

Microarrays

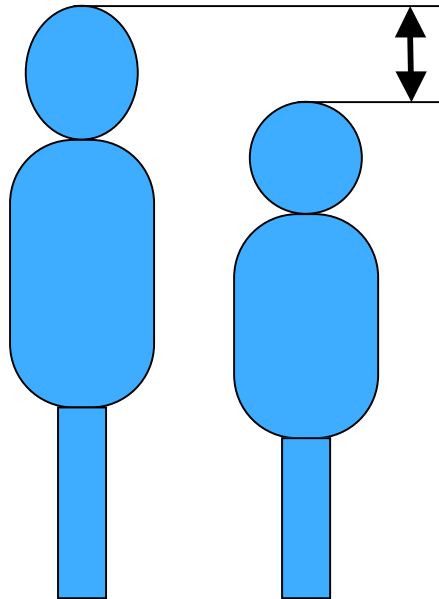
Γρηγόρης Αμούτζιας

Microarrays

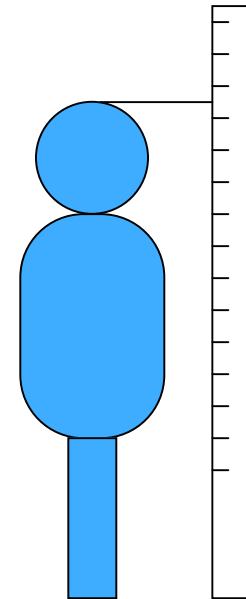
- Chips ή αντικειμενοφόροι πλάκες που μετράνε την γονιδιακή έκφραση (και όχι μόνο).
- Δεν μετράνε απόλυτες τιμές συγκεντρώσεων μεταγραφημάτων.
- Μετράνε σχετικές αλλαγές.
 - Για μια συνθήκη, πόσο πιο αυξημένη ή μειωμένη είναι η έκφραση ενός γονιδίου, σε σχέση με μια άλλη συνθήκη.
 - Π.χ. Πόσο πιο αυξημένη/μειωμένη είναι η έκφραση ενός γονιδίου A μετά από άσκηση (control: αμέσως πριν την άσκηση).
- Δεν μπορούμε να συγκρίνουμε το γονίδιο A με το γονίδιο B
- Στο Chip υπάρχουν για το κάθε μεταγράφημα:
 - cDNAs ολόκληρου του μεταγραφώματος (spotting) (PCR)
 - ολιγομερή (τα λεγόμενα oligo-probes), με μήκος 25-70 βάσεις.
 - Spotting
 - In situ synthesis

Σχετικές διαφορές

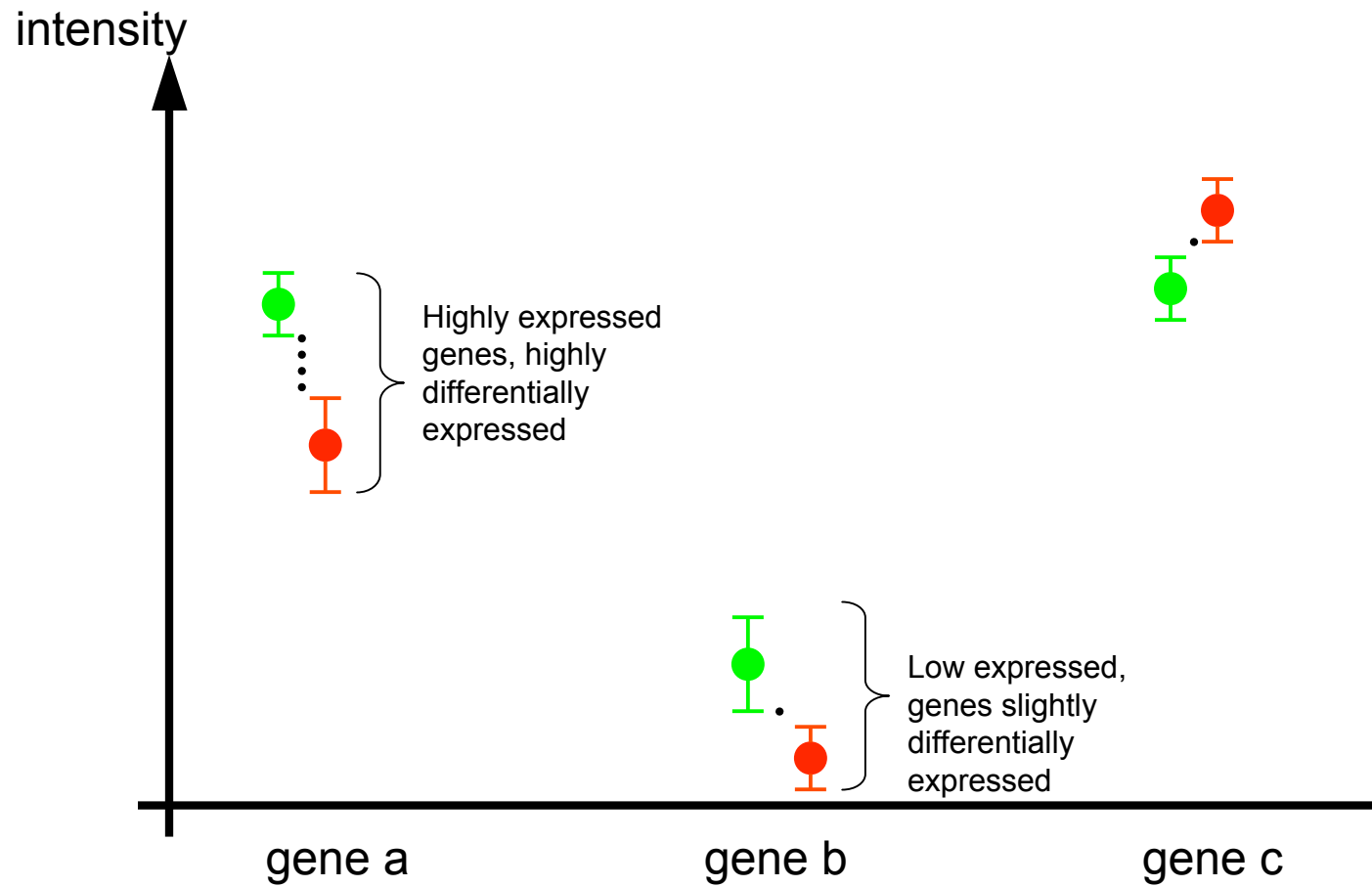
relative measurement 2
COLORS (Cy5 / Cy3)



absolute measurement
1 COLOR (Cy3)



Σχετικές διαφορές



Όλιγο-Μικροσυστοιχίες (oligo-microarrays)

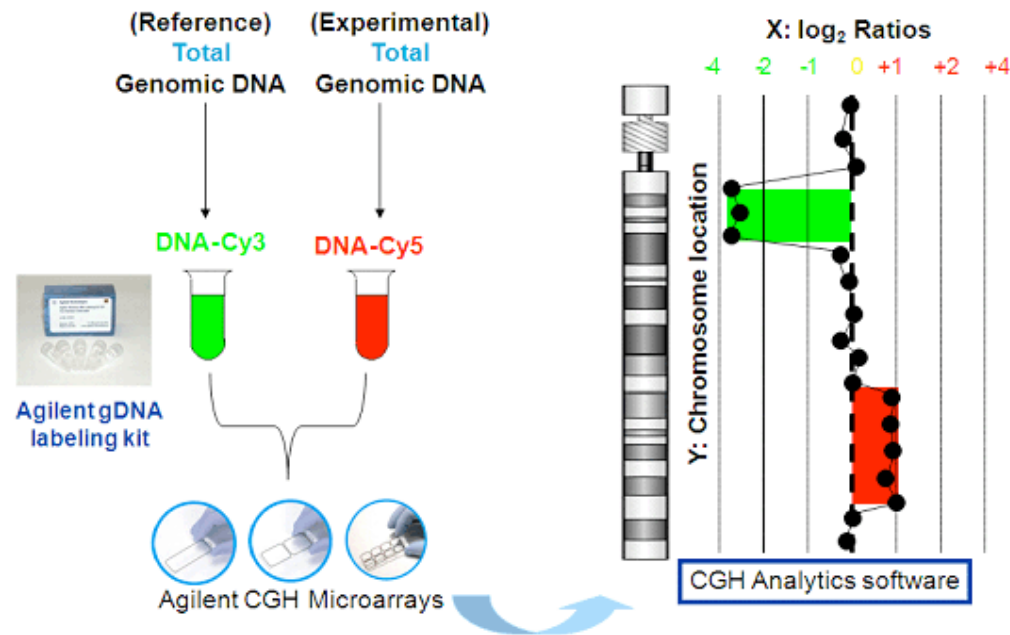
- Chip στο οποίο υπάρχουν DNA ολιγομερή (τα λεγόμενα probes), με μήκος 25-70 βάσεις που αντιπροσωπεύουν το γονιδίωμα ενός οργανισμού.
- Περισσότερα από ένα διαφορετικά probes μπορεί να υπάρχουν για ένα γονίδιο.
- Πάνω σε αυτά τα probes υβριδίζονται τα συμπληρωματικά τους cDNAs.
- Το κάθε cDNA είναι συνδεδεμένο με φθορίζουσα χρωστική.
- Κάθε κηλίδα (spot) στο chip αντιστοιχεί σε ένα συγκεκριμένο είδος probes.
- Η ένταση της φθορίζουσας χρωστικής υποδηλώνει την ένταση με την οποία εκφράζεται το mRNA για τον αντίστοιχο probe.

Τι μπορούν να ελέγξουν

- Έκφραση mRNA.
- Έκφραση micro-RNA.
- CGH (comparative genomic hybridization).
- SNPs.
- DNA μεθυλίωση.
- CHIP-on-chip.

CGH

- Comparative genome hybridization.
- Βλέπει αν μια γενωμική περιοχή σε ένα δείγμα έχει πολλαπλασιαστεί ή χαθεί σε σχέση με ένα control.



Τα βήματα

From: [PLoS Comput Biol. 2010 May; 6\(5\): e1000786.](#)
Published online 2010 May 27. doi: 10.1371/journal.pcbi.1000786.
[Copyright/License](#)

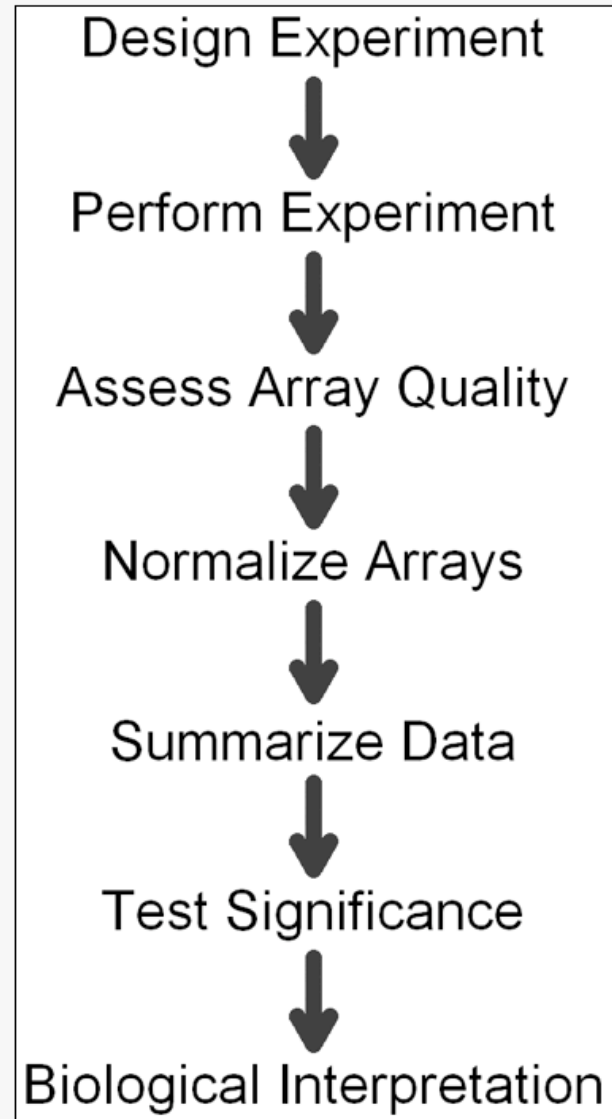
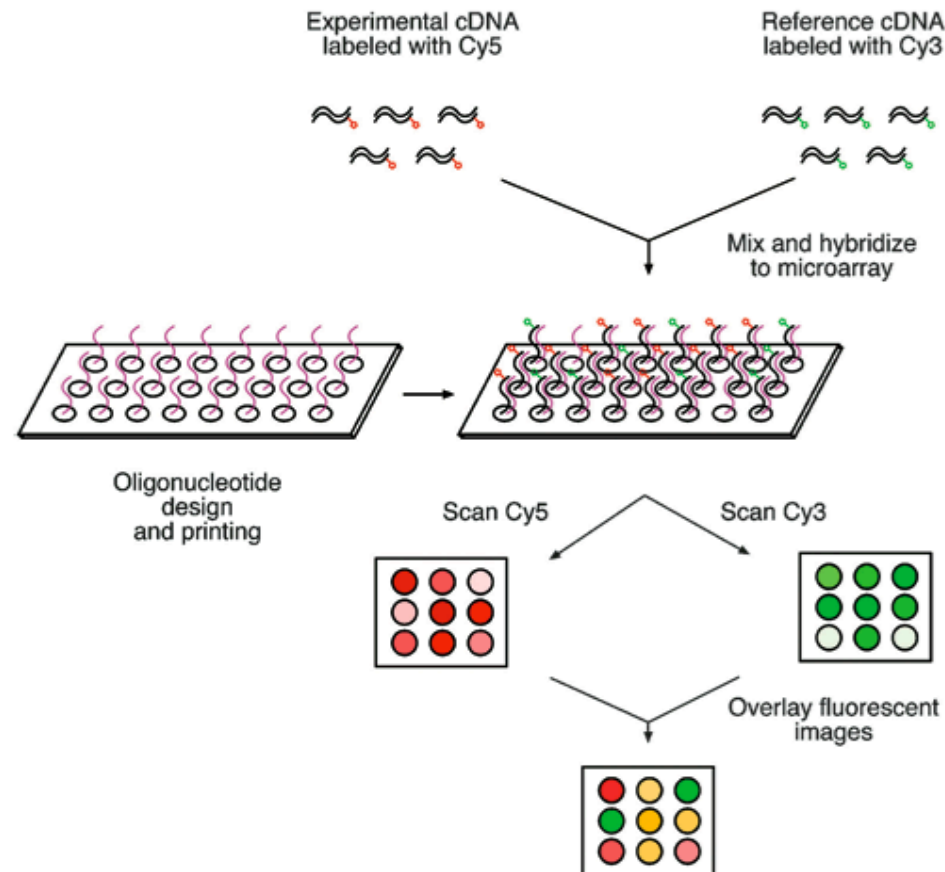


Figure 1
Steps in a typical microarray analysis.

