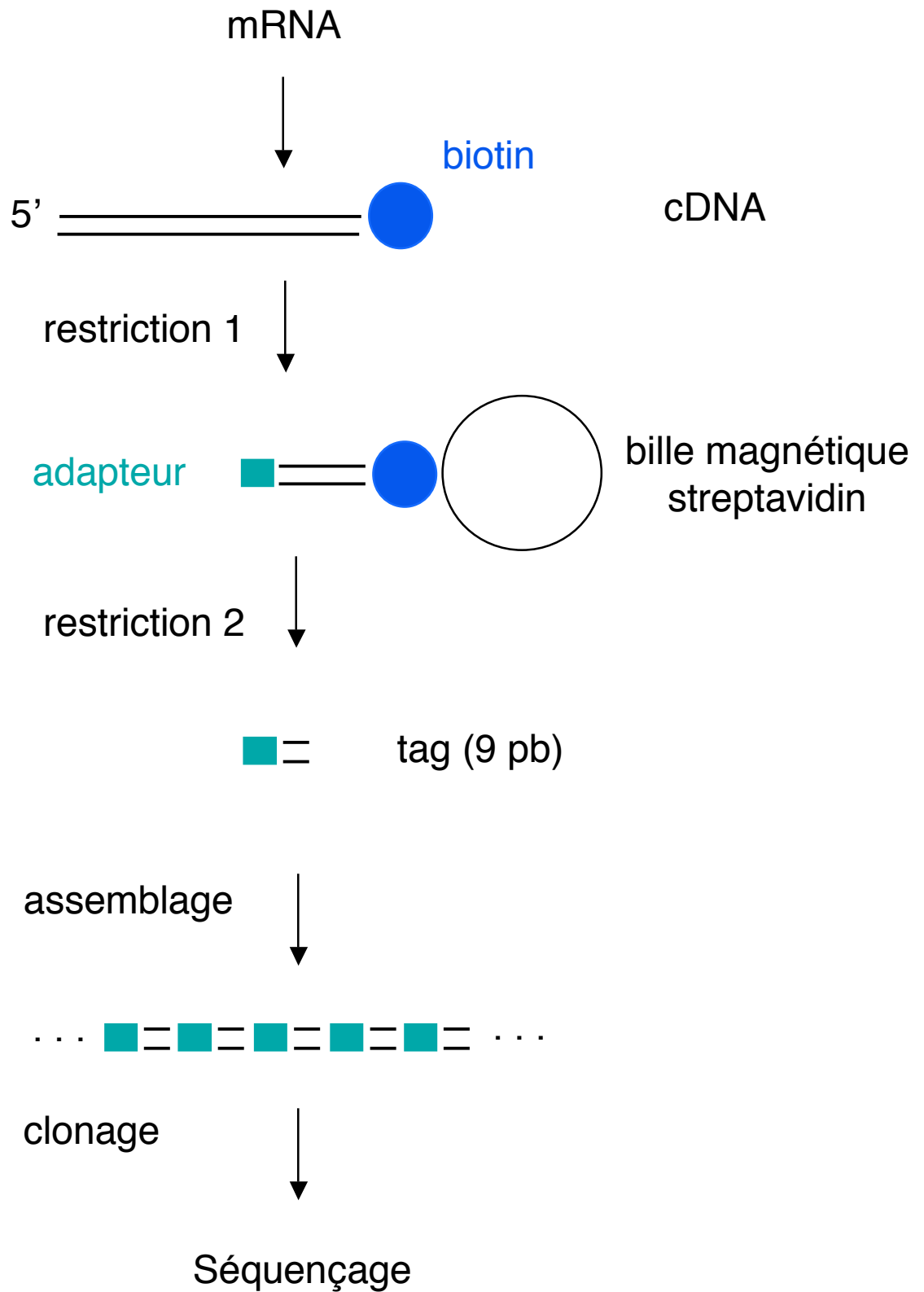
NimbleGen oligonucleotide chip



Impression d'ADN	Synthèse d'ADN <i>in situ</i>
"DNA microarrays"	"DNA chips", "oligonuclotide chips", "Affymetrix chips"
Déposition par contact ou jet d'encre	Synthèse sur la lame (photolithogravure ou jet d'encre).
<ul style="list-style-type: none"> -fragments PCR -clones -longs oligonucléotides (50-70 mer) 	-oligonucléotides (25mer)
Connaissance de la séquence pas nécessaire à l'avance	Connaissance de la séquence nécessaire
Double marquage (Cy3 + Cy5)	Simple marquage (biotin-avidin-fluorescein)
> 60'000 spots par lame	> 500'000 oligos par lame (env. 40'000 gènes)
Robot, scanner	Appareil à hybrider, scanner
Prix abordable (gènes à préparer)	Très cher

SAGE

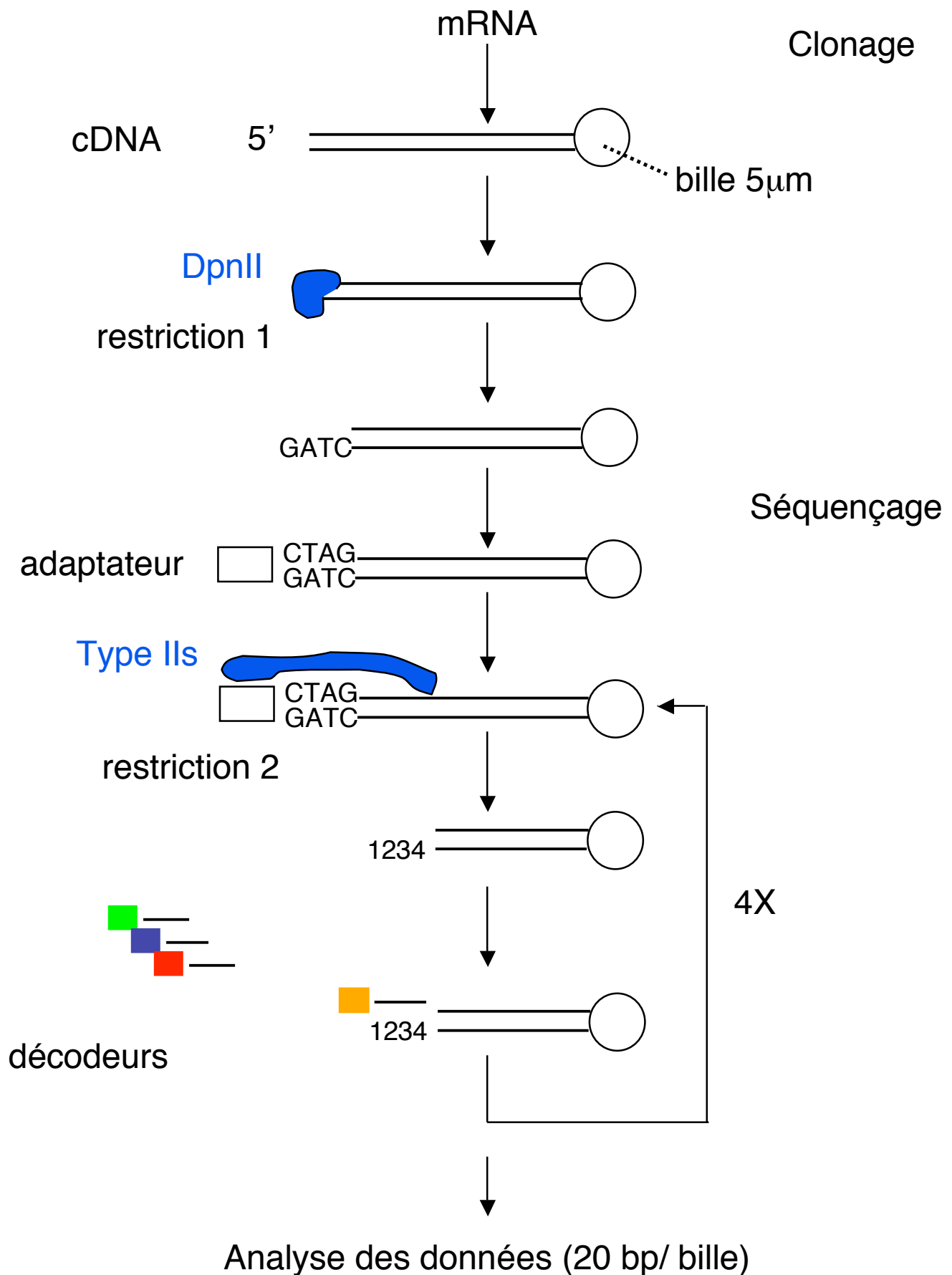
(Serial Analysis of Gene Expression)



