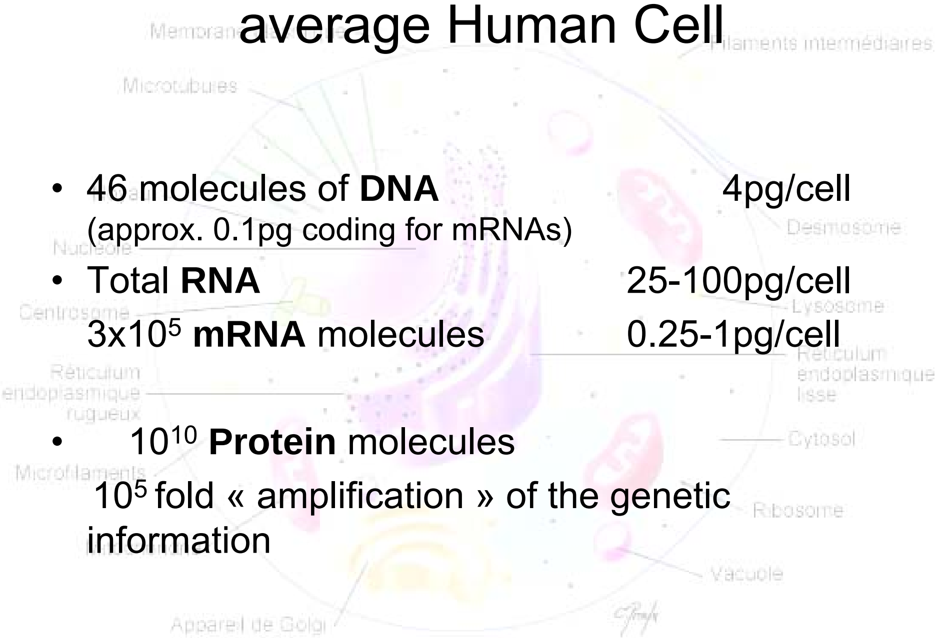
What you should know about : Affymetrix microarrays

October 8th 2010

A. Paillusson

DNA, RNA and Proteins in an average Human Cell

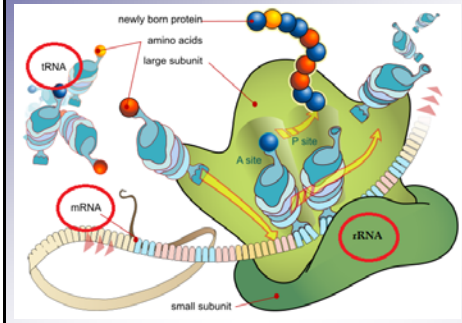


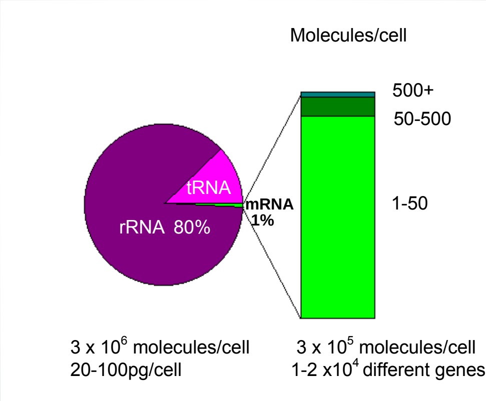
• 46 molecules of **DNA** (approx. 0.1pg coding for mRNAs)	4pg/cell
• Total **RNA**	25-100pg/cell
3×10^5 **mRNA** molecules	0.25-1pg/cell

- 10^{10} **Protein** molecules

10^5 fold « amplification » of the genetic information

mRNA Abundance in Mammalian Cells





Steady State Concentration of mRNA

- The steady state of eukaryotic mRNA concentration is the result of the equilibrium between synthesis and decay including all the possible intermediate processing and transport steps.

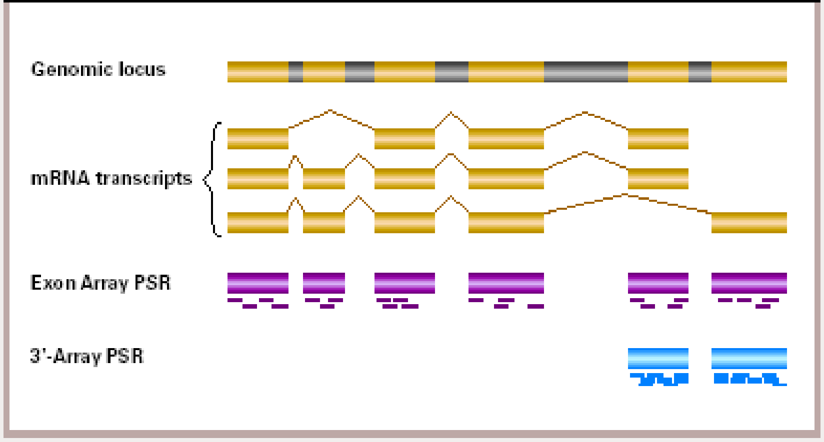
Some Gene Expression Facts for Eukaryotes