Μικροσυστοιχίες (microarrays)

- Με δύο χρώματα (πράσινο/κόκκινο)
- Το ένα χρώμα είναι για μια συγκεκριμένη συνθήκη και το άλλο χρώμα για μια άλλη (συνήθως control).
- Εξάγεται το ολικό RNA για την κάθε συνθήκη.
- Ελέγχεται η ποσότητα και ποιότητα του RNA.
- Γίνεται σήμανση του κάθε δείγματος RNA με συγκεκριμένη χρωστική (π.χ. Cy5:red - Cy3:green).
- Τα 2 δείγματα αναμιγνύονται και το μίγμα υπόκειται σε υβριδισμό πάνω στο chip.
- Μετρείται η ένταση της κάθε μιας από τις 2 χρωστικές, για κάθε κηλίδα.
- Υπολογίζεται ο λόγος των εντάσεων των 2 χρωστικών, στην κάθε κηλίδα.

Μικροσυτοιχίες



Μικροσυστοιχίες (microarrays)

- Αν το γονίδιο εκφράζεται περισσότερο στην A συνθήκη (κόκκινη χρωστική) από ότι στην control (πράσινη χρωστική), τότε ο λόγος συνθήκη_A/control (κόκκινη/πράσινη) θα είναι $\lambda > 1$, αλλιώς σε αντίθετη περίπτωση $0 < \lambda < 1$.
- Αν το γονίδιο εκφράζεται με διπλάσια ένταση στην συνθήκη A, σε σχέση με την συνθήκη control, τότε ο λόγος θα είναι $\lambda = 2$.
- Αν το γονίδιο εκφράζεται με τη μισή ένταση στην συνθήκη A, σε σχέση με την συνθήκη control, τότε ο λόγος θα είναι $\lambda = 0.5$.
- Μετατρέποντας τους λόγους σε \log_2 , έχουμε:
 - $\lambda = 2 \rightarrow \log_2 \lambda = 1$
 - $\lambda = 0.5 \rightarrow \log_2 \lambda = -1$
 - Με την κανονικοποίηση σε \log_2 τα δεδομένα γίνονται συμμετρικά.

Μικροσυστοιχίες (microarrays)

- Πότε θεωρούμε ότι ένα γονίδιο υπερ/υπό-εκφράζεται σε μια συγκεκριμένη συνθήκη.
 - $\text{Log}_2\lambda > 1$ ή $\text{Log}_2\lambda < -1$ (διπλάσια/υποδιπλάσια έκφραση σε σχέση με τη συνθήκη control).
 - Με στατιστικές μεθόδους (t-test, ANOVA).

microarrays

- Κατανομή των σημάτων.
- Μετατροπή - κανονικοποίηση των σημάτων.

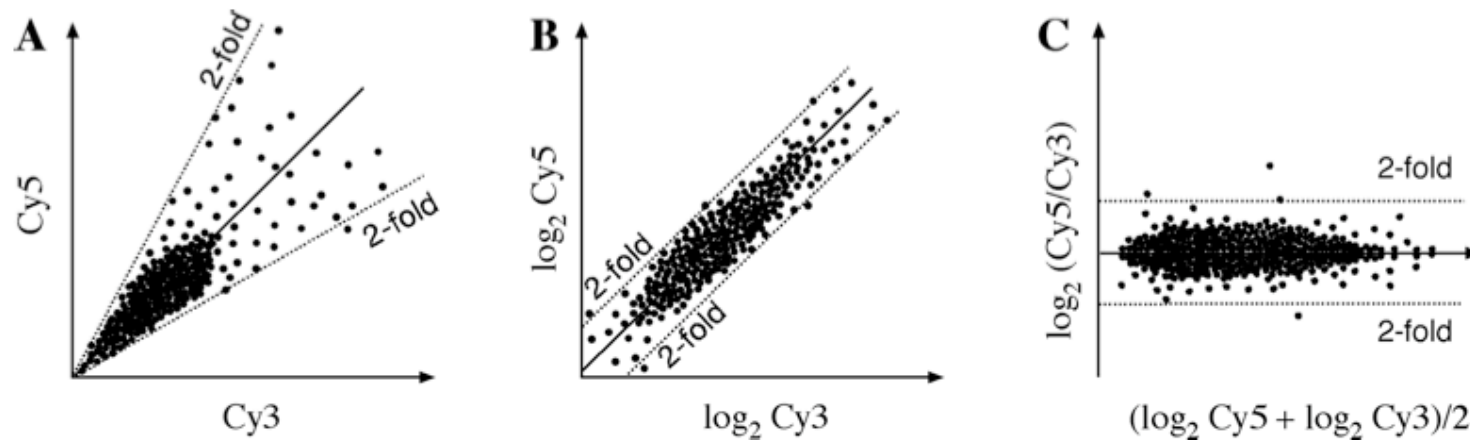


Figure 18.5: Scatter plot of gene expression analysis showing the process of data normalization. The solid line indicates linear regression of the data points; dashed lines show the cutoff for a twofold change in expression. **(A)** Plot of raw fluorescence signal intensities of Cy5 versus Cy3. **(B)** Plot of the same data after log transformation to the base of 2. **(C)** Plot of mean log intensity versus log ratio of the two fluorescence intensities, which shifts the data points to around the horizontal axis, making them easier to visualize.

Ομαδοποίηση γονιδίων με την ίδια συμπεριφορά.

- Χρειαζόμαστε αρκετά σημεία (διαφορετικές συνθήκες ή χρονικές στιγμές)
- Με μεθόδους αποστάσεων, όπου οι μετρήσεις ενός γονιδίου για διαφορετικές συνθήκες αποτελούν ένα διάνυσμα.
- Υπολογίζουμε αποστάσεις μεταξύ διαφορετικών διανυσμάτων (γονιδίων).
 - Ευκλείδια απόσταση
 - Συντελεστής συσχέτισης Pearson (Pearson correlation coefficient).
 - Δημιουργείται πίνακας αποστάσεων μεταξύ των γονιδίων.
 - Το αντίστοιχο μπορεί να γίνει και για να ομαδοποιήσουμε κοινές συνθήκες.

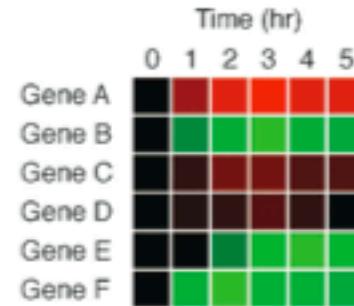
Μικροστοιχίες

	0 hr	1 hr	2 hr	3 hr	4 hr	5 hr
Gene A	1	4	6	8	6	6
Gene B	1	0.6	0.3	0.1	0.3	0.4
Gene C	1	2	4	4	3	3
Gene D	1	1.5	2	3	2	1
Gene E	1	1	0.5	0.2	0.1	0.2
Gene F	1	0.3	0.1	0.2	0.3	0.4

convert to false colors



\log_2 conversion



	Gene B	Gene C	Gene D	Gene E	Gene F
Gene A	-0.82	0.96	0.65	-0.68	-0.79
Gene B		-0.85	-0.86	0.66	0.67
Gene C			0.70	-0.65	-0.87
Gene D				-0.41	-0.72
Gene E					0.26

calculating Pearson correlation coefficients between genes



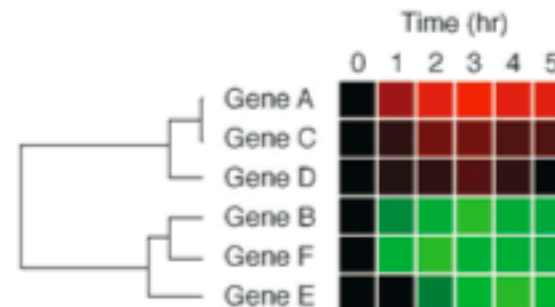
	0 hr	1 hr	2 hr	3 hr	4 hr	5 hr
Gene A	0	2	2.6	3	2.6	2.6
Gene B	0	-0.7	-1.7	-3.3	-1.7	-1.3
Gene C	0	1	2	2	1.6	1.6
Gene D	0	0.6	1	1.6	1	0
Gene E	0	0	-1	-2.3	-3.3	-2.3
Gene F	0	-1.7	-3.3	-2.3	-1.7	-1.3



conversion of coefficients to positive distance values

	Gene B	Gene C	Gene D	Gene E	Gene F
Gene A	1.82	0.04	0.35	1.68	1.79
Gene B		1.85	1.86	0.34	0.33
Gene C			0.30	1.65	1.87
Gene D				1.41	1.72
Gene E					0.74

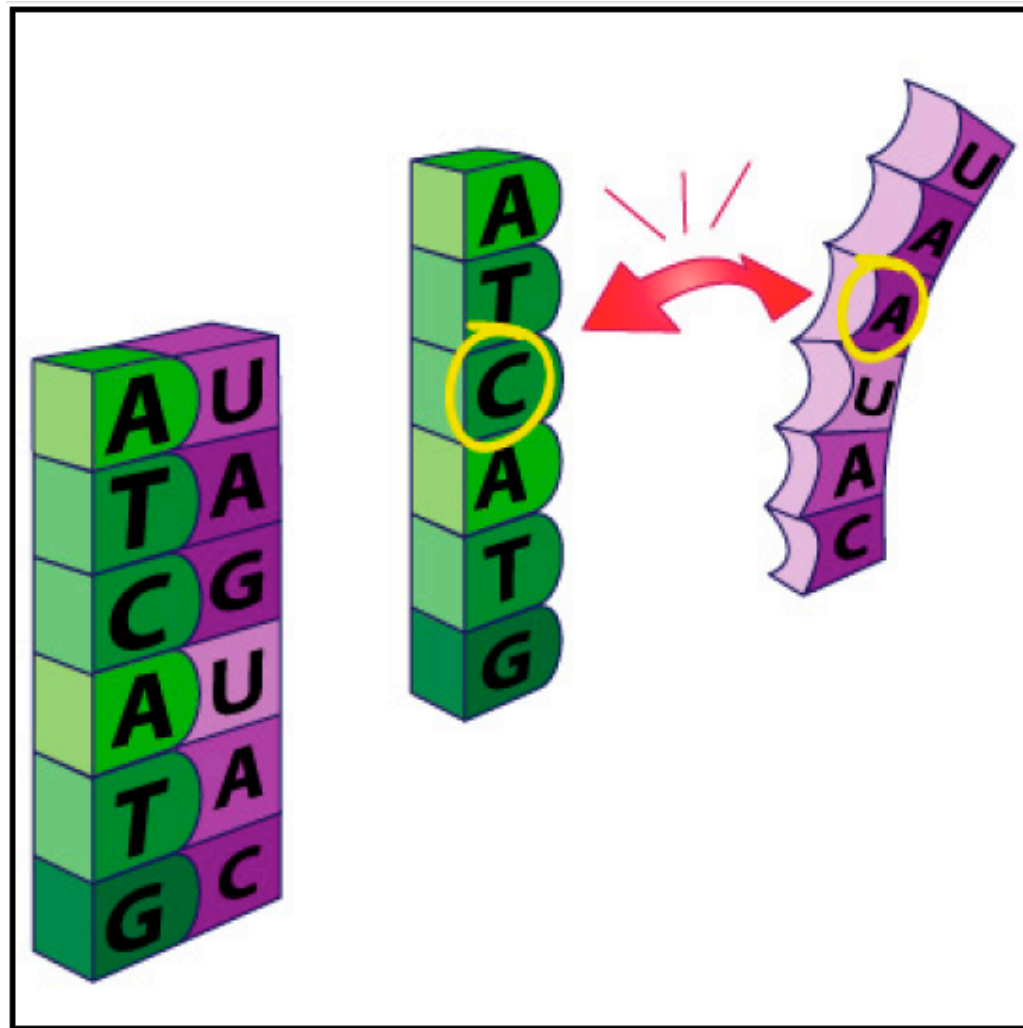
hierarchical clustering



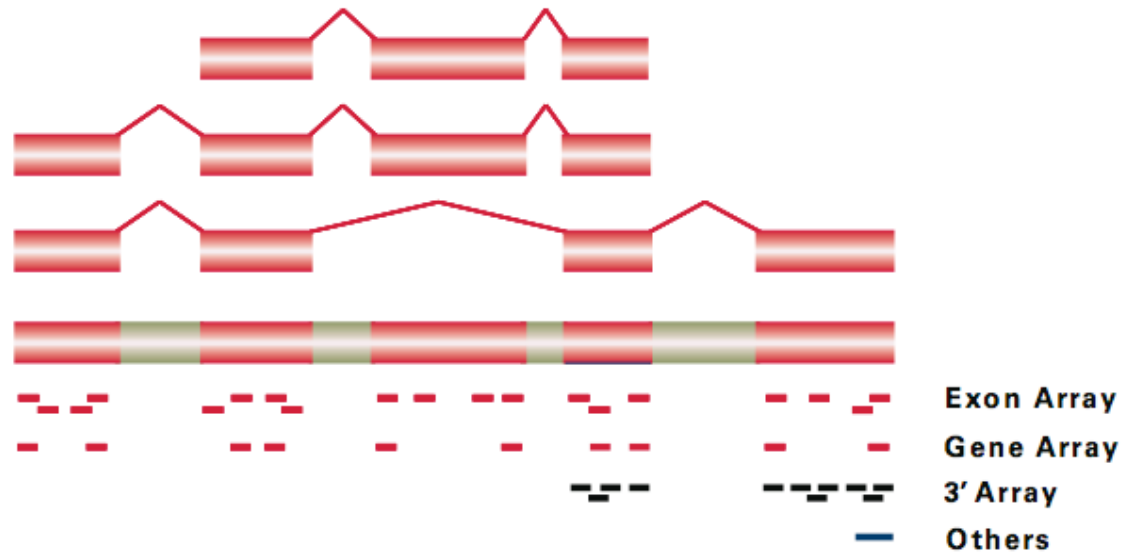
Μήκος των probes

- Affymetrix: 25mer
- Agilent: 60mer
- Illumina: 50mer
- Όσο μεγαλύτερο το probe, τόσο πιο ευαίσθητο (μπορεί να ανιχνεύσει targets σε μικρότερες συγκεντρώσεις).
- Όμως, μεγαλύτερο probe μπορεί και να ανεχτεί περισσότερους πολυμορφισμούς (ή mismatches).