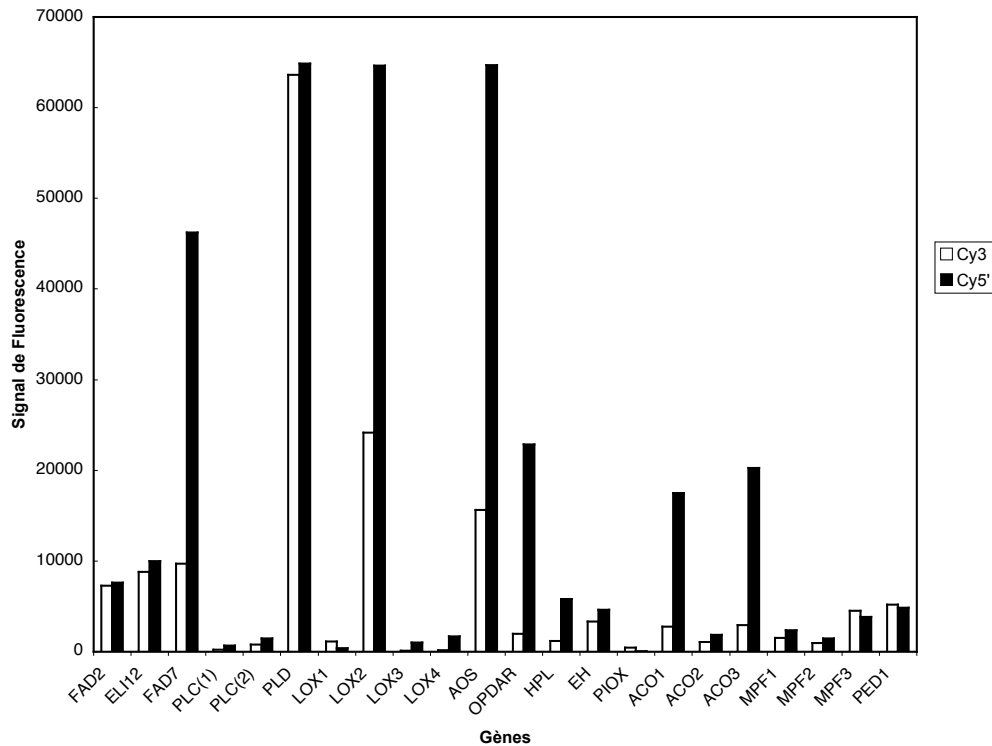
SAGE: exemple de résultats

Tag	Gene	n	Percent
GAGCACACC	Procarboxypeptidase A1 (X67318)	64	7.6
TTCTGTGTG	Pancreatic trypsinogen 2 (M27602)	46	5.5
GAACACAAA	Chymotrypsinogen (M24400)	37	4.4
TCAGGGTGA	Pancreatic trypsin 1 (M22612)	31	3.7
GCGTGACCA	Elastase III B (M18692)	20	2.4
GTGTGTGCT	Protease E (D00306)	16	1.9
TCATTGGCC	Pancreatic lipase (M93285)	16	1.9
CCAGAGAGT	Procarboxypeptidase B (M81057)	14	1.7
TCCTCAAAA	No match (see Table, 2, P1)	14	1.7
AGCCTTGGT	Bile salt stimulated lipase (X54457)	12	1.4
GTGTGCGCT	No match	11	1.3
TGCGAGACC	No match (see Table 2 P2)	9	1.1
GTGAAACCC	21 Alu entries	8	1.0
GGTGACTCT	No match	8	1.0
AAGGTAACA	Secretory trypsin inhibitor (M11949)	6	0.7
TCCCTGTGTG	No match	5	0.6
GTGACCACG	No match	5	0.6
CCTGTAATC	M91159, M29366, 11 Alu entries	5	0.6
CACGTTGGA	No match	5	0.6
AGCCCTACA	No match	5	0.6
AGCACCTCC	Elongation factor 2 (Z11692)	5	0.6
ACGCAGGGA	No match (see Table 2, P3)	5	0.6
AATTGAAGA	No match (see Table 2, P4)	5	0.6
TTCTGTGGG	No match	4	0.5
TTCATACAC	No match	4	0.5
GTGGCAGGC	NF- κ B (X61499), Alu entry (S94541)	4	0.5
GTAAAACCC	TNF receptor II (M55994), Alu entry (X01448)	4	0.5
GAACACACA	No match	4	0.5
CCTGGGAAG	Pancreatic mucin (J05582)	4	0.5
CCCATCGTC	Mitochondrial CytC oxidase (X15759)	4	0.5
SAGE tags occurring:	Greater than three times	380	45.2
	Three times (15 \times 3 =)	45	5.4
	Two times (32 \times 2 =)	64	7.6
	One time	351	41.8
	Total SAGE tags	840	100.0

MPSS (Massively Parallel Signature Sequencing)



Analyse des données

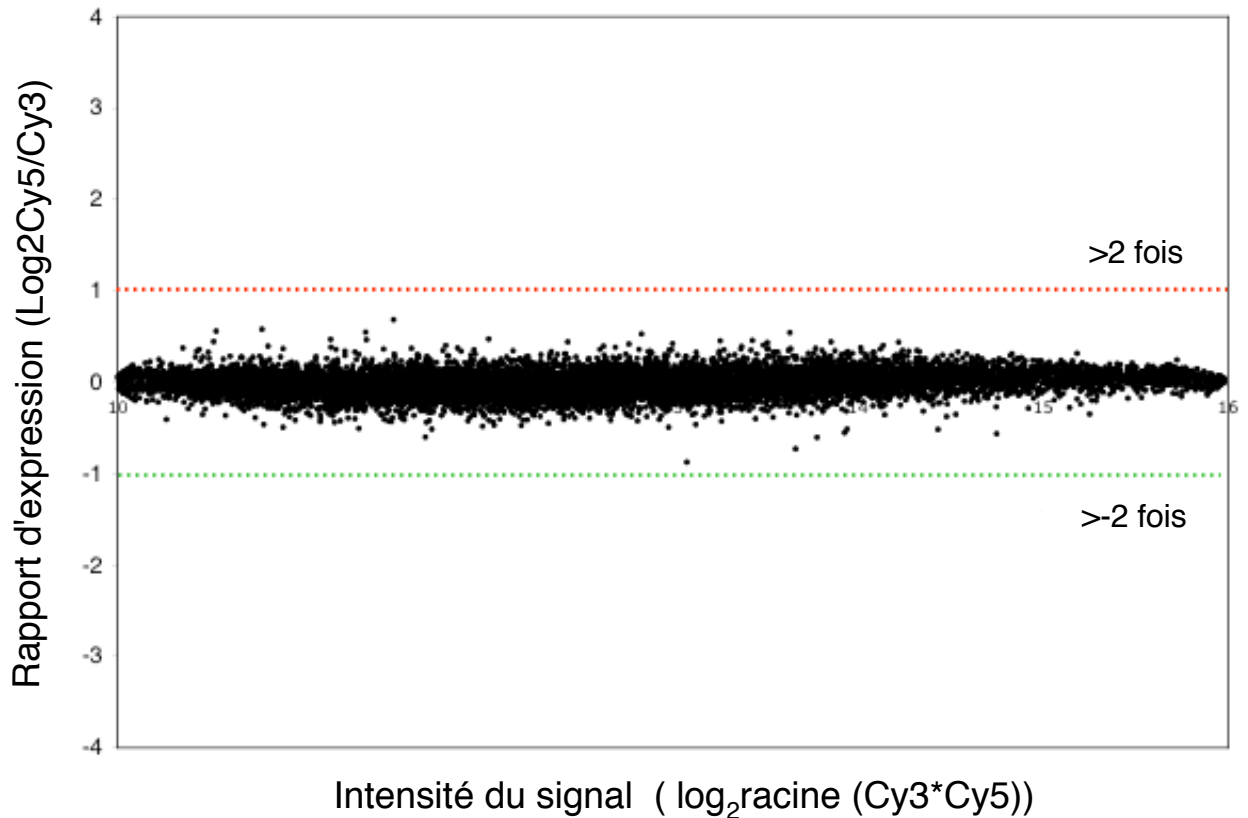


Comment identifier un gène induit ou réprimé?

1. Seuil arbitraire (signal 2 fois plus grand dans la condition 1 que dans la condition 2)
2. Répliquer l'expérience, test student t $H_0: \mu_i = 1$
Si μ_i différent de 1, le gène est induit avec $P < 0.05$

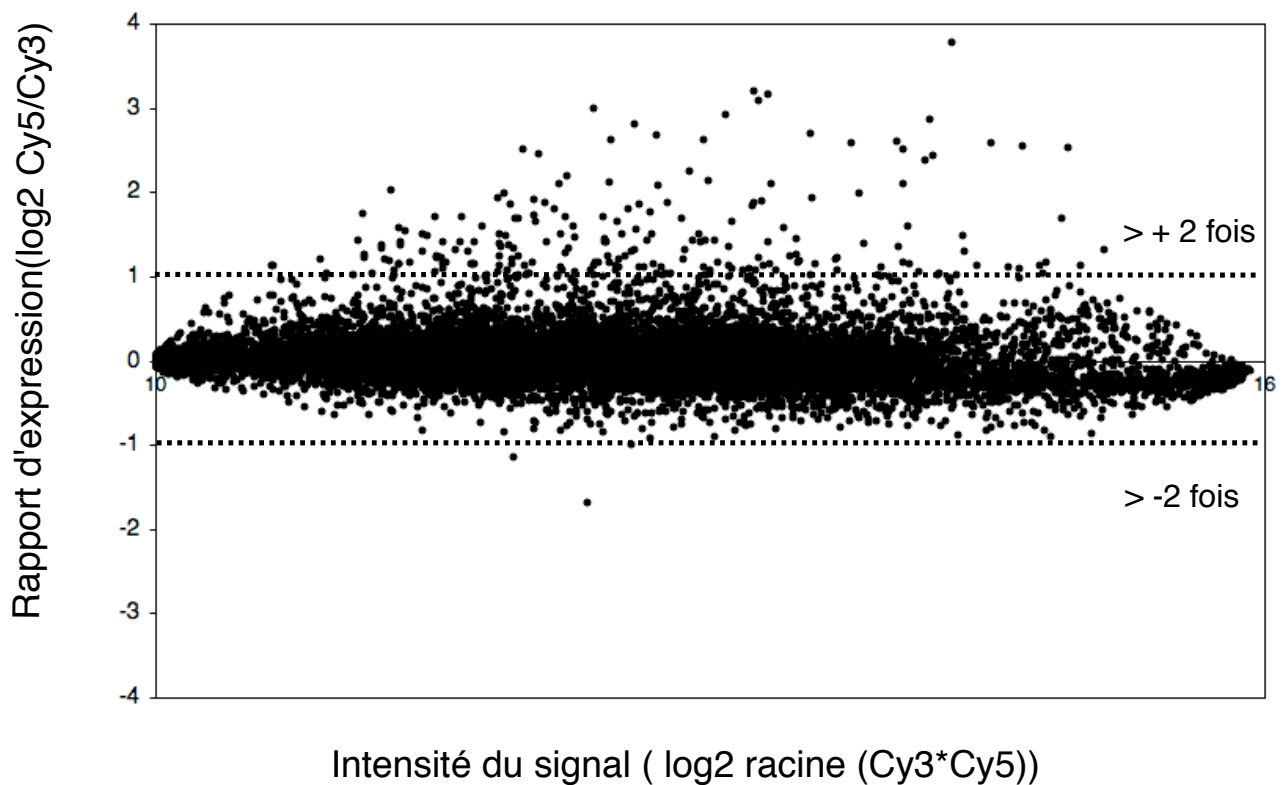
CTL1 vs CTL2

(moyenne de 6 réplicats indépendants)