- Genes are expressed at varying levels, not just on/off
- A gene can give rise to different mRNAs (alternative promoters, splicing, polyA sites)
- Each cell type (and most probably each cell) has a different transcriptome
- When we measure mRNA concentrations we do not necessarily measure transcription

Transcriptome Analysis Technologies

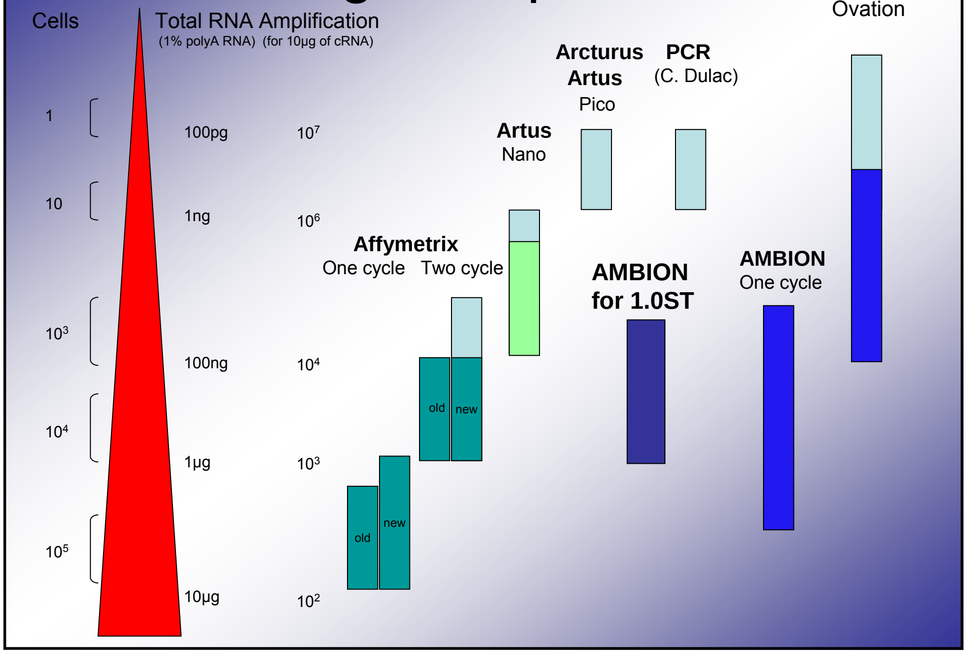
- All gene expression technologies have their advantages and disadvantages
- They should be viewed as complementary to each other and used for validation of results

Figure 2. Schematic for coverage of probe sets across the entire length of the transcript. Golden regions are exons whereas the grey regions represent introns that are removed during splicing. The short dashes underneath the exon regions for the Exon Array and 3'-Array PSR (probe selection region) indicate individual probes representing that PSR.