Υβριδισμός των probes



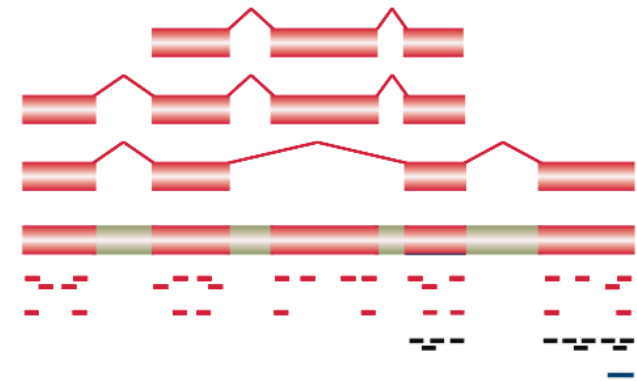
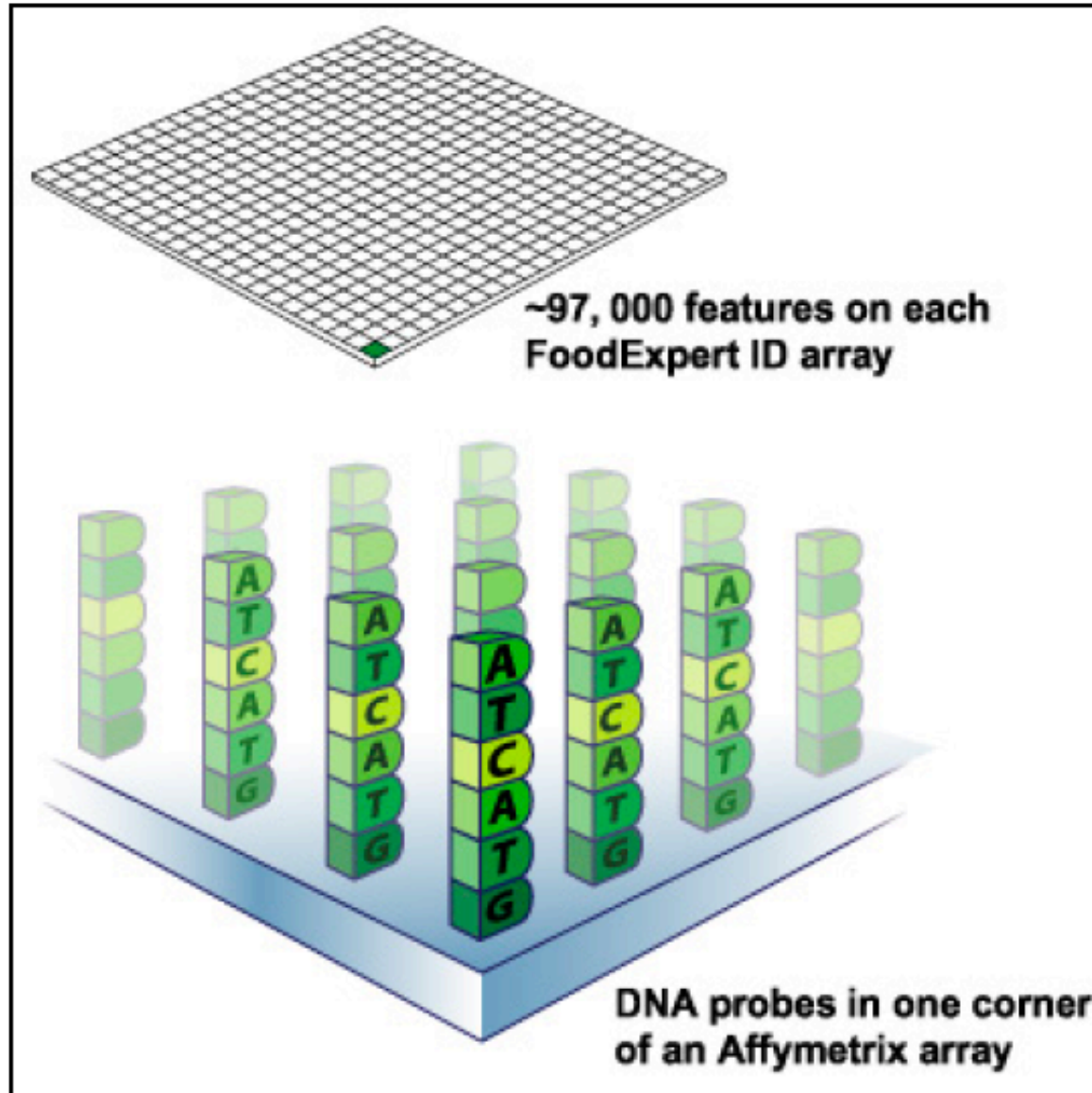
Probes & chips



Σχεδιασμός των probes

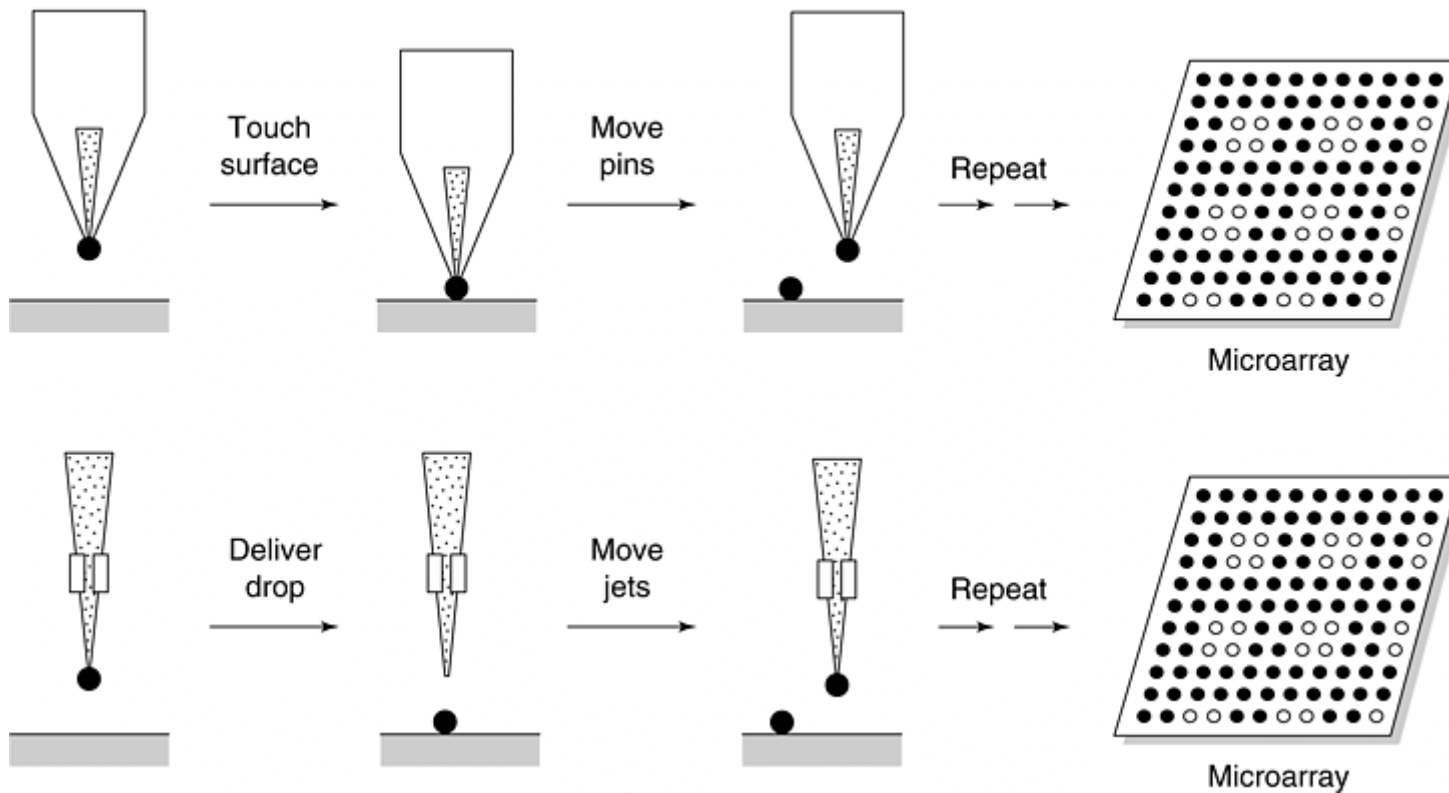
- Θα πρέπει η ακολουθία τους να είναι μοναδική ώστε να αντικατοπτρίζει την έκφραση για το γονίδιο που σχεδιάστηκαν.
- Όλοι οι probes θα πρέπει να έχουν παρόμοια T_m (βέλτιστη θερμοκρασία υβριδισμού).
- Τα probes δεν θα πρέπει να σχηματίζουν δευτεροταγείς δομές που τα εμποδίζουν να υβριδιστούν με τον στόχο.

spots



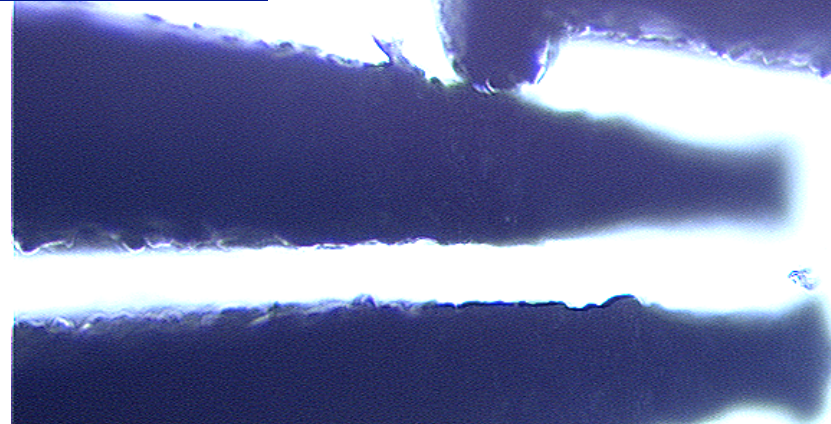
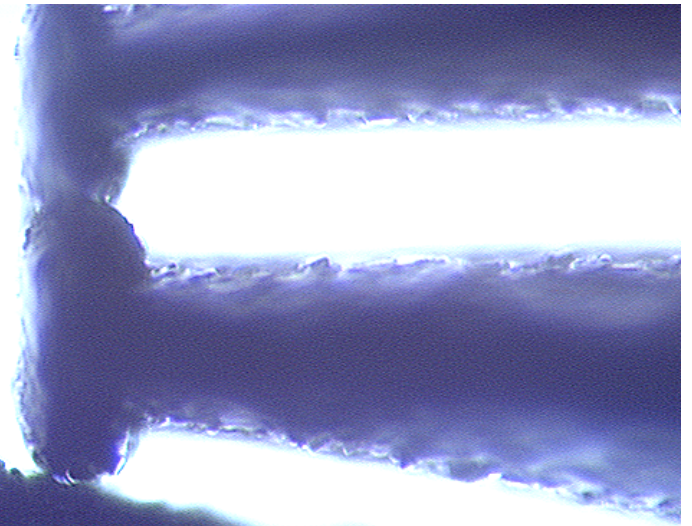
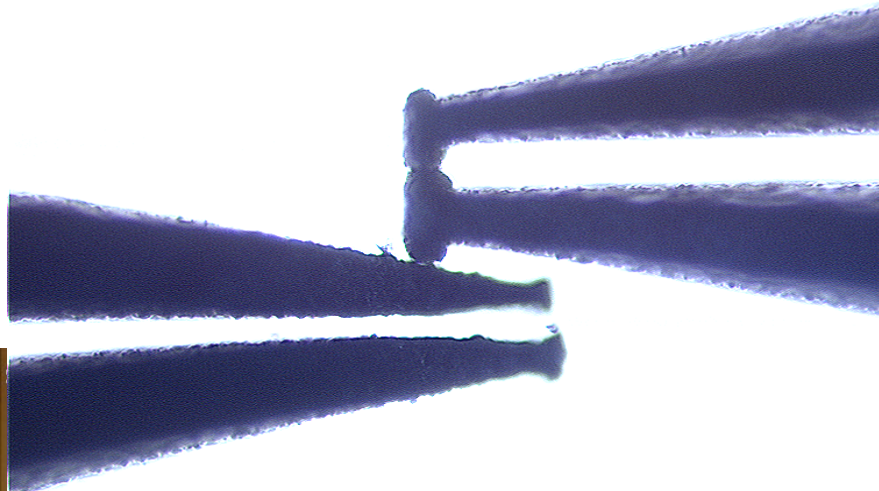
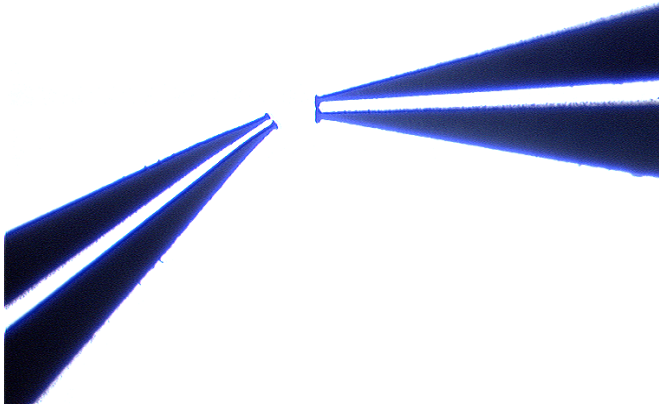
Array Spotting

- Με βελόνες (pins) (απώθηση έτοιμων cDNAs ή oligos).
- Με inkjet (Agilent: spin-off της Hewlett-Packard) (in situ synthesized oligos)