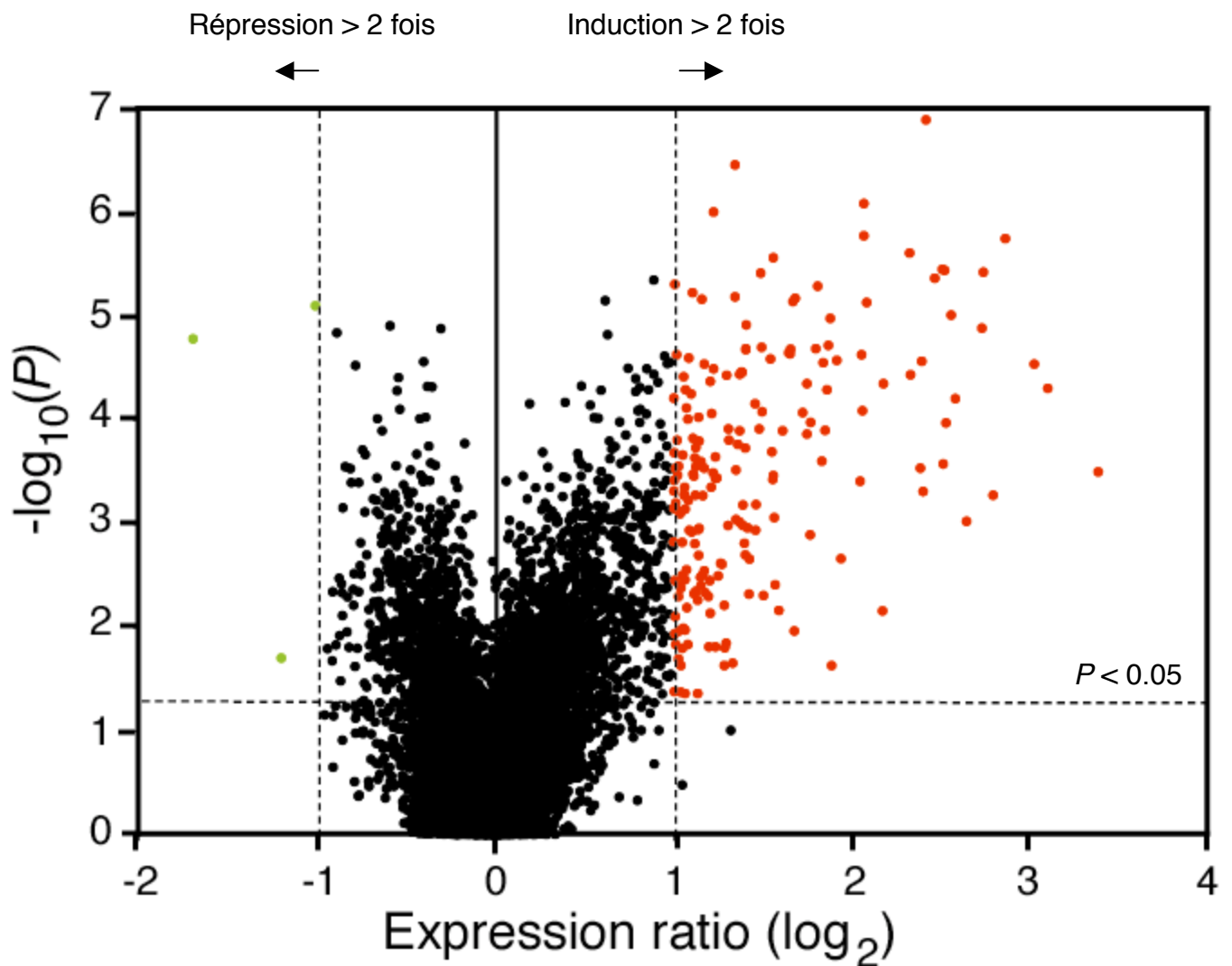
CTL vs Pieris

Effet de *Pieris rapae* sur l'expression génétique chez *Arabidopsis thaliana*
(moyenne de 5 réplicats indépendants)