PSR : **P**robe **S**election **R**egion

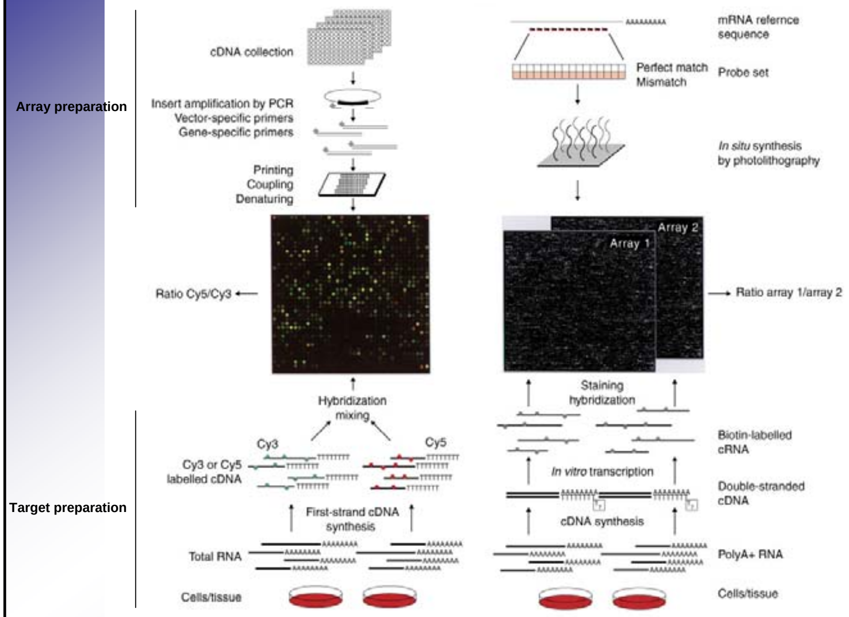
Target Amplification

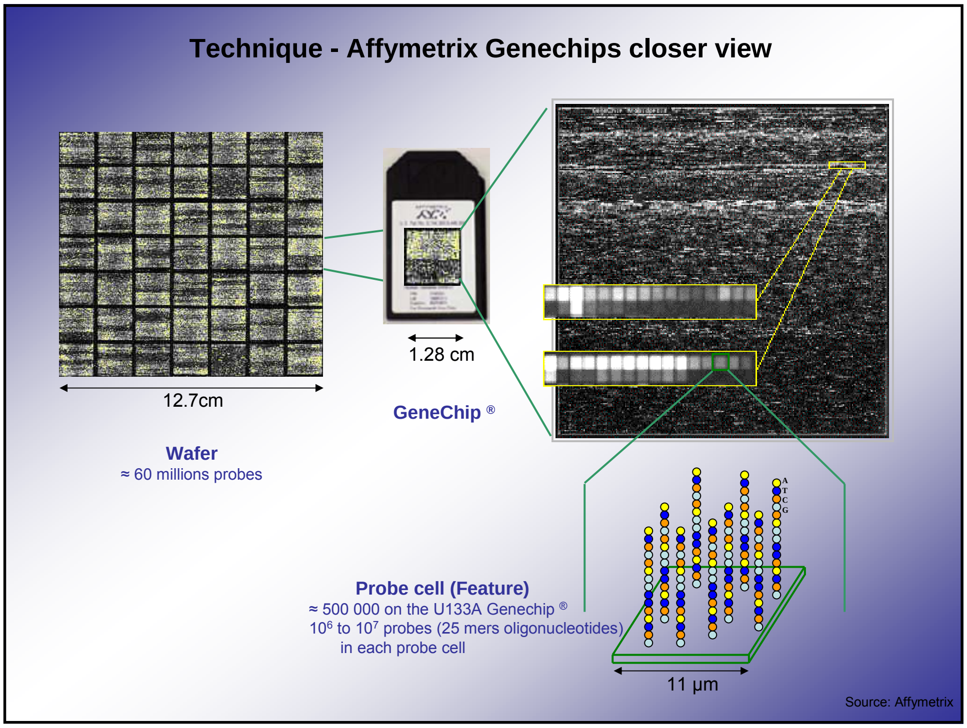


Terminology

Affymetrix uses 11-20 oligonucleotides of 25 nucleotides (25mer) to represent a gene.
Target: labeled RNA or DNA representing the transcriptome used for hybridization to the probes.
Some target labeling procedures rely on the polyA tail at the 3'end of mRNAs.
Probe: an oligonucleotide (25mer).
Perfect match (PM): A 25mer complementary to the reference sequence of a gene.
Mismatch (MM): same as PM but with a single nucleotide change in the middle (13th nt). The purpose of the MM is to measure non-specific hybridization and background noise.
Probe-pair: a PM,MM pair.
Probe set: a collection of probe-pairs (11-20) representing a gene.
Affy-ID: an identifier for a probe-set.