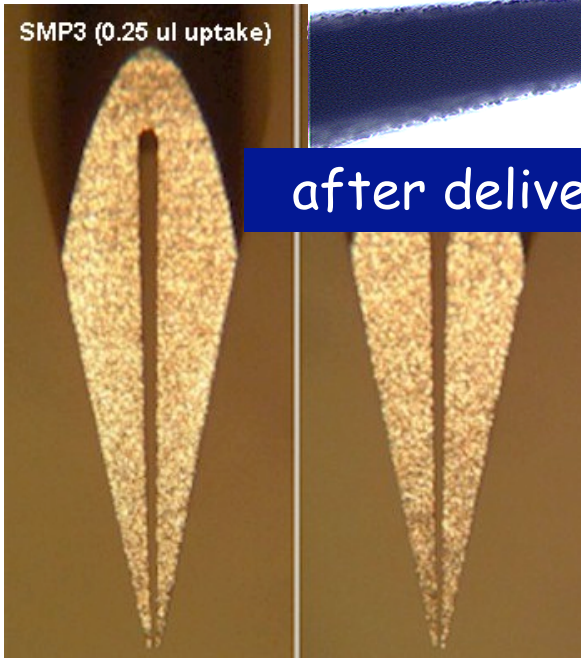
spotting pin quality decline

after delivery of 5×10^5 spots



SMP3 (0.25 ul uptake)

after delivery of 3×10^5 spots



Inkjet printing - Agilent

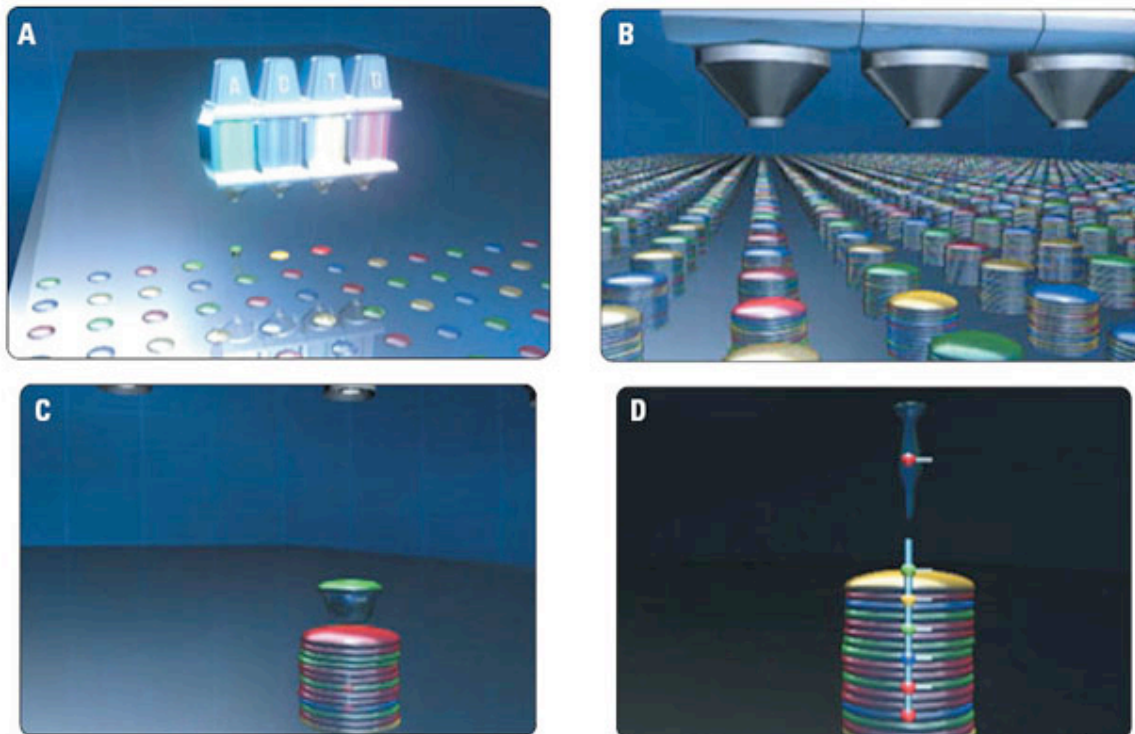


Figure 1. These four images communicate the general mechanism for oligo synthesis via inkjet printing. A: the first layer of nucleotides is deposited on the activated microarray surface. B: growth of the oligos is shown after multiple layers of nucleotides have been precisely printed. C: close-up of one oligo as a new base is being added to the chain, which is shown in figure D.

Agilent phosphoramidite chemistry

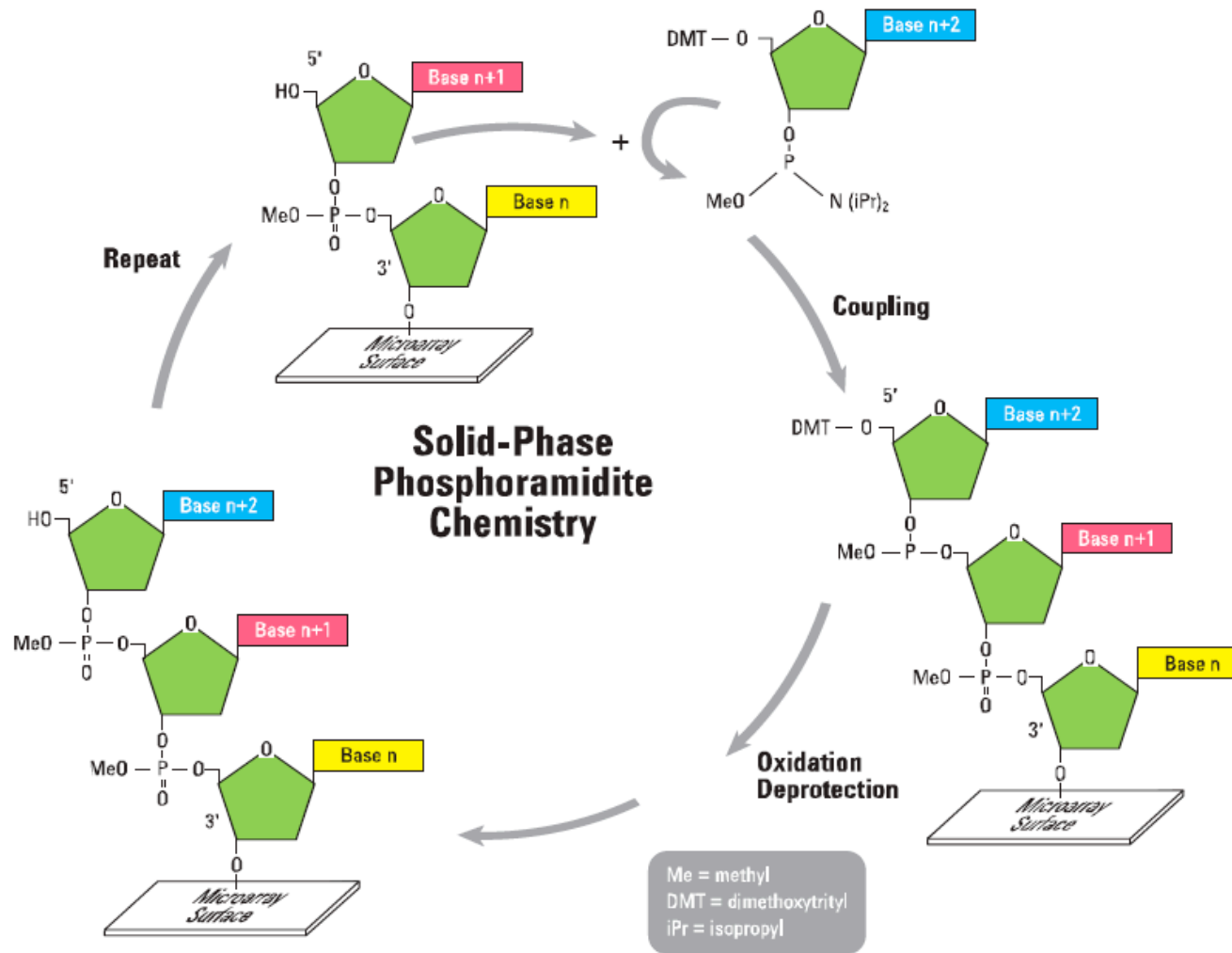


Figure 4. This figure shows the general cycle of oligo synthesis via phosphoramidite chemistry.

Φωτολιθογραφία - Affymetrix

Youtube:

<http://www.youtube.com/watch?v=ui4BOtwJEXs&feature=related>

Affymetrix:

http://media.affymetrix.com:80/about_affymetrix/outreach/lesson_plan/downloads/student_manual_activities/activity3/activity3_manufacturing_background.pdf