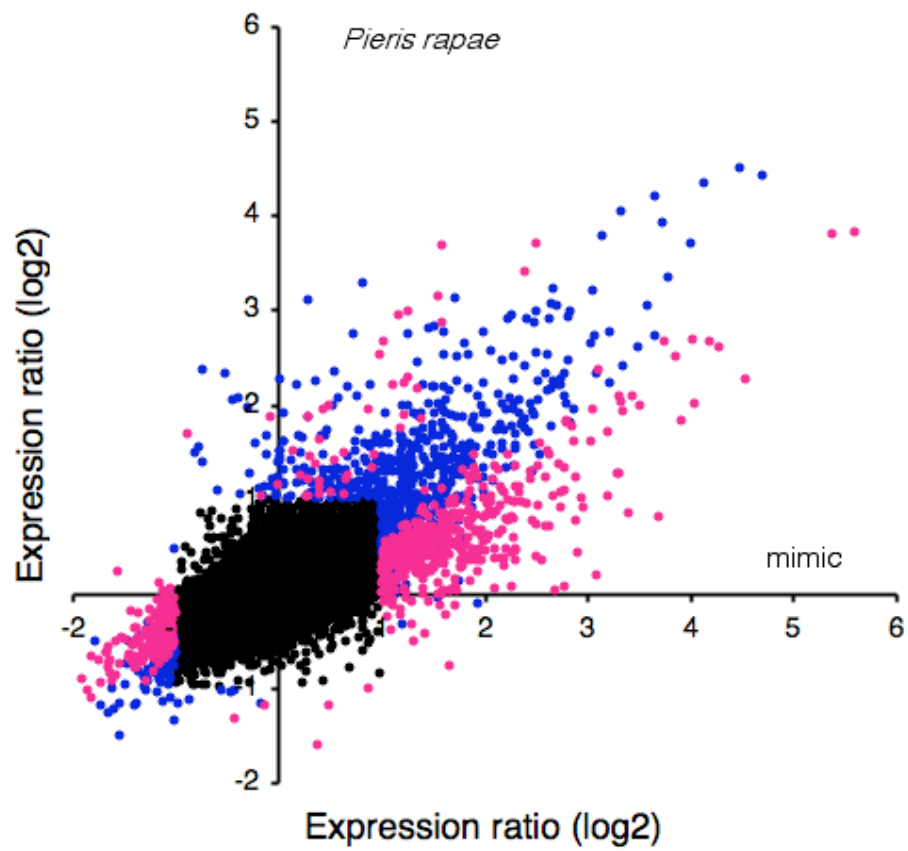


Analyse statistique du changement d'expression

Effet de *Pieris rapae* sur l'expression génétique chez *Arabidopsis thaliana*

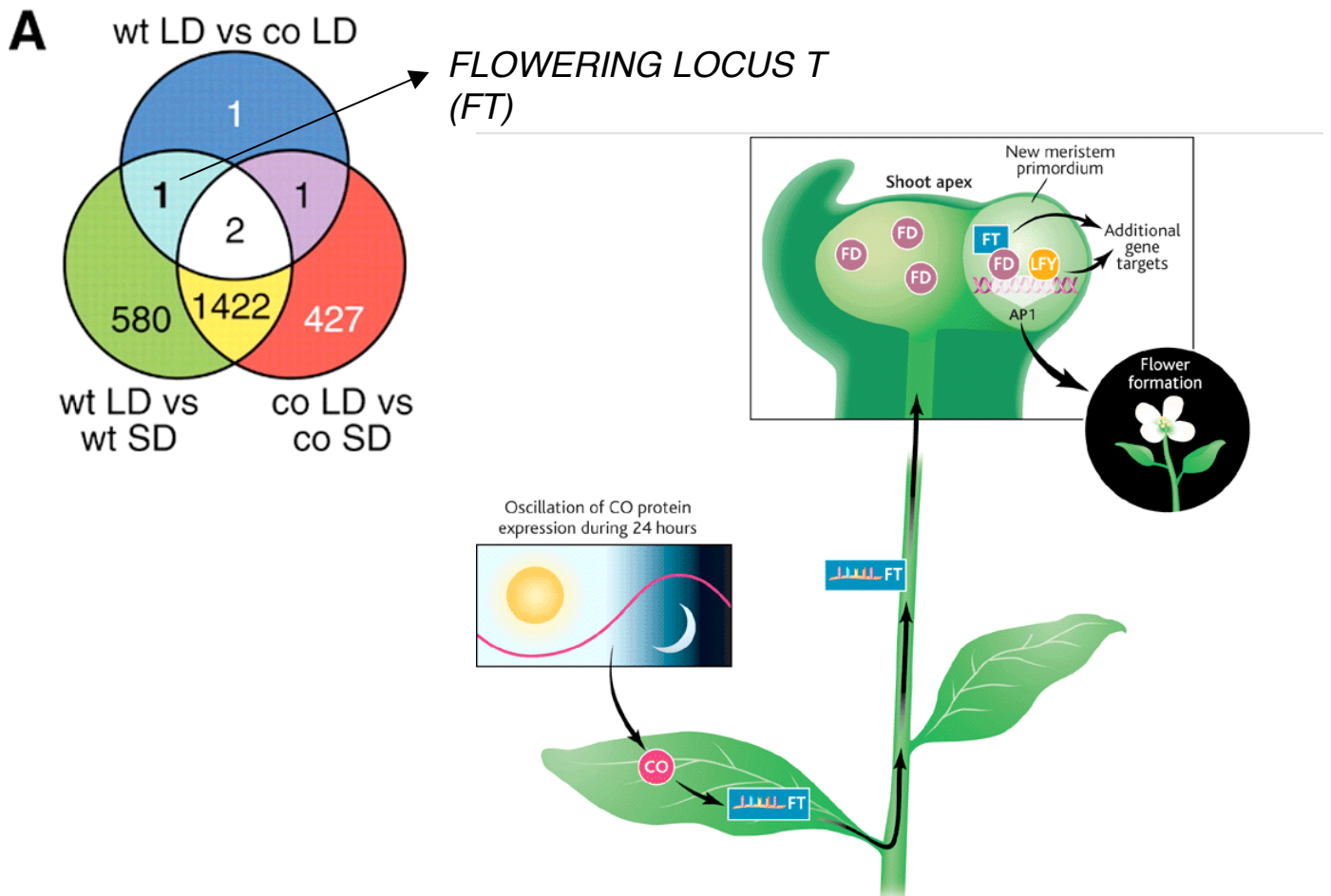
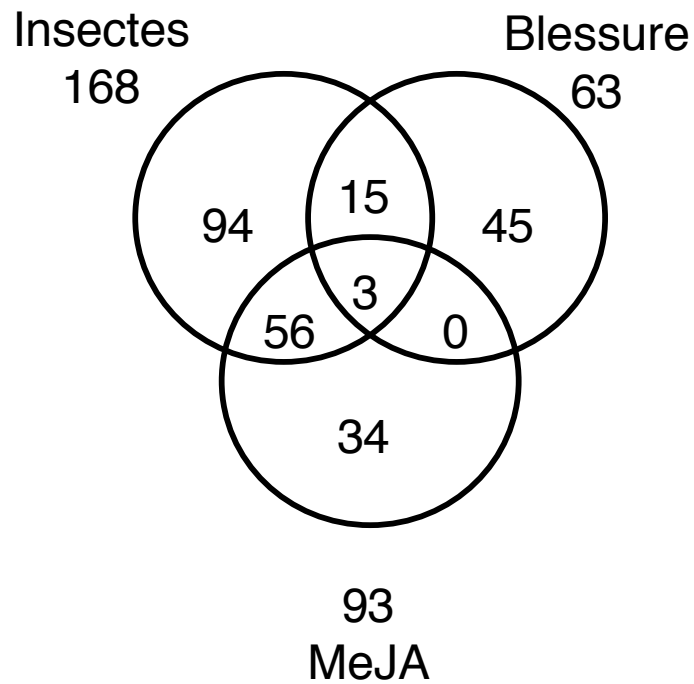


12K *Arabidopsis* microarray, t-test sur 5 réplicats indépendants



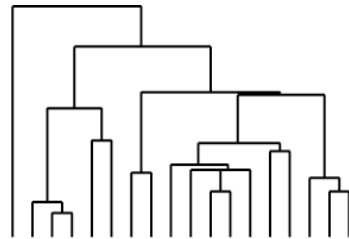
Test $t_{\mu_1 = \mu_2}$ et rapport $>$ ou $<$ 2

Diagramme de Venn

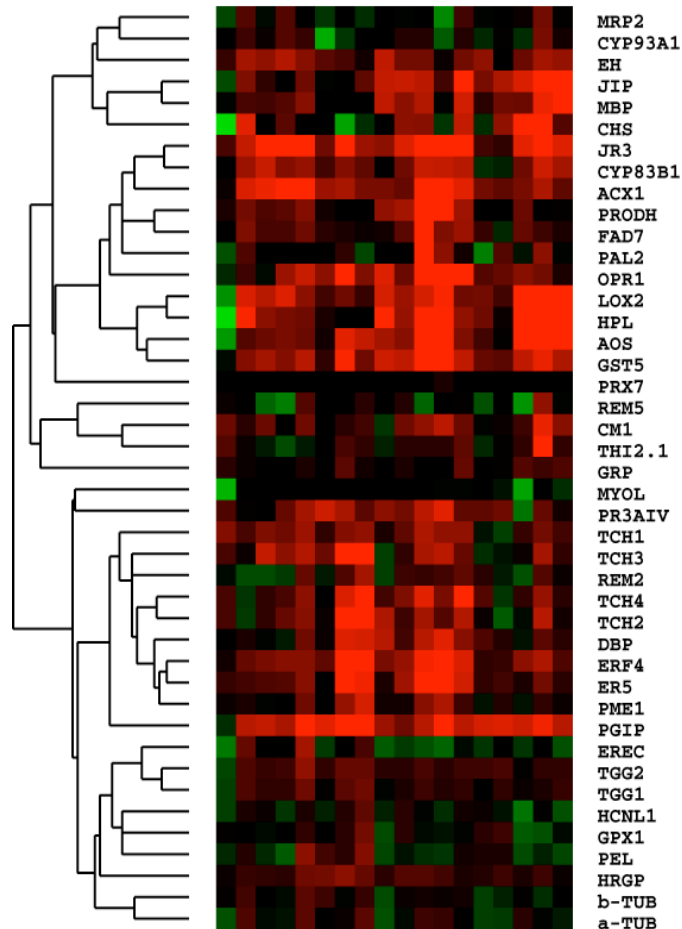


Hierarchical clustering

Expériences

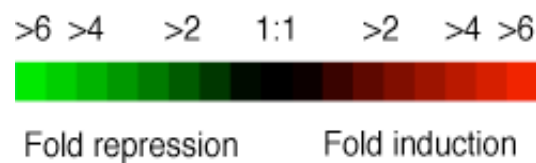


SA24hr
Ecc48hr
Ecc30hr
Ecc16hr
Ecc8hr
Ecc4hr
W30 min
W15 min
W6 hr
P. rapae
W90 min
W60 min
W3 hr
W24 hr
W9 hr
MeJA3hr
MeJA16hr
MeJA6hr



MRP2
CYP93A1
EH
JIP
MBP
CHS
JR3
CYP83B1
ACX1
PRODH
FAD7
PAL2
OPR1
LOX2
HPL
AOS
GST5
PRX7
REM5
CM1
THI2.1
GRP
MYOL
PR3AIV
TCH1
TCH3
REM2
TCH4
TCH2
DBP
ERF4
ER5
PME1
PGIP
EREC
TGG2
TGG1
HCNL1
GPX1
PEL
HRGP
b-TUB
a-TUB