Printed DNA Arrays   Affymetrix Gene Chips





Technique - Affymetrix Genechips closer view

12.7cm

Wafer
≈ 60 millions probes

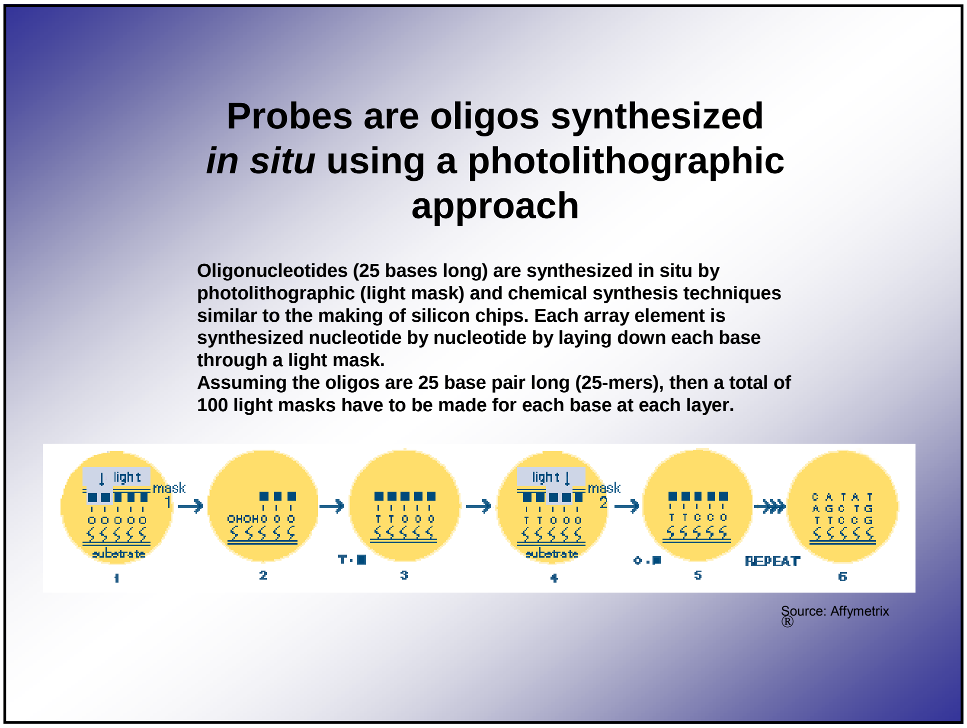
1.28 cm

GeneChip ®

Probe cell (Feature)
≈ 500 000 on the U133A Genechip ®
10^6 to 10^7 probes (25 mers oligonucleotides)
in each probe cell

11 µm

Source: Affymetrix



Probes are oligos synthesized *in situ* using a photolithographic approach

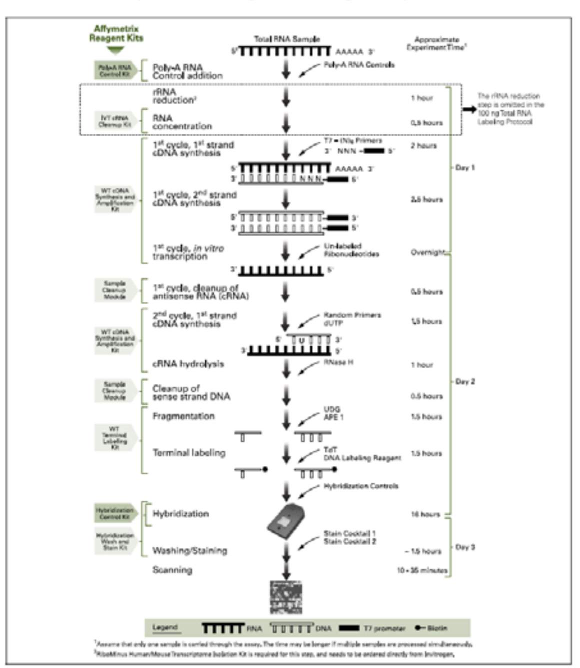
Oligonucleotides (25 bases long) are synthesized in situ by photolithographic (light mask) and chemical synthesis techniques similar to the making of silicon chips. Each array element is synthesized nucleotide by nucleotide by laying down each base through a light mask.
Assuming the oligos are 25 base pair long (25-mers), then a total of 100 light masks have to be made for each base at each layer.

Source: Affymetrix ®

Organisms coverage

- Human mouse rat (3' and exon arrays)
- A. thaliana, pseudomonas,
- Dog, pig,drosophila, monkey
- Custom arrays
- Other existing companies: Illumina,