Φωτολιθογραφία

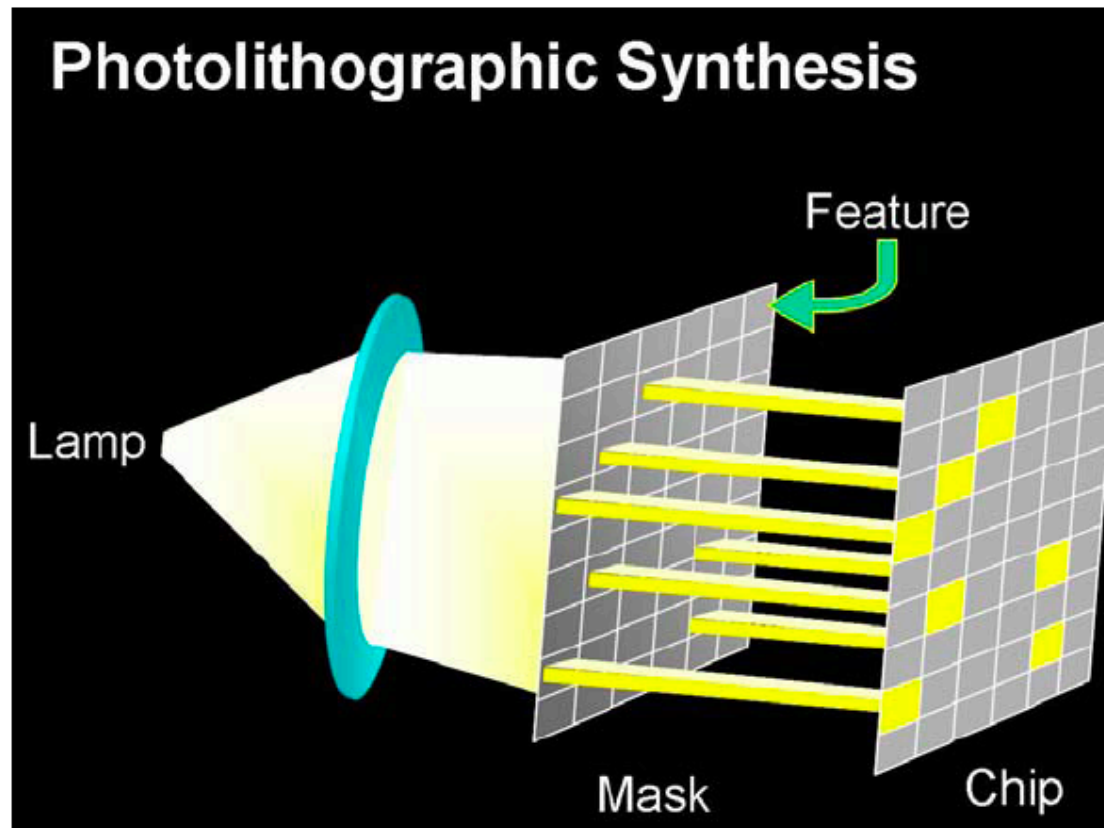


Fig. 1 - The mask only allows light to pass to specific features on the chip

Φωτολιθογραφία

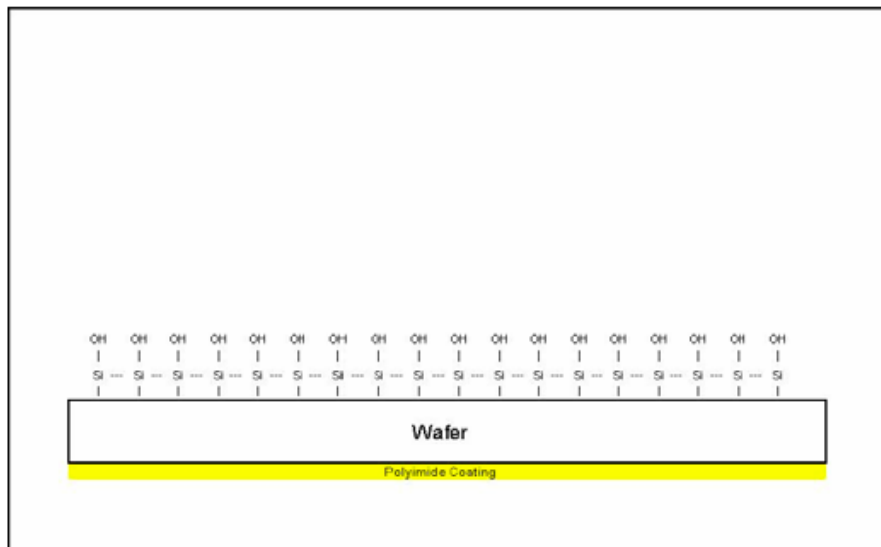


Fig. 3 - Silanation - each "Si" is a starting point

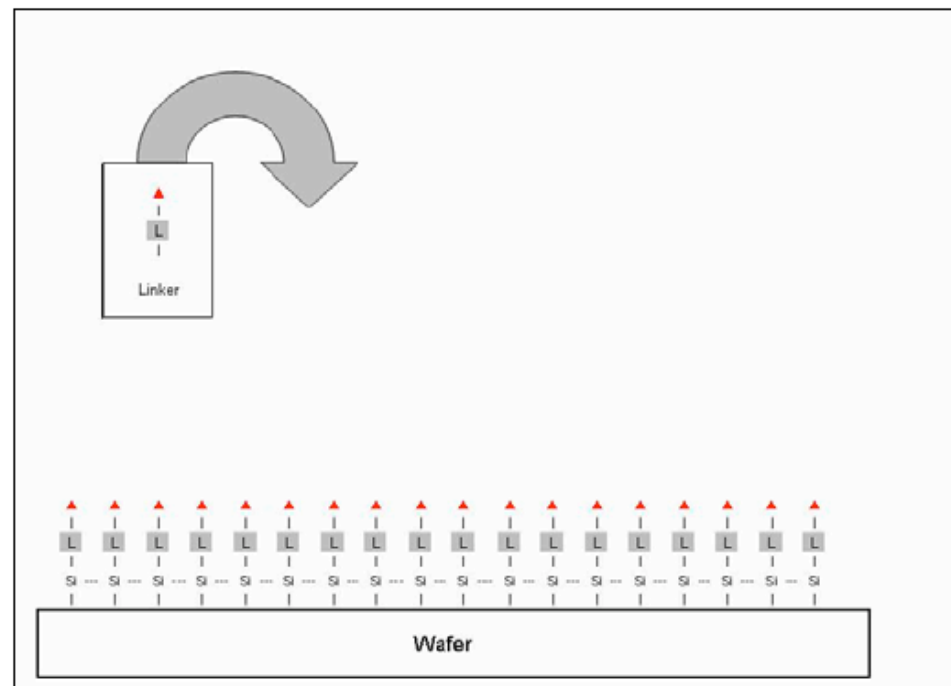


Fig. 4 - The addition of the linker molecule

Φωτολιθογραφία

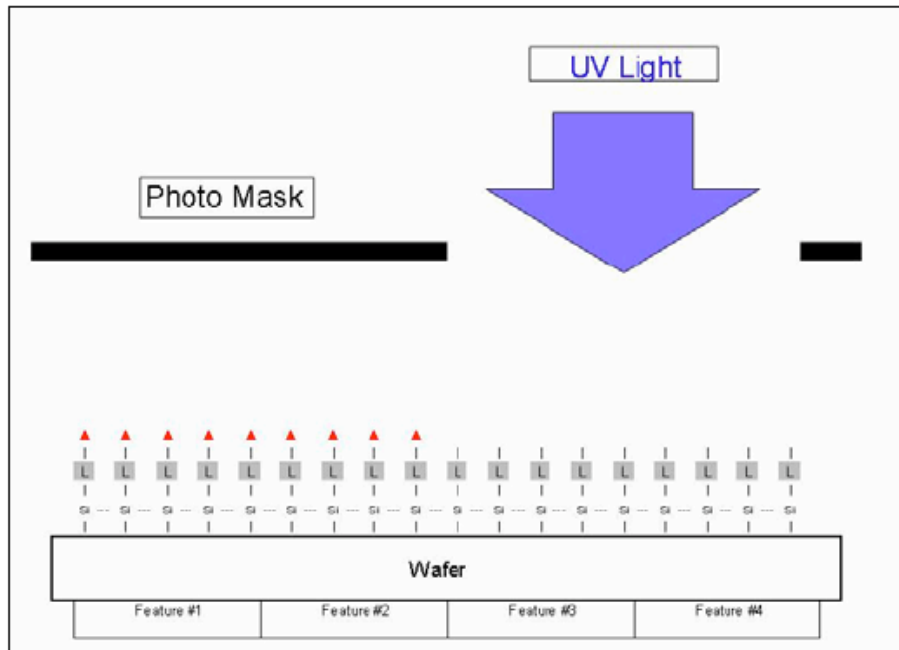


Fig. 5 - Deprotection of Feature #3 and 4

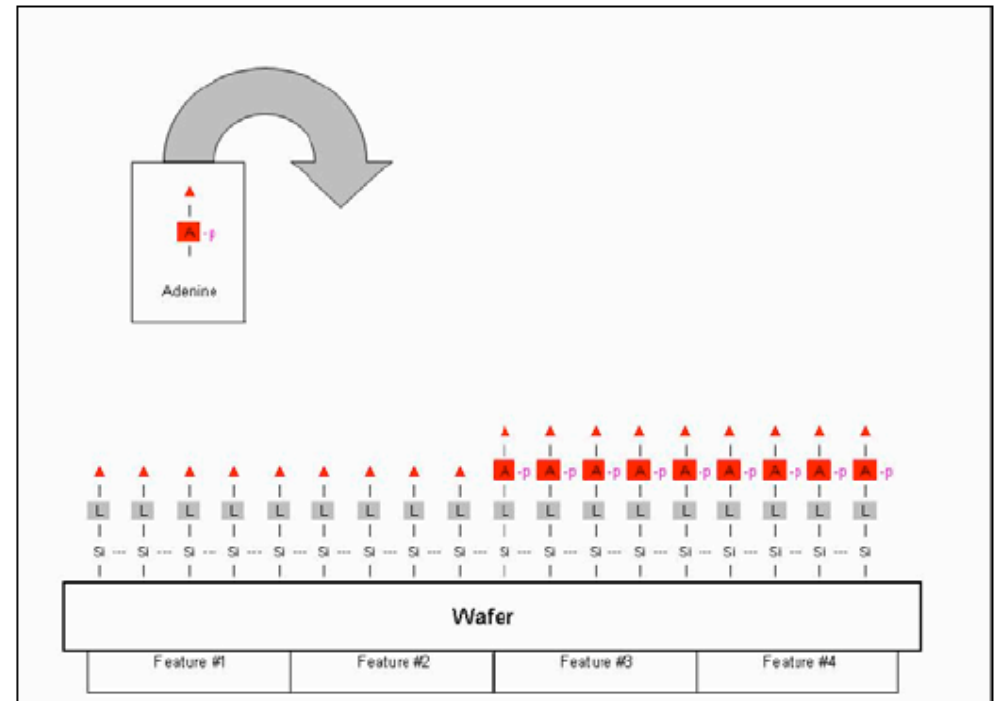


Fig.6 - Addition of Adenine (A) Nucleotides

Φωτολιθογραφία

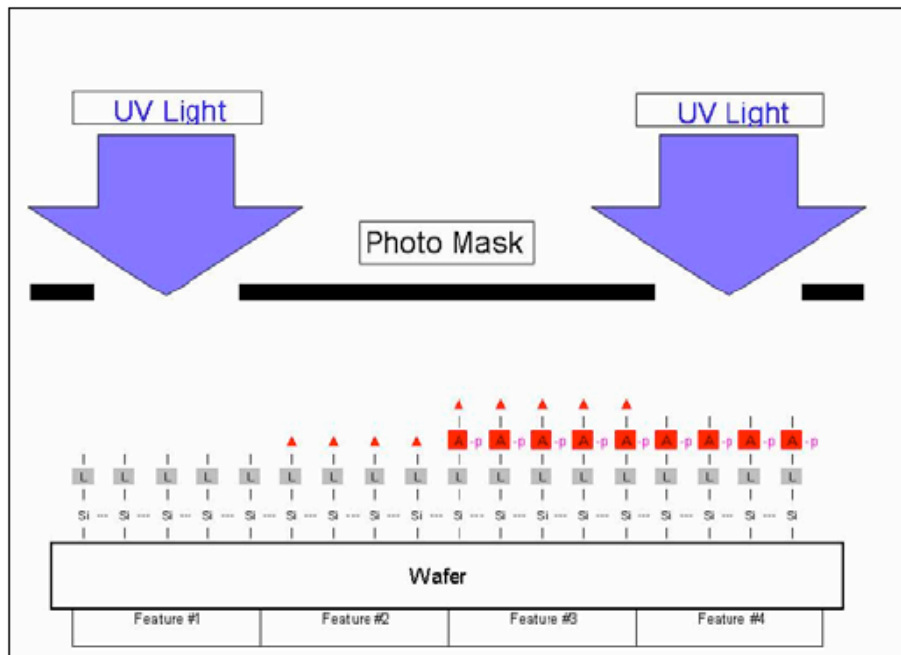


Fig.7 - 2nd mask (deprotecting Feature 1 and 4)

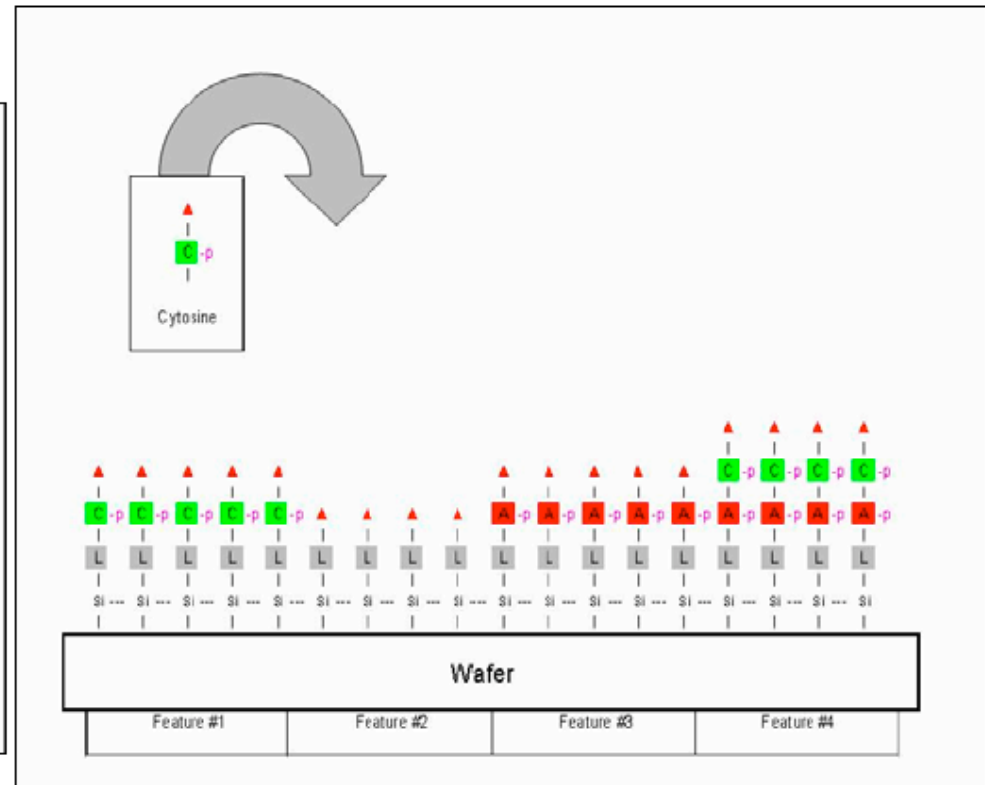


Fig. 8 - Addition of the 2nd nucleotide (C)