Gènes

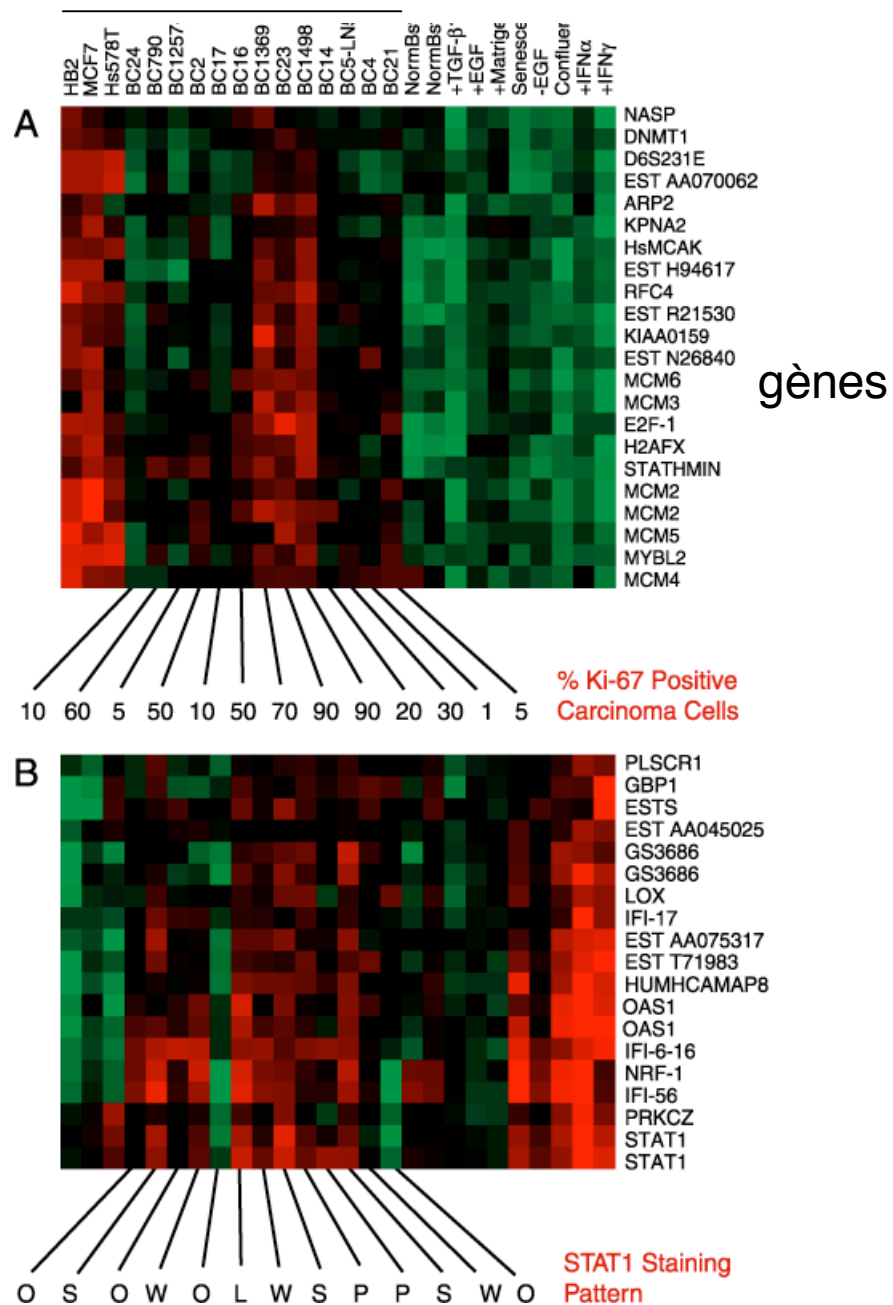


<http://rana.lbl.gov/EisenSoftware.htm>

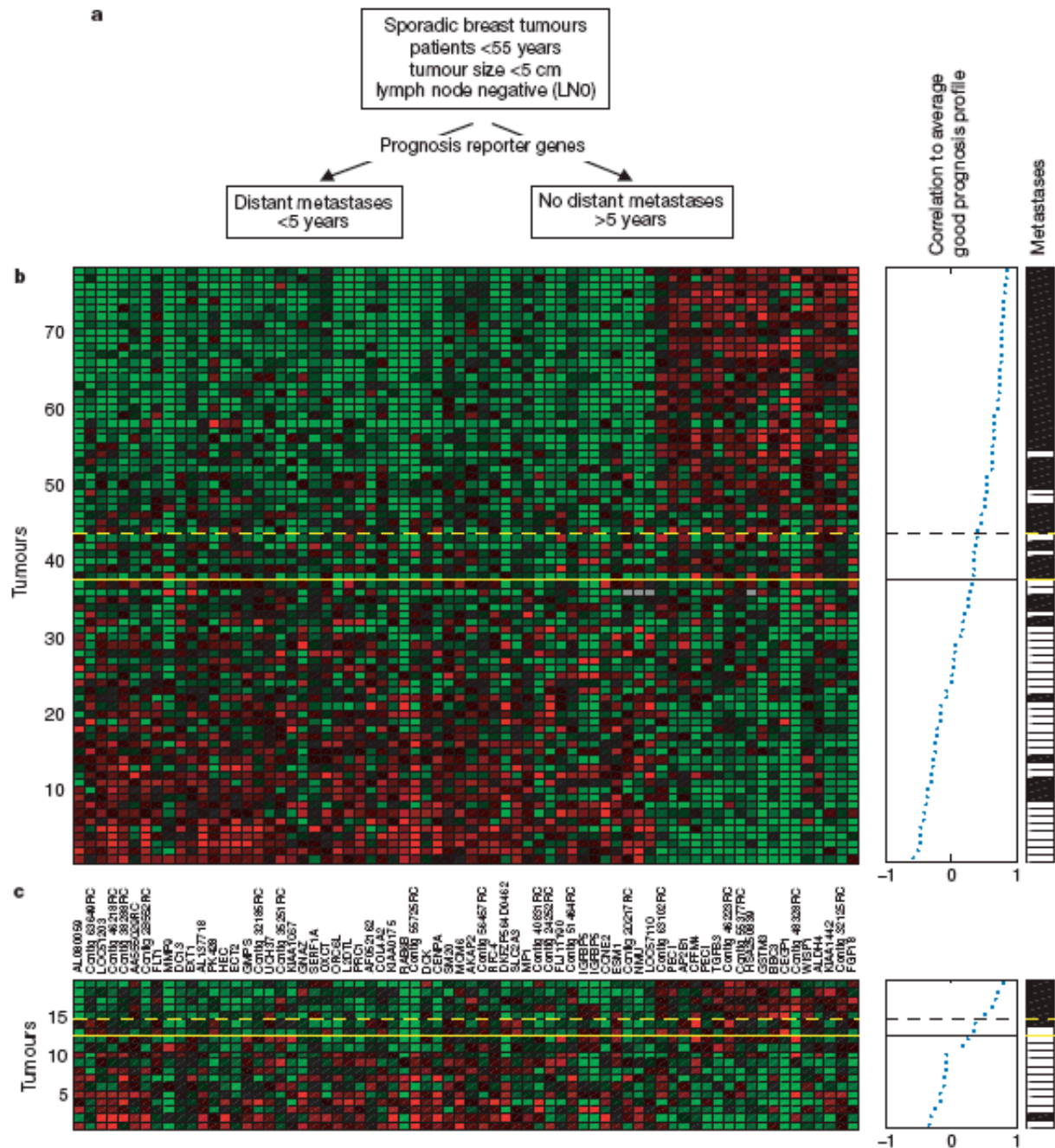
Cancer du sein

Tumeurs impossibles à différencier
par analyses classiques

patients



Le microarray comme outil diagnostique



Analyse de correspondance

(Phylip, <http://evolution.genetics.washington.edu/phylip.html>)

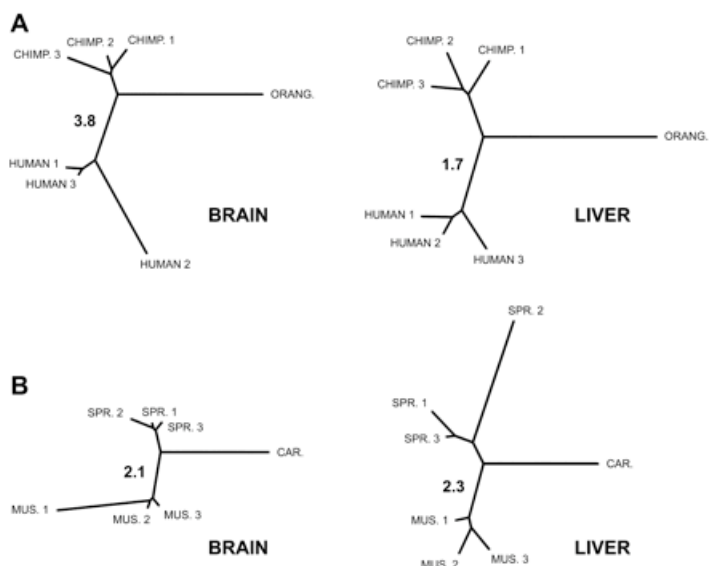
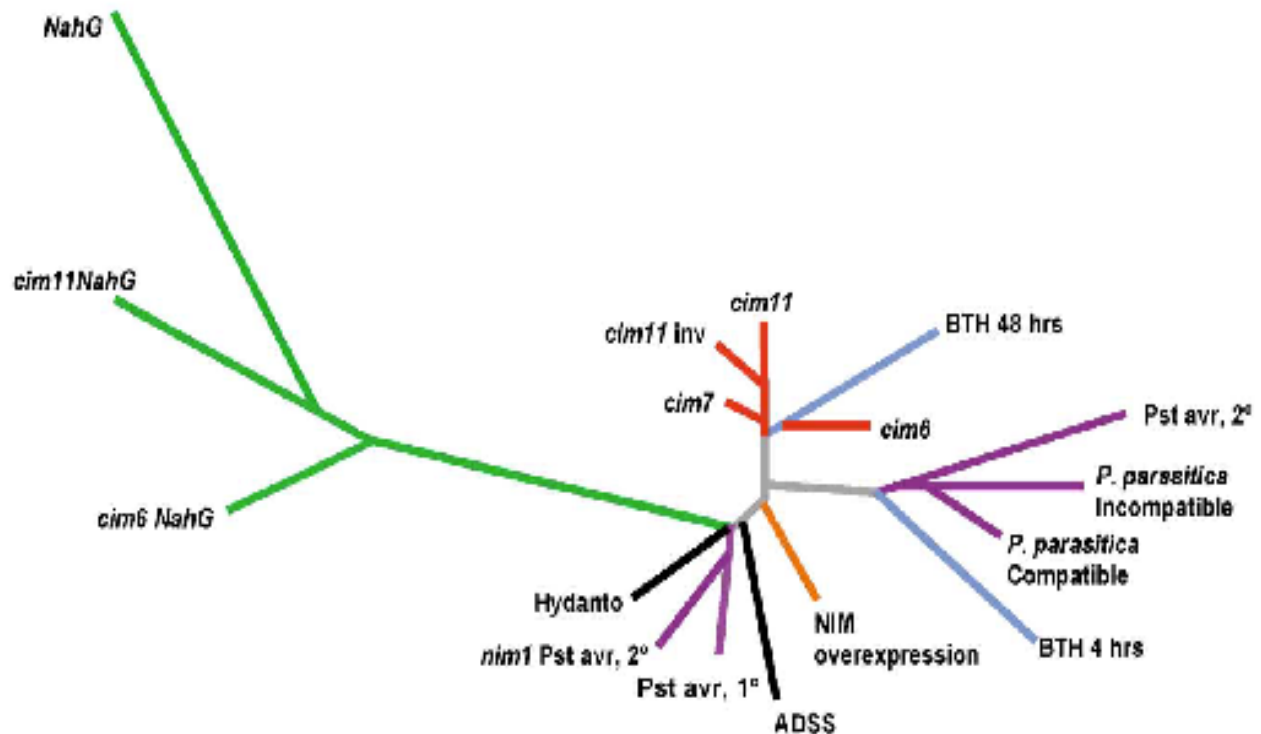


Figure 1. Distance trees representing the relative extent of expression changes in brain and liver among (A) three primate and (B) three mouse species: MUS., *M. musculus*; SPR., *M. spretus*; and CAR, *M. caroli* (6). Numbers refer to the ratio between the changes common to humans and chimpanzees, and *M. musculus* and *M. spretus*, respectively.