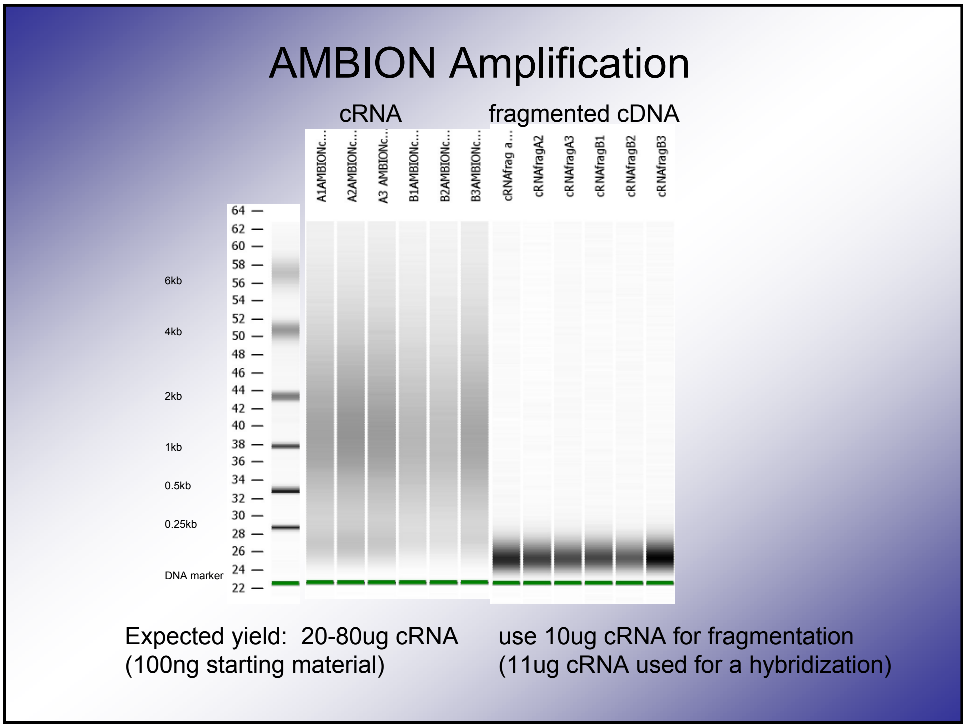
nimblegen



Whole Transcript Sense Target Labeling Assay Schematic

[1] Assume that only one sample is carried through the assay. The time may be longer if multiple samples are processed simultaneously.

[2] RiboMinus Human/Mouse Transcriptome Isolation Kit is required for this step, and needs to be ordered directly from Invitrogen.



AMBION Amplification

Expected yield: 20-80ug cRNA
(100ng starting material)

use 10ug cRNA for fragmentation
(11ug cRNA used for a hybridization)

Controls for Sample Processing and Hybridization

- **Internal Controls**

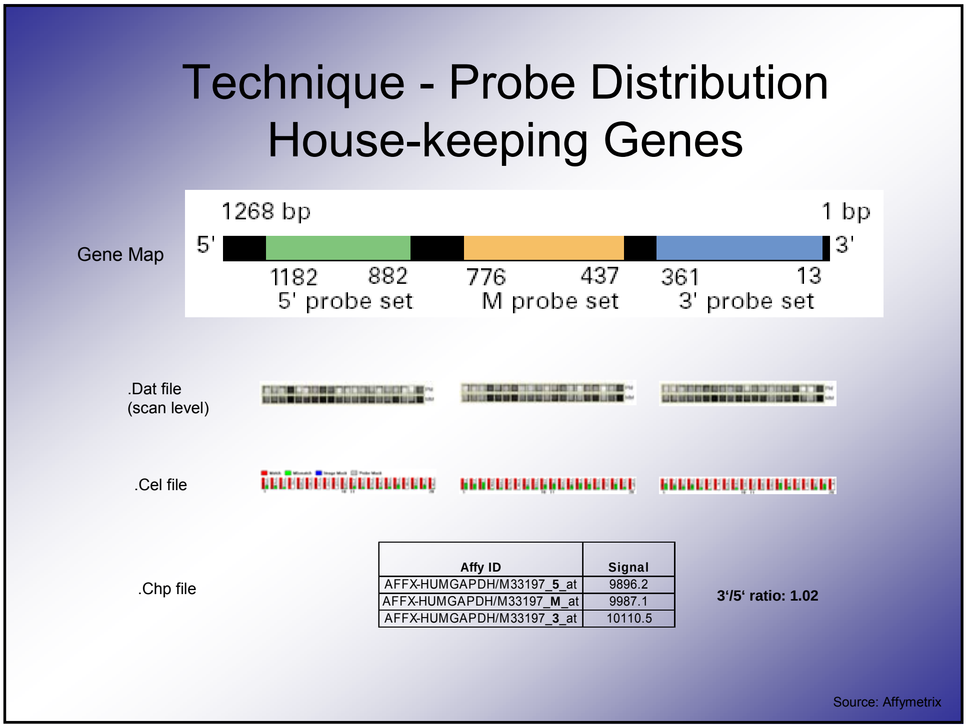
  RNAs in the sample known to be (or not) expressed and known to differ or be equal between different samples (eg. GAPDH).