Φωτολιθογραφία

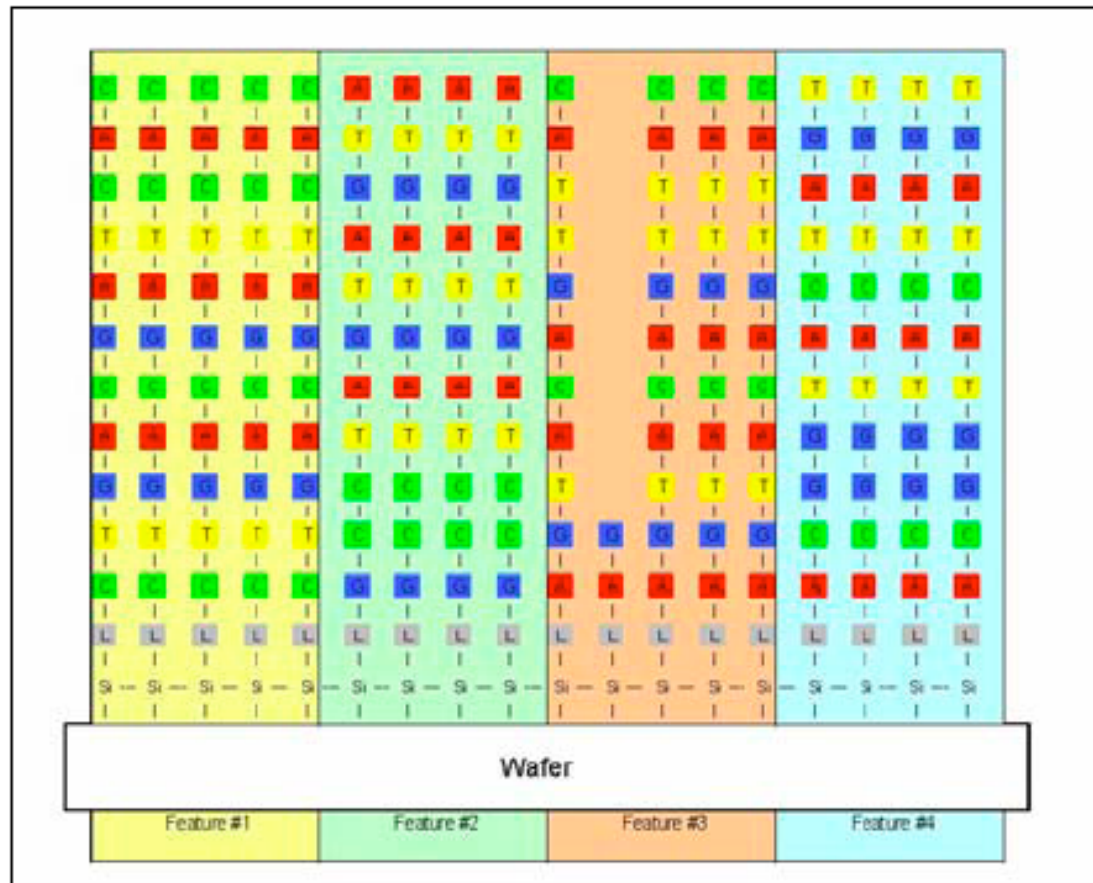


Fig. 13 - Final Results after removal of protection molecule and capping agent

Feature

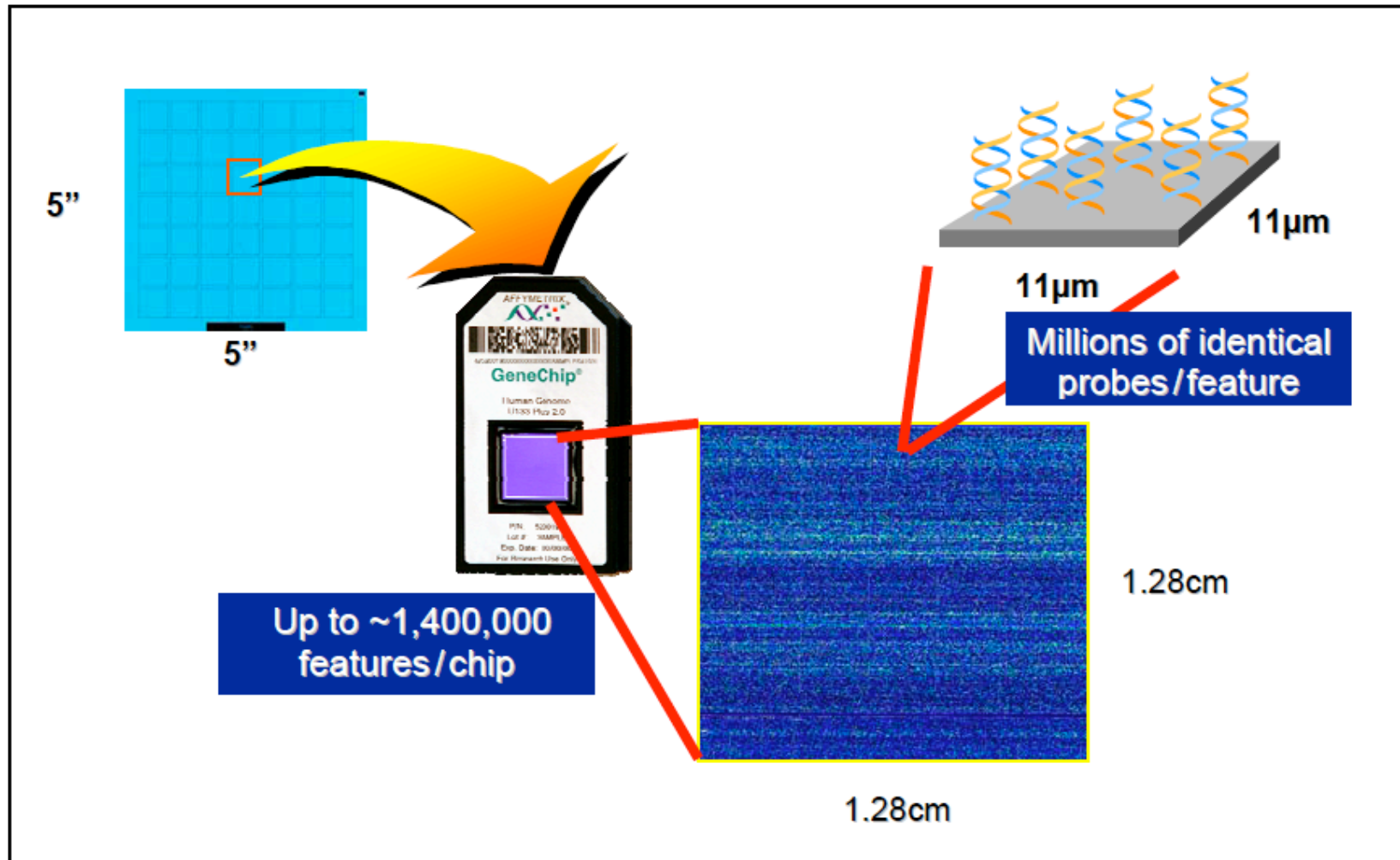
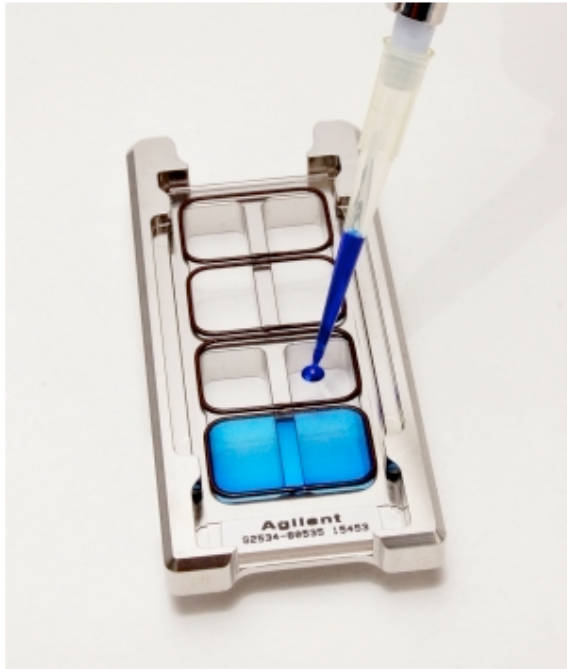
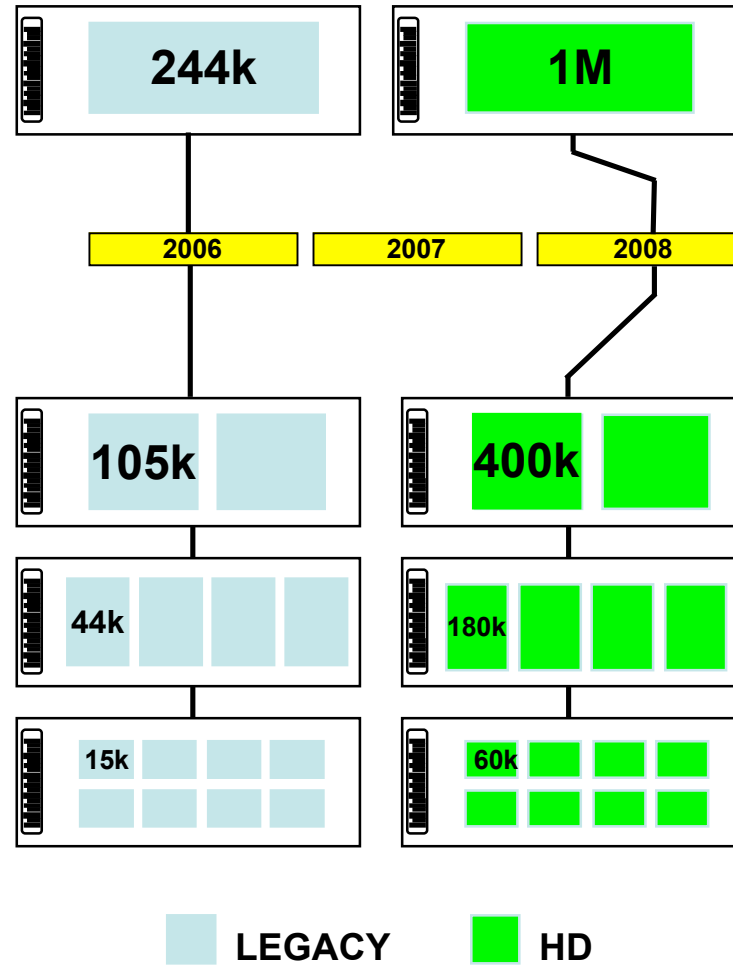


Fig. 17 - Dimension of a GeneChip: Wafer to Chip to a Single Feature

Agilent chips



Agilent chips



Agilent platform

- Φούρνος υβριδισμού



- Laser Scanner



Affymetrix Arrays/strips/plates



Figure 1: GeneChip® Human Genome U133 Arrays shown in cartridge and plate formats.



Affymetrix platform



Affymetrix platforms



GeneAtlas™ System & array strips



GeneTitan® System & array plates

Έλεγχος Ποσότητας/Ποιότητας ολικού RNA

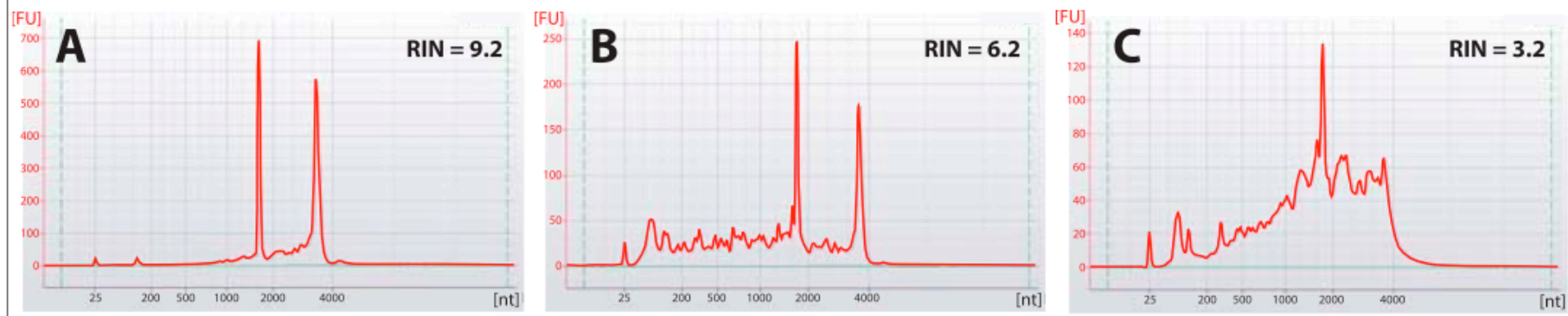
- Nanodrop (φωτόμετρο που χρησιμοποιεί ελάχιστο όγκο δείγματος - 1λ) για ακριβής ποσοτικοποίηση.
 - A260 nm -> νουκλεοτίδια, RNA, ssDNA, dsDNA
 - A280 nm -> πρωτεΐνες, φαινόλες κ.α.
 - A230 nm -> EDTA, φαινόλες, υδατάνθρακες.
 - A260/A280 nm ~ 1.8 - 2.1
 - 260/230 nm ~ 2 - 2.2



Έλεγχος Ποσότητας/Ποιότητας ολικού RNA

- Agilent 2100 Bioanalyzer για έλεγχο ποιότητας του RNA (τυχόν αποδόμιση). (~ €20.000)
- Ηλεκτροφόρηση σε τριχοειδή
 - Λόγος 28S/18S rRNA 2:1
 - RNA integrity number (RIN)

Figure 1.2 Example Agilent Bioanalyzer Electropherograms from three different total RNAs of varying integrity. Panel [A] represents a highly intact total RNA (RIN = 9.2), panel [B] represents a moderately intact total RNA (RIN = 6.2), and panel [C] represents a degraded total RNA sample (RIN = 3.2).



NOTE: Total RNAs with lower RIN values may require increased input amounts to generate enough aRNA for hybridization to an array.

Έλεγχος Ποσότητας/Ποιότητας ολικού RNA

1 Fast and easy operation Add sample



- Ready-to-use reagent kits
- Quick-start instructions
- Chip preparation in less than 5 minutes
- Minimal use of hazardous chemicals and waste disposal
- Sample volumes in the μL -range

2 Automation Start chip run



- Start analysis at the press of a button
- Predefined protocols
- System uses internal standards to calculate results

3 Digital data in 30 minutes Watch real-time data display



- Automated data analysis
- Digital data can be filed in a database or shared
- No user-dependent data interpretation

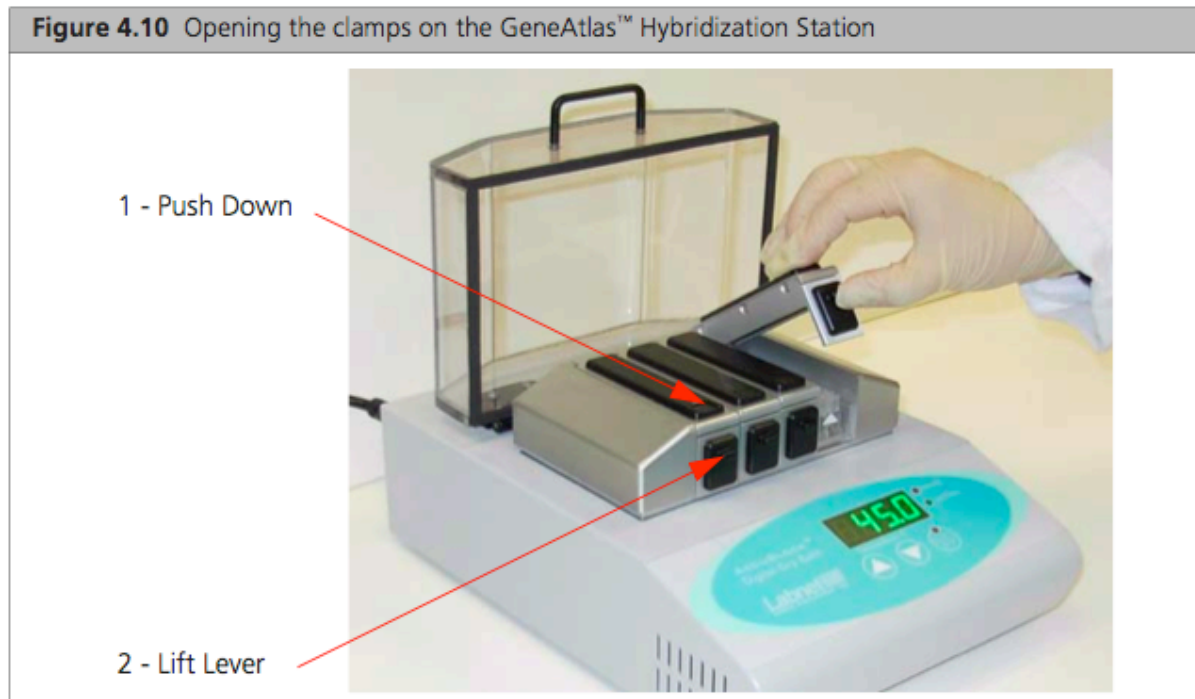
Χημικές αντιδράσεις

Υβριδισμός

Όλα τα probes υβριδίζονται στην ίδια θερμοκρασία.