Self organizing maps (SOM)

(www-genome.wi.mit.edu/GeneCluster/GenCluster.html)

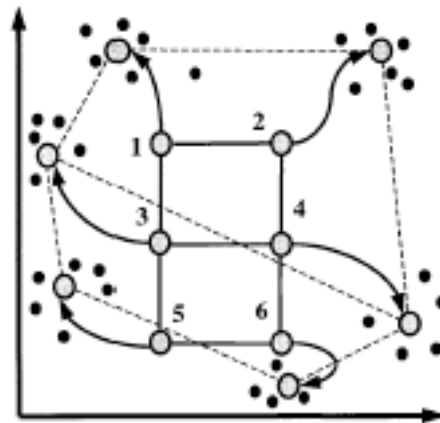
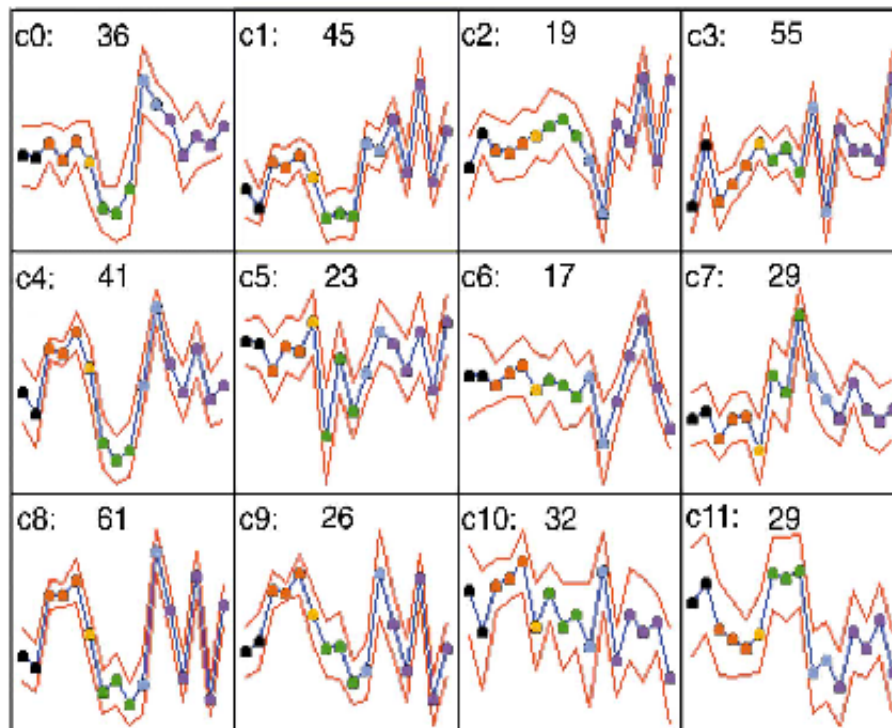
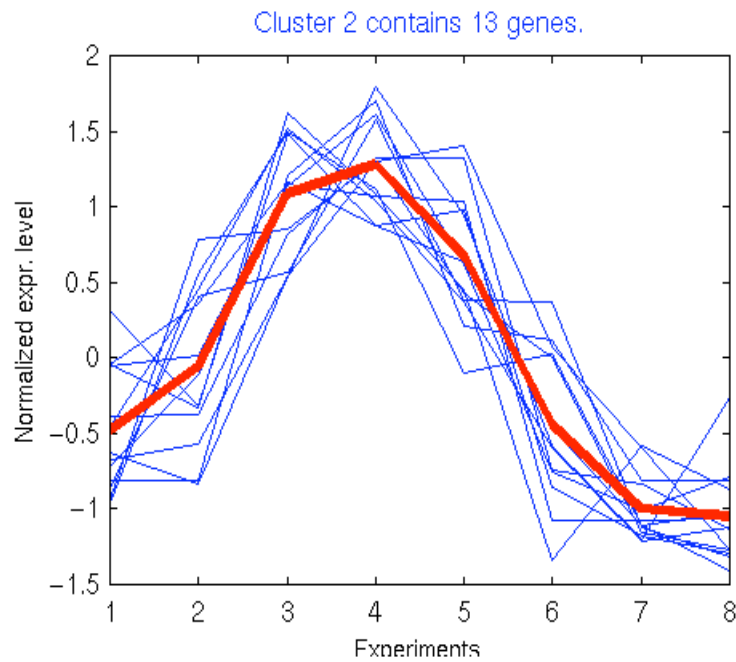


FIG. 1. Principle of SOMs. Initial geometry of nodes in 3×2 rectangular grid is indicated by solid lines connecting the nodes. Hypothetical trajectories of nodes as they migrate to fit data during successive iterations of SOM algorithm are shown. Data points are represented by black dots, six nodes of SOM by large circles, and trajectories by arrows.



Analyse de promoteurs

Cluster 2



Accession Number Gene Name

AF057044	ACX1
M92353	ASA1
L22585	ASB
Z26519	CM1
D78598	CYP83B1
D14007	FAD7
Z26426	GST1
D44465	GST5
Y13577	JR3
Y10617	OPR1
U59508	PRODH
U18993	TSA
M23872	TSB

[Fetch Sequences](#)

Motif 3: CTACCKAC

	1	2	3	4	5	6	7	8
A	0.0156	0.0156	0.9724	0.2890	0.0156	0.0156	0.6990	0.0156
C	0.9664	0.0096	0.0096	0.6930	0.9664	0.0096	0.0096	0.9664
G	0.0053	0.0053	0.0053	0.0053	0.0053	0.5520	0.0053	0.0053
T	0.0127	0.9695	0.0127	0.0127	0.0127	0.4227	0.2860	0.0127

7 Sites



Analyse de promoteurs: un exemple

Harmer et al. (2000) Science 290:2110-2113

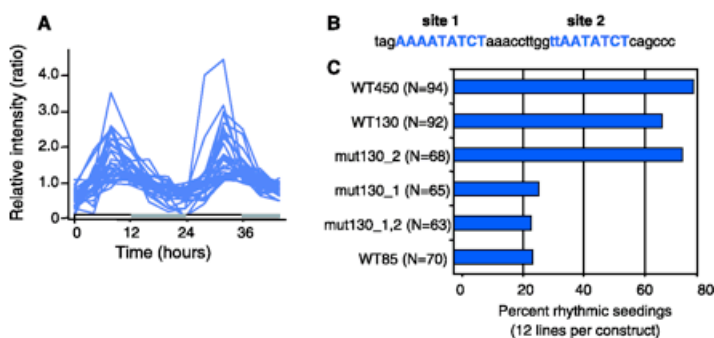
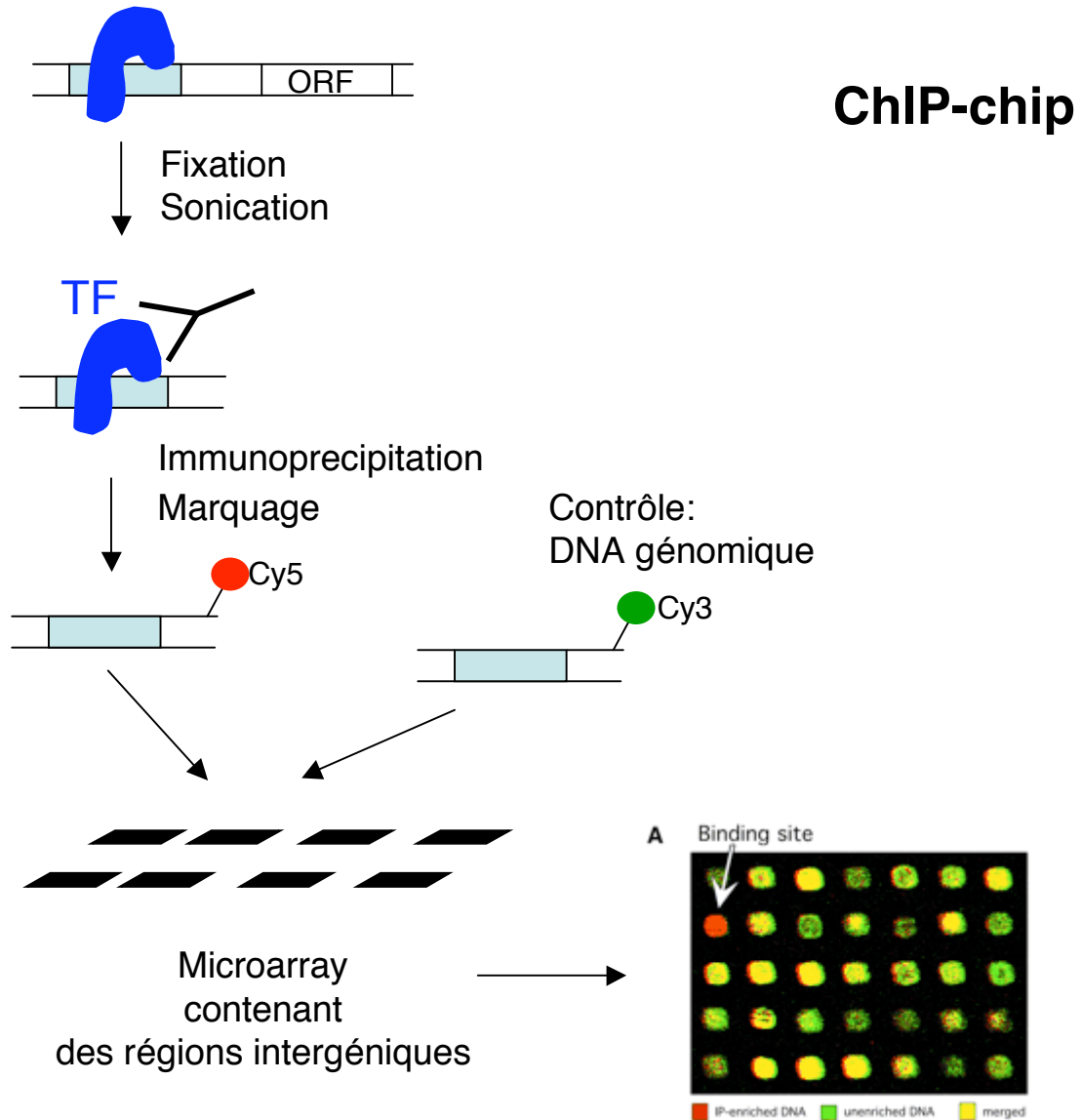