  They indicate whether the RNA population has been faithfully represented.

- **External Controls (PolyA controls; Hybridization controls)**

  RNAs added in known amounts to the sample that aren’t otherwise there (e.g. bacterial RNAs in mammmalian samples).

  They indicate how well target labeling and array processing worked.



Technique - Probe Distribution House-keeping Genes

Gene Map

1268 bp

1 bp

5'

3'

1182 882 5' probe set

776 437 M probe set

361 13 3' probe set

.Dat file
(scan level)

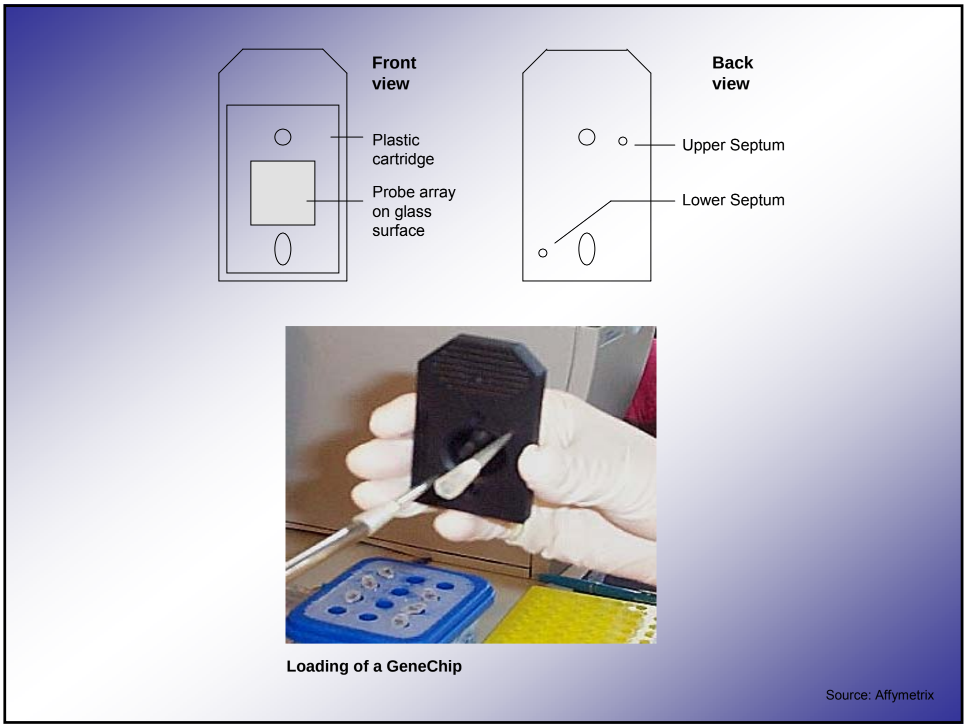
.Cel file

.Chp file

Affy ID	Signal
AFFX-HUMGAPDH/M33197_**5**_at	9896.2
AFFX-HUMGAPDH/M33197_**M**_at	9987.1
AFFX-HUMGAPDH/M33197_**3**_at	10110.5