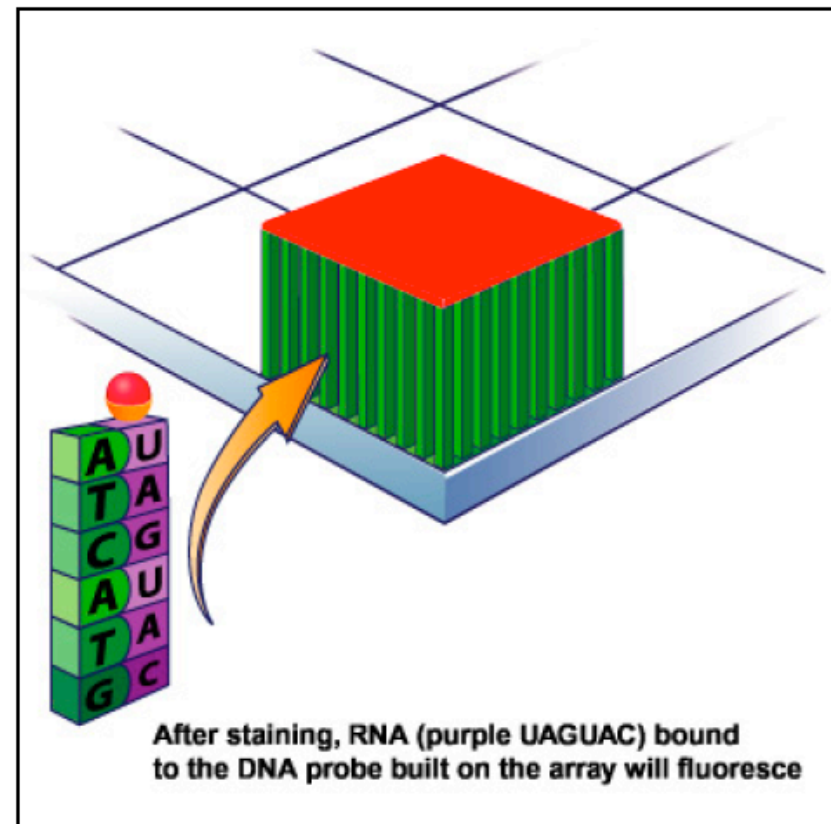
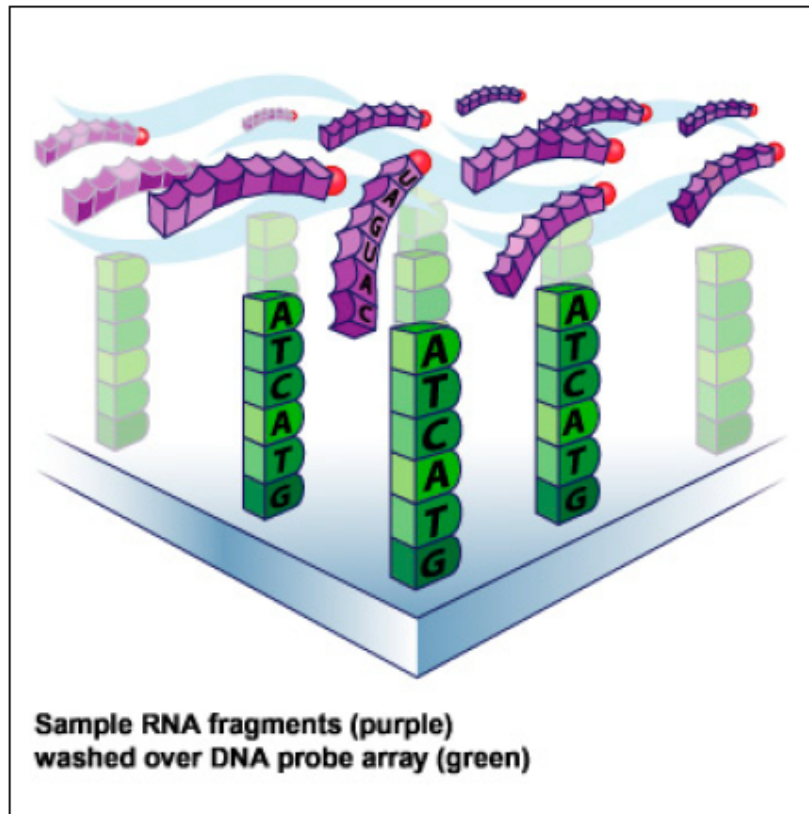
Όμως, όλα τα probes δεν έχουν την ίδια βέλτιστη θερμοκρασία υβριδισμού.

Άρα, υπάρχουν προβλήματα για επιμέρους probes



Φούρνος υβριδισμού του Affymetrix GeneAtlas

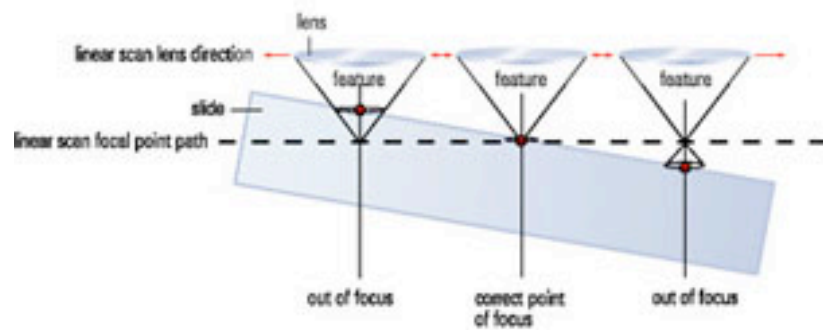
Υβριδισμός



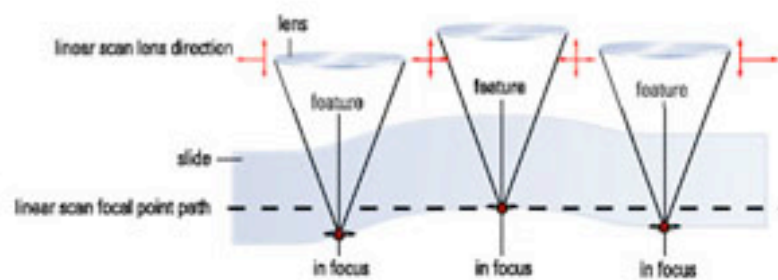
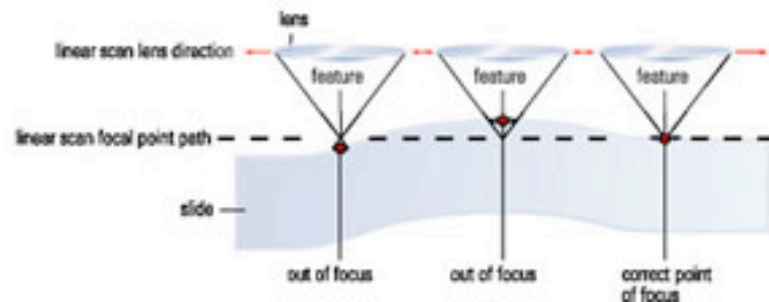
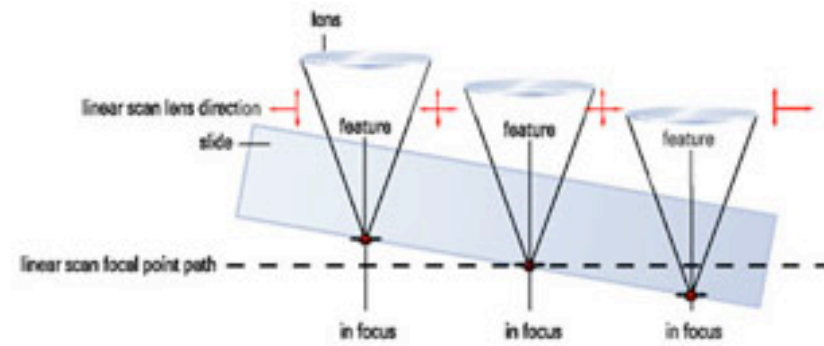
Laser scanning

Agilent autofocus

Single Point Focus
Out of focus, resulting in questionable data.



Agilent Dynamic Autofocus
Always in focus, for results you can trust.



Μέτρηση σήματος στα spots

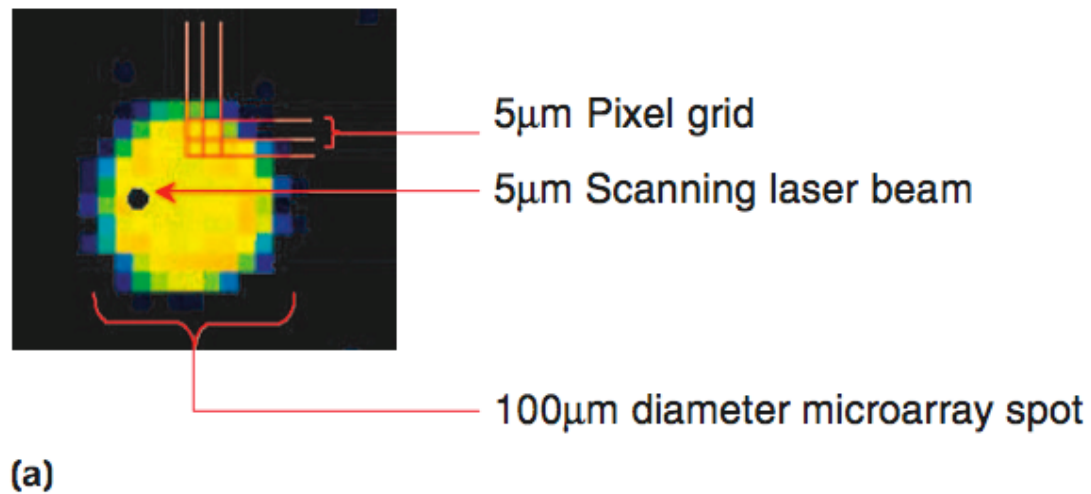


Figure 1.12: The pixels comprising a feature. (a) A false-colour image of the pixels from a single scan of a 100- μm microarray feature. The size of the laser spot is 5 μm . The pixel size has been set to 5 μm so that each pixel represents the area from the size of the laser spot. **(b)** and **(c)** See pp. 9 and 10.

Συνένωση των σημάτων

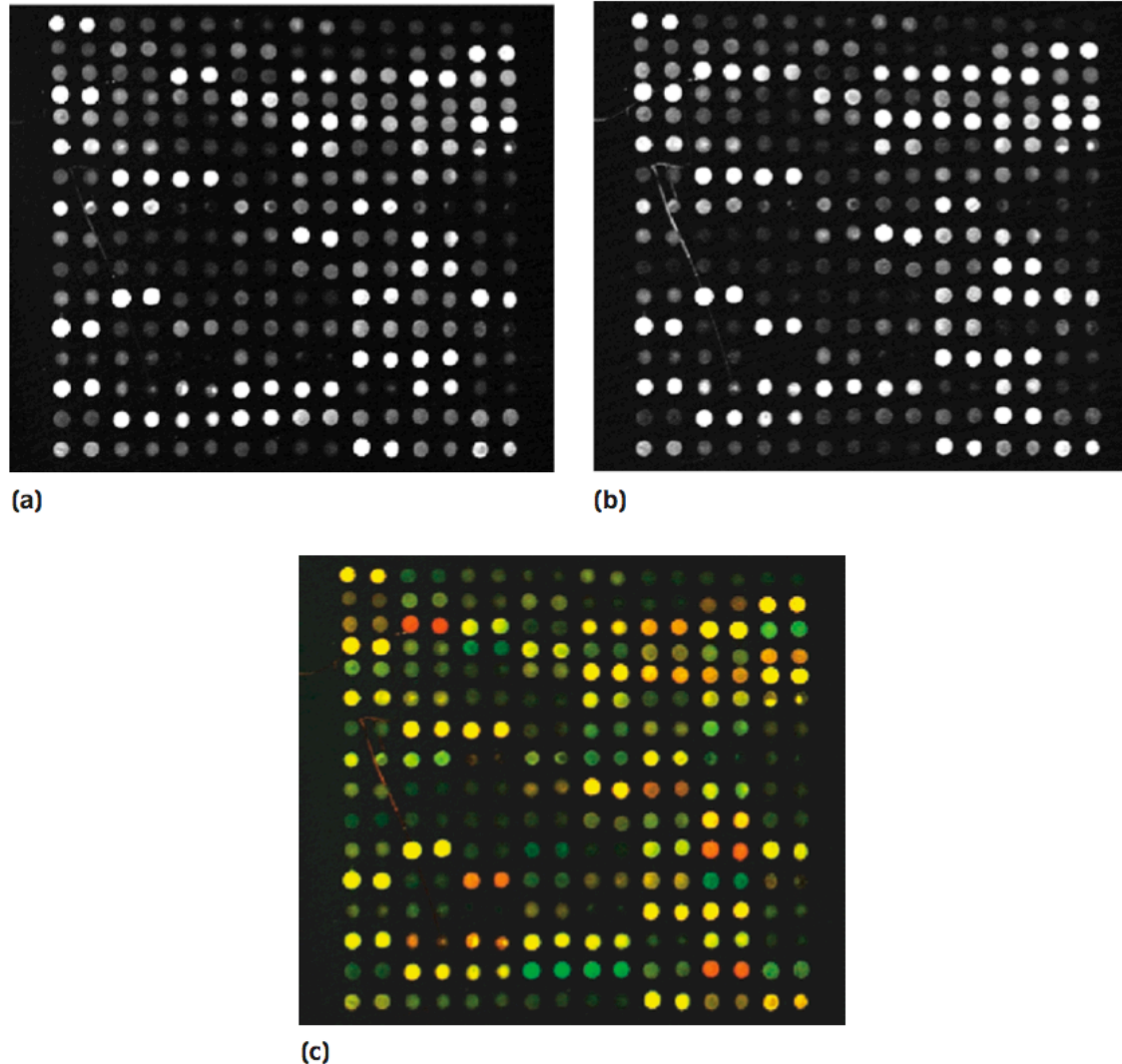


Figure 1.13: Output of scanners. (a) This is the scanner output for a part of a microarray – in this case one of twelve 16×16 blocks of features. This is the monochrome image of the Cy3 (green) channel. (b) The scanner output for the same part of the array but using the Cy5 (red) channel. (c) It is usual to combine the two monochrome images into a composite false-colour image of the array. Green features correspond to features that are expressed more in the sample labelled with Cy3 than the sample labelled with Cy5, and so will be bright in (a) and dark in (b). Similarly, red spots will be bright in (b) and dark in (a). Yellow features have a similar level of expression in both samples. Dark features are low expressed in both samples.