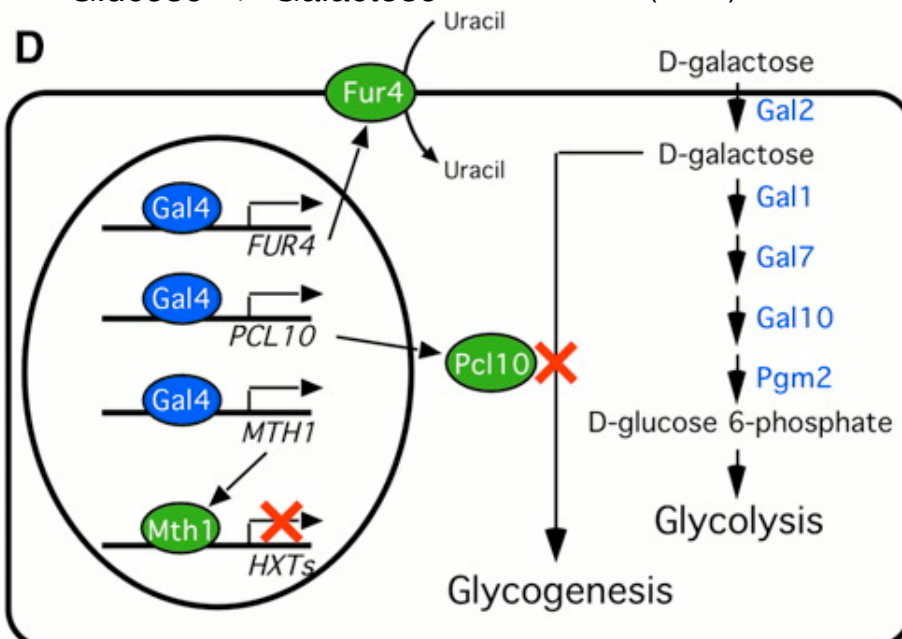
Figure 5. A conserved "evening element" essential for circadian rhythmicity. **(A)** A cluster of 31 clock-controlled genes containing an AAAATATCT promoter evening element. Promoters of clock-controlled genes were scanned for overrepresented elements using AlignACE and ScanACE (37). The evening element was not overrepresented in any other circadian phase cluster. **(B)** Sequences of sites 1 and 2 are indicated in blue. **(C)** Mutation of the conserved AAAATATCT, but not a closely related motif, greatly reduced the ability of a promoter to confer circadian rhythmicity on a luciferase reporter gene in plants. Fusions to the firefly luciferase gene consisted of 450 bp (WT450), 130 bp (WT130, mut130_1, mut130_2, mut130_1,2), and 85 bp (WT85) of the *CCR2* promoter upstream of the putative transcriptional start site. Site 1 was replaced by gagcagctgc in constructs mut130_1 and mut130_1,2; site 2 was replaced by gagcagctgc in constructs mut130_2 and mut130_1,2. Constructs were introduced in *Arabidopsis* plants via *Agrobacterium*-mediated gene transfer (39). Luciferase assays were conducted and analyzed as described (5). Twelve T2 lines were examined for each construct.