3'/5' ratio: 1.02

Source: Affymetrix



Front view

Plastic cartridge

Probe array on glass surface

Back view

Upper Septum

Lower Septum

Loading of a GeneChip

Source: Affymetrix

Technique - Affymetrix Platform

GeneChip Hybe Oven 640

Fluidics Station 450

GeneArray Scanner 3000 and workstation

Target Hybridization

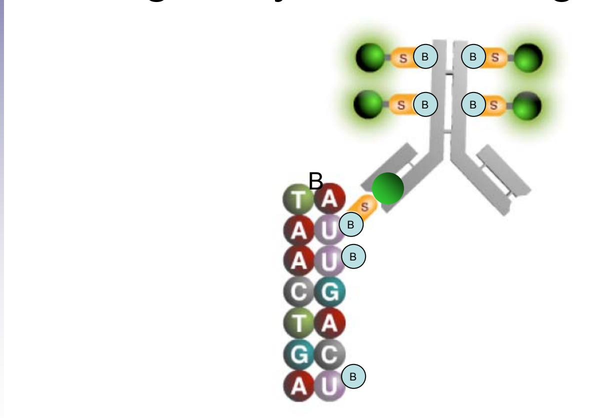
Washing and staining

Scanning

Image processing

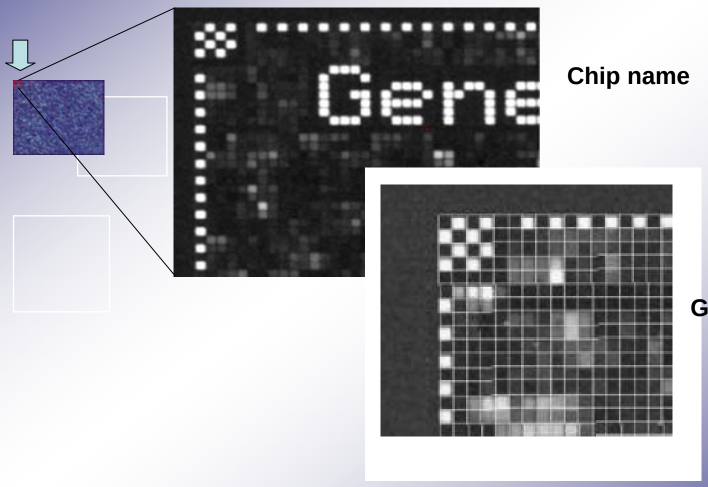
Source: Affymetrix

Staining of Hybridized Target



Streptavidin conjugated to Phycoerythrin (SAPE Fluorochrome) binds to the biotinylated UMP of the target; the signal is amplified by the SAPE which is binding to the multiple Biotins on the anti-Streptavidin antibody