Ποιότητα των spots

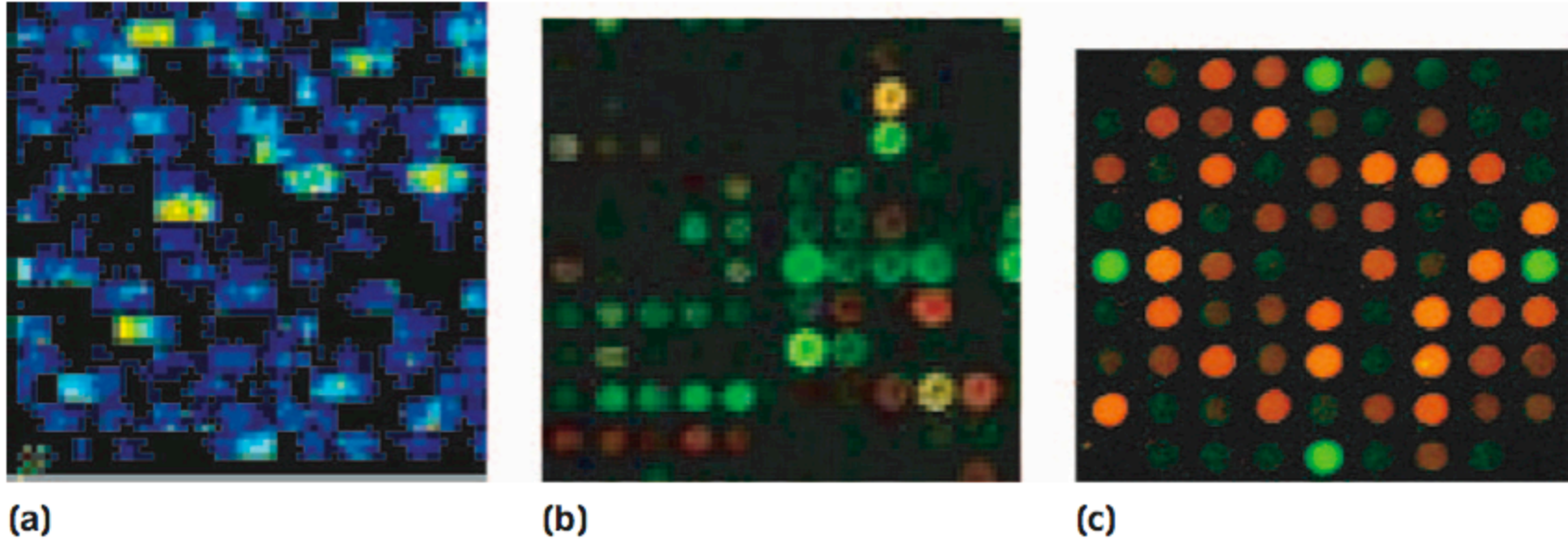


Figure 1.7: Array quality. (a) On Affymetrix arrays the features are rectangular regions. The masks refract light, so there is leakage of signal from one feature to the next. The Affymetrix image-processing software compensates for this by using only the interior portions of the features. (b) Spotted arrays produce spots of variable size and quality. This image shows some of this variation; we cover image processing of spotted arrays in detail in Chapter 4. (c) Inkjet arrays tend to be of the highest quality, with regular, even spots.

Ανάλυση προβληματικών Spots

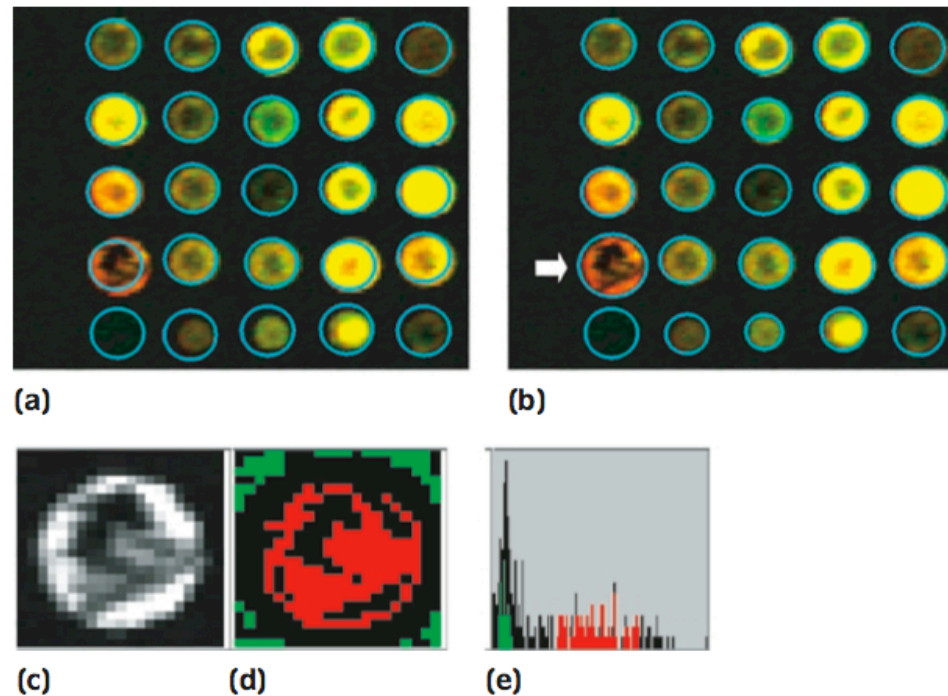


Figure 4.3: (a) Fixed circle segmentation. A circle of the same size is placed on every feature on the array and the pixels inside the circle are used to determine the intensity of the feature. This is not a good method because the circle will be too large for some features and too small for others. (b) Variable circle segmentation. A circle of different size is applied to each feature and the pixels inside the circle are used to determine the intensity of the feature. This performs better on different size features but does not perform so well on features with irregular shapes, for example, the irregular red feature that is marked with an arrow. (c) Zoom in on the red channel of the irregularly shaped feature marked with the arrow in (b). Note the black region where there is no hybridisation, probably because there is no probe attached to the glass in that area. (d) Histogram method applied to that feature. The red pixels are the ones that have been used to calculate the feature signal; the green pixels have been used to calculate the feature background. The black pixels are unused. The area corresponding to the black region in (c) is not used for calculating the feature intensity. The brightest features have also been excluded. The red-to-green ratio of this feature calculated by fixed circle segmentation is 1.8, variable circle segmentation is 1.9, and histogram segmentation is 2.6; so the measured differential gene expression between the samples is different with the different algorithms. Because of the irregular shape of the feature, the histogram method probably gives the most realistic measurement. (e) Histogram of the intensities of the pixels in the irregularly shaped feature. The red bars represent pixels used for the signal intensity; the green bars represent pixels used for the background intensity; the black bars are unused pixels. The brightest and darkest pixels are not used, thus giving a better measurement of hybridisation intensity.

Υπολογισμός & αφαίρεση θορύβου υποβάθρου

IMAGE PROCESSING

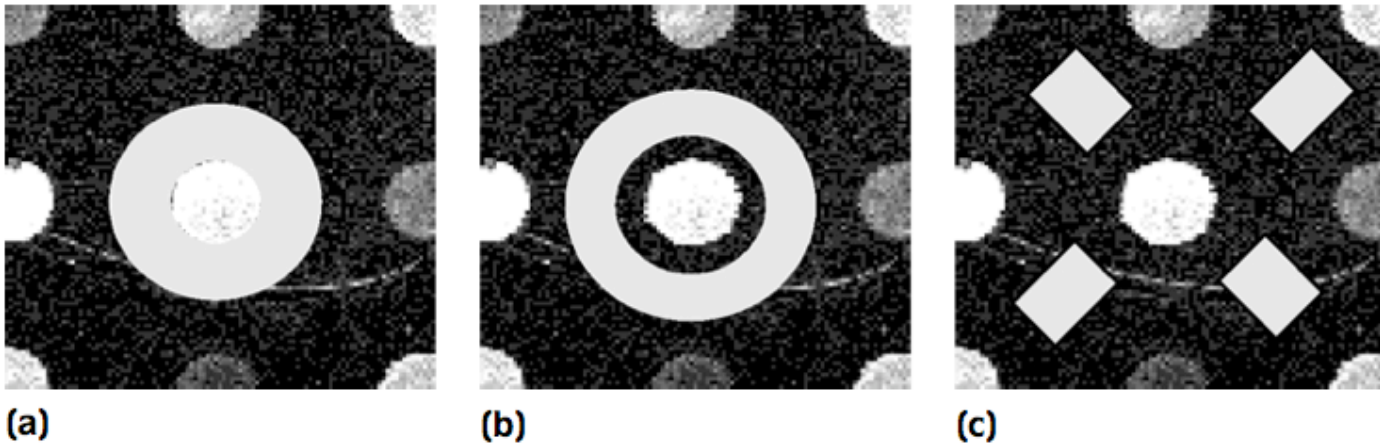


Figure 4.4: Background regions used by different software. Different software packages use different pixel regions surrounding the feature to determine the background intensity. **(a)** ScanAlyze: the region is adjacent to the feature. This will be inaccurate if the feature is larger than the fixed size of the circle used for segmentation. **(b)** ImaGene: there is a space between the feature and the background. This is a better method than (a). **(c)** Spot and GenePix: the background region is in between the features. This is also a good method.

Παράγοντες που επηρεάζουν τα microarrays

- Περιβαλλοντικοί
 - Όζον
 - Υγρασία
 - Υπεριώδης ακτινοβολία
 - Στατικός ηλεκτρισμός
 - Μόλυνση της ατμόσφαιρας (οξειδωτικοί παράγοντες).
 - Σκόνη (μπορεί να επηρεάσει τις μετρήσεις του Laser)
 - Πούδρα από τα γάντια.
- Ενδογενείς
 - Πρόβλημα σε κάποιο βήμα των χημικών αντιδράσεων.
 - Ο τύπος του ιστού.
- Dye swar για να αποκλείσουμε παράγοντες που πιθανών επηρέασαν το πείραμα

Microarrays & Ozone

- Η χρωστική Cy5(red) είναι συνήθως πιο ασταθής σε σχέση με τη χρωστική Cy3(green).
- Το όζον μπορεί να προκαλέσει περισσότερη αστάθεια της Cy5.
- Σε ένα πείραμα που το control χρωματίστηκε με Cy5 και το δείγμα με Cy3, αν υποθέσουμε ότι το όζον προκάλεσε την αποδόμιση του Cy5, τότε περιμένουμε να δούμε (από λάθος) τα γονίδια του δείγματος να:
 - A) υπερεκφράζονται σε σχέση με το control?
 - B) υποεκφράζονται σε σχέση με το control?

Microarrays & ozone

[Anal Chem.](#) 2003 Sep 1;75(17):4672-5.

Effects of atmospheric ozone on microarray data quality.

[Fare TL](#), [Coffey EM](#), [Dai H](#), [He YD](#), [Kessler DA](#), [Killian KA](#), [Koch JE](#), [LeProust E](#), [Marton MJ](#), [Meyer MR](#), [Stoughton RB](#), [Tokiwa GY](#), [Wang Y](#).

Rosetta Inpharmatics LLC, 12040 115th Avenue NE, Kirkland, Washington 98034, USA.

Abstract

A data anomaly was observed that affected the uniformity and reproducibility of fluorescent signal across DNA microarrays. Results from experimental sets designed to identify potential causes (from microarray production to array scanning) indicated that the anomaly was linked to a batch process; further work allowed us to localize the effect to the posthybridization array stringency washes. Ozone levels were monitored and highly correlated with the batch effect. Controlled exposures of microarrays to ozone confirmed this factor as the root cause, and we present data that show susceptibility of a class of cyanine dyes (e.g., Cy5, Alexa 647) to ozone levels as low as 5-10 ppb for periods as short as 10-30 s. Other cyanine dyes (e.g., Cy3, Alexa 555) were not significantly affected until higher ozone levels (> 100 ppb). To address this environmental effect, laboratory ozone levels should be kept below 2 ppb (e.g., with filters in HVAC) to achieve high quality microarray data.

PMID: 14632079 [PubMed - indexed for MEDLINE]

Microarrays & ozone

From: [BMC Biotechnol. 2007; 7: 8.](#)

Published online 2007 February 12. doi: 10.1186/1472-6750-7-8.

[Copyright/License](#)

[Permissions](#)

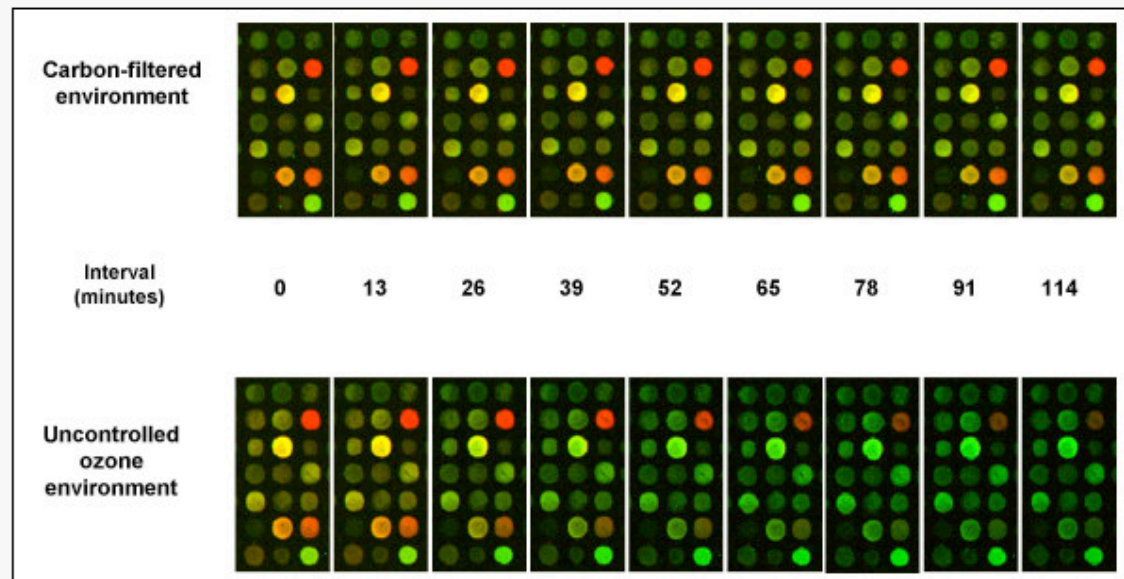


Figure 2

Reduction of Cy5 signal for in-house fabricated microarrays kept in a none-ozone controlled environment. These figures show a region of two 20 K mouse microarrays, selected because of their wide range of both Cy5 and Cy3 signals. The microarray images were from scans made during a 114 minute experiment interval. While the microarrays maintained in the carbon-filtered laboratory environment (ozone ~2–4 ppb) remained relatively unchanged with time (top images), reduction in Cy5 in the uncontrolled ozone environment (ozone ~25 ppb) causes the microarray image to have a predominantly green cast (bottom images).

Ανάλυση δεδομένων

Μετατροπή έντασης σήματος σε \log_2

SECTION 5.2 DATA CLEANING AND TRANSFORMATION