Facteur de transcription

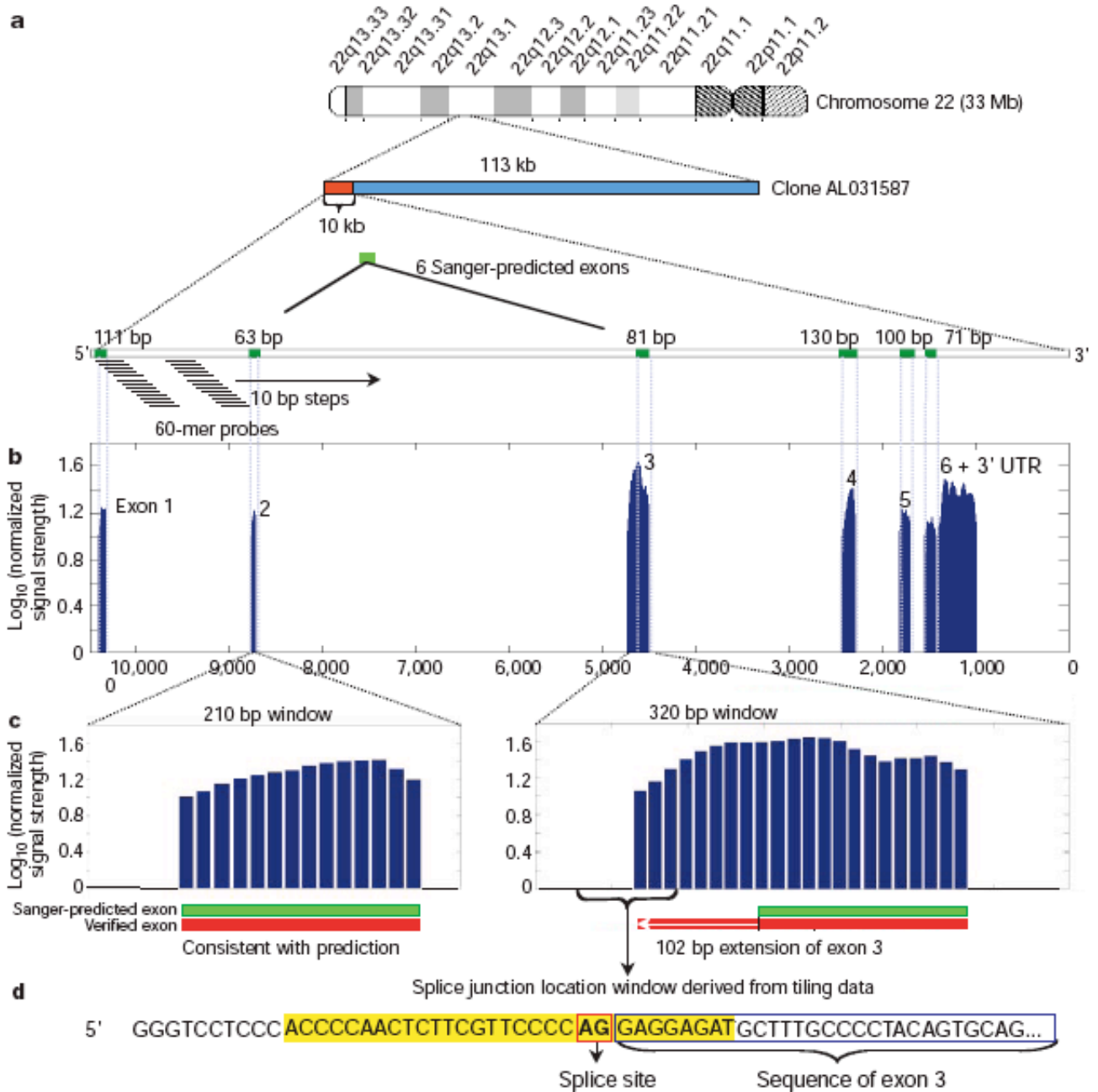


Glucose --> Galactose

Ren et al. (2000) Science 290:2306

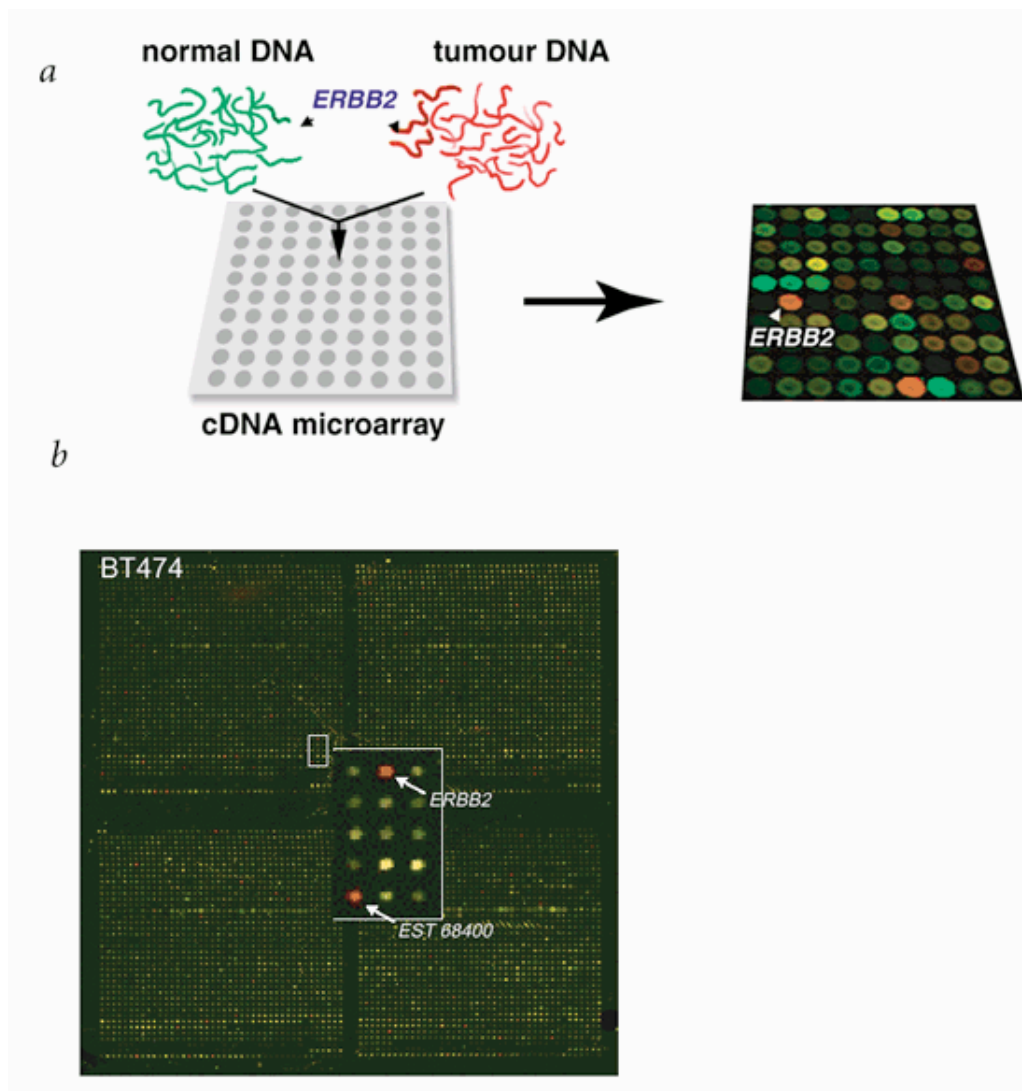


Annotation expérimentale du génome (tiling arrays)



Genome-wide analysis of DNA copy-number changes using cDNA microarrays

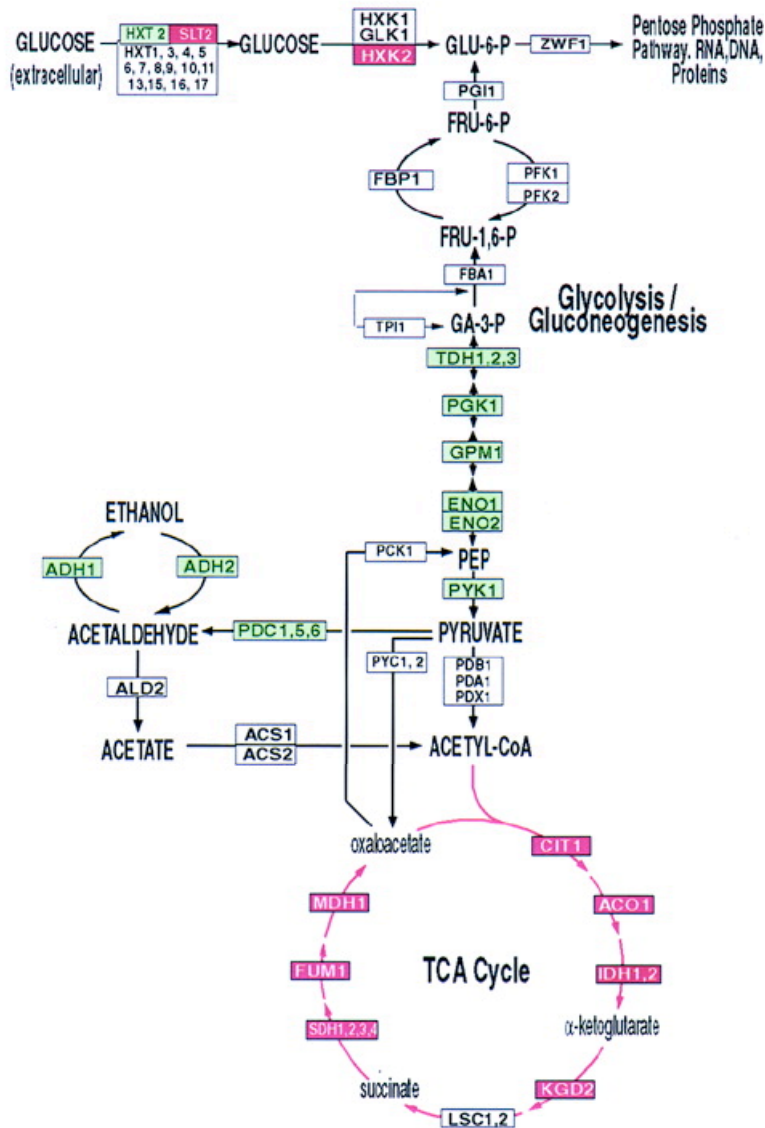
Jonathan R. Pollack¹, Charles M. Perou², Ash A. Alizadeh³, Michael B. Eisen², Alexander Pergamenschikov², Cheryl F. Williams², Stefanie S. Jeffrey⁴, David Botstein² & Patrick O. Brown^{1, 3}



Analyse de changements d'expression au cours de l'évolution

a

Parent vs Evolved strains



Evolution expérimentale

(250 générations)

Levures dans un milieu pauvre en glucose

Analyse globale de l'expression génétique (tous les gènes: 6124)

Résultats

3% de changement d'expression

Utilisation optimale du glucose pour la production d'ATP

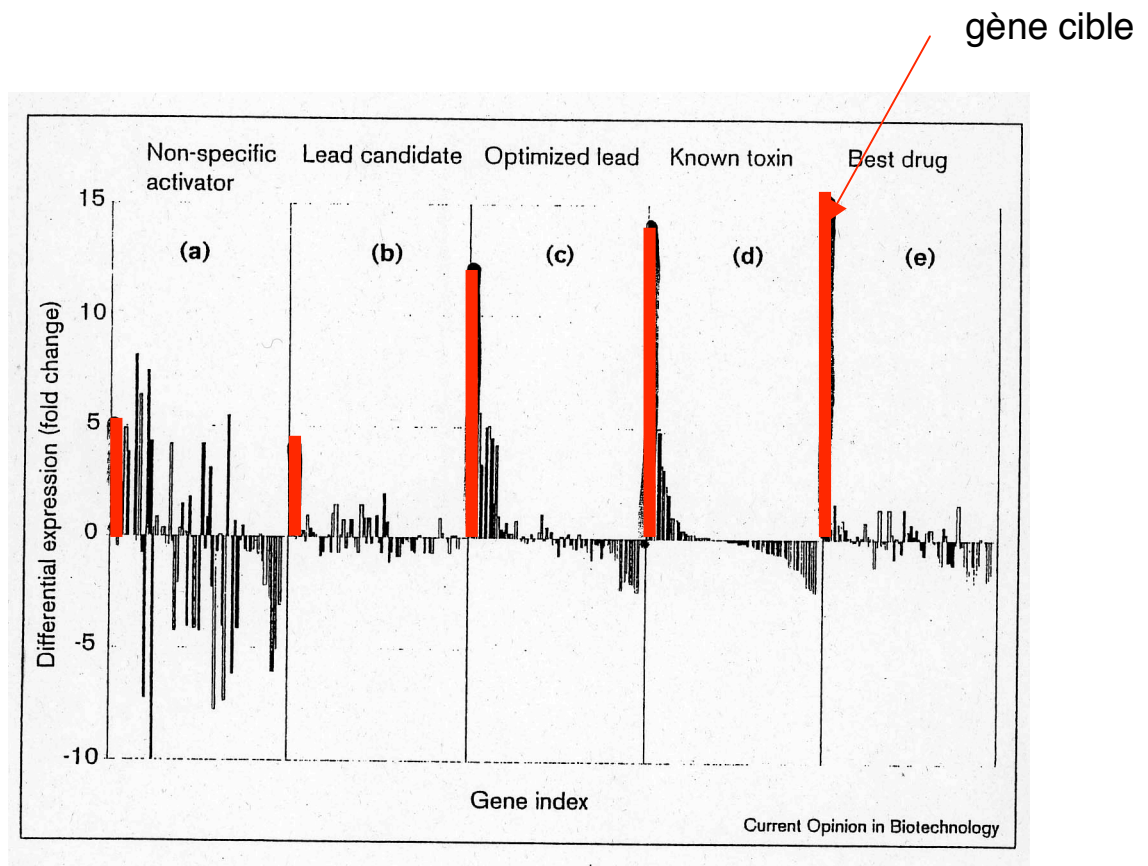
-diminution de la fermentation

-augmentation de la respiration

Discussion

L'identité des mutations n'est pas connue et ne peut être mesurée: on mesure le résultat et pas la cause

Développement de médicaments



Références

Publications

<http://www.nslj-genetics.org/microarray/>

nature genetics supplement -vol 21- january 1999

Eisen MB and Brown PO (1999) DNA arrays for analysis of gene expression.
Methods in Enzymology 303: 179-205

Hedge P et al. (2000) A concise guide to cDNA microarray analysis. BioTechniques
29: 548-562

Shena M. (2003) Microarray analysis. John Wiley & Sons, Hoboken, New Jersey.

Académique

<http://cmgm.stanford.edu/pbrown/>

<http://web.wi.mit.edu/young/>

<http://www.unil.ch/ibpv/microarrays.htm>

<http://www.affymetrix.com>

<http://www.catma.org>

Bases de données

<http://www.ncbi.nlm.nih.gov/geo/>

<http://www.ebi.ac.uk/arrayexpress/>