B2 Oligo

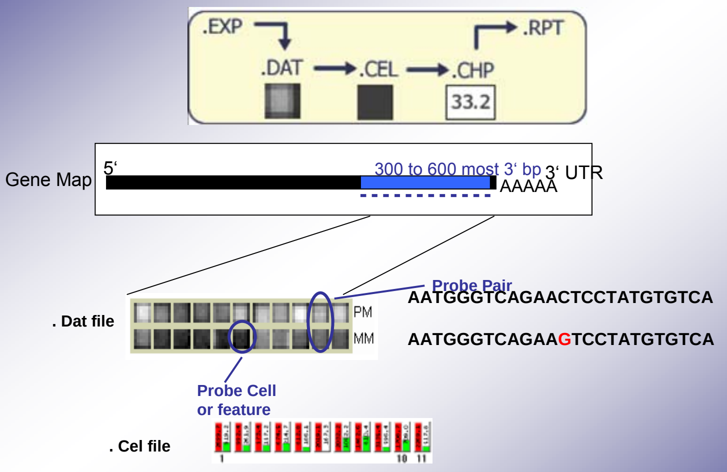


Chip name

Grid alignment

Source: Affymetrix

Technique - Probe Distribution Standard Genes

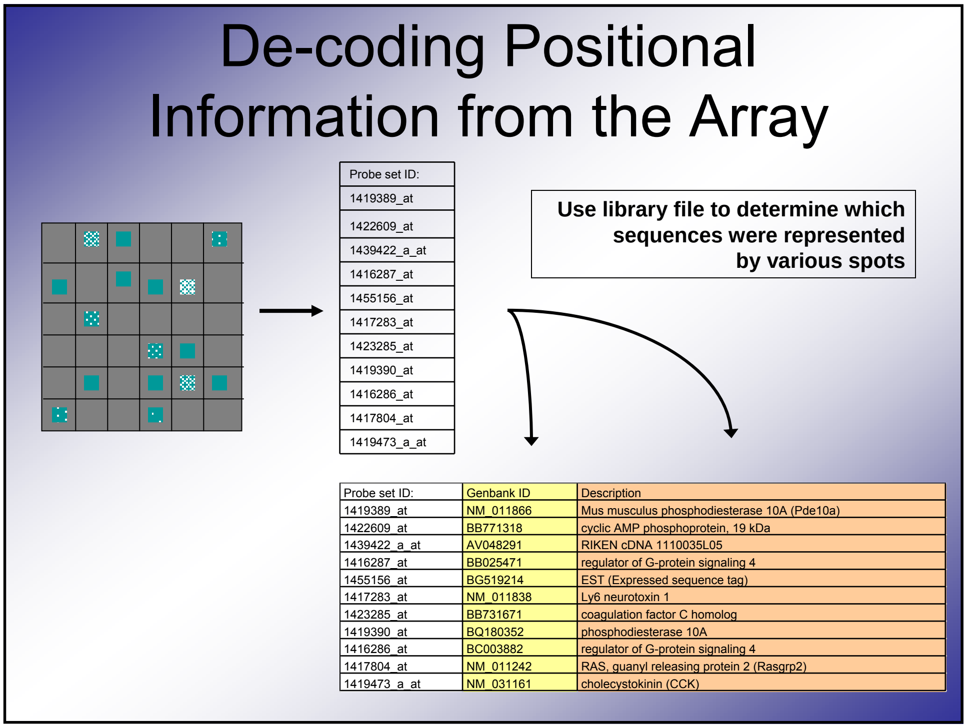


AATGGGTCAGAACTCCTATGTGTCA

AATGGGTCAGAAGTCCTATGTGTCA

. Chp file

Affy ID	Stat Pairs	Stat Pairs used	Signal	Flag	Detection p-value
222081_at	11	11	7.7	A	0.633789
222080_s_at	11	11	1025.2	P	0.432373
222079_at	11	11	48.6	P	0.000244
222078_at	11	11	2.7	A	0.850342
222077_s_at	11	11	68.7	P	0.005859
222076_at	11	11	25	M	0.72583

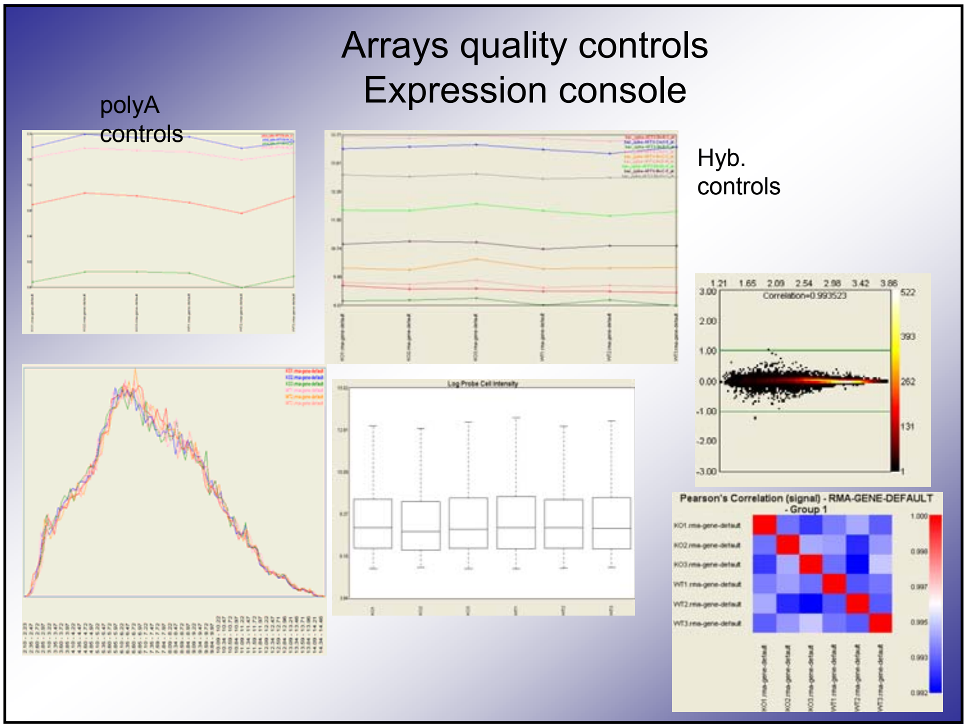


De-coding Positional Information from the Array

Probe set ID:
1419389_at
1422609_at
1439422_a_at
1416287_at
1455156_at
1417283_at
1423285_at
1419390_at
1416286_at
1417804_at
1419473_a_at

Use library file to determine which sequences were represented by various spots

Probe set ID:	Genbank ID	Description
1419389_at	NM_011866	Mus musculus phosphodiesterase 10A (Pde10a)
1422609_at	BB771318	cyclic AMP phosphoprotein, 19 kDa
1439422_a_at	AV048291	RIKEN cDNA 1110035L05
1416287_at	BB025471	regulator of G-protein signaling 4
1455156_at	BG519214	EST (Expressed sequence tag)
1417283_at	NM_011838	Ly6 neurotoxin 1
1423285_at	BB731671	coagulation factor C homolog
1419390_at	BQ180352	phosphodiesterase 10A
1416286_at	BC003882	regulator of G-protein signaling 4
1417804_at	NM_011242	RAS, guanyl releasing protein 2 (Rasgrp2)
1419473_a_at	NM_031161	cholecystokinin (CCK)



Arrays quality controls Expression console

polyA controls

Hyb. controls

Pathway analysis

- Statistical tools (R, Genespring…)
- IPA Ingenuity pathway analysis
- GeneGO
- Free software : panther, DAVID
- Genevestigator
- …
- Thank you for your attention!
- Questions?
- Visit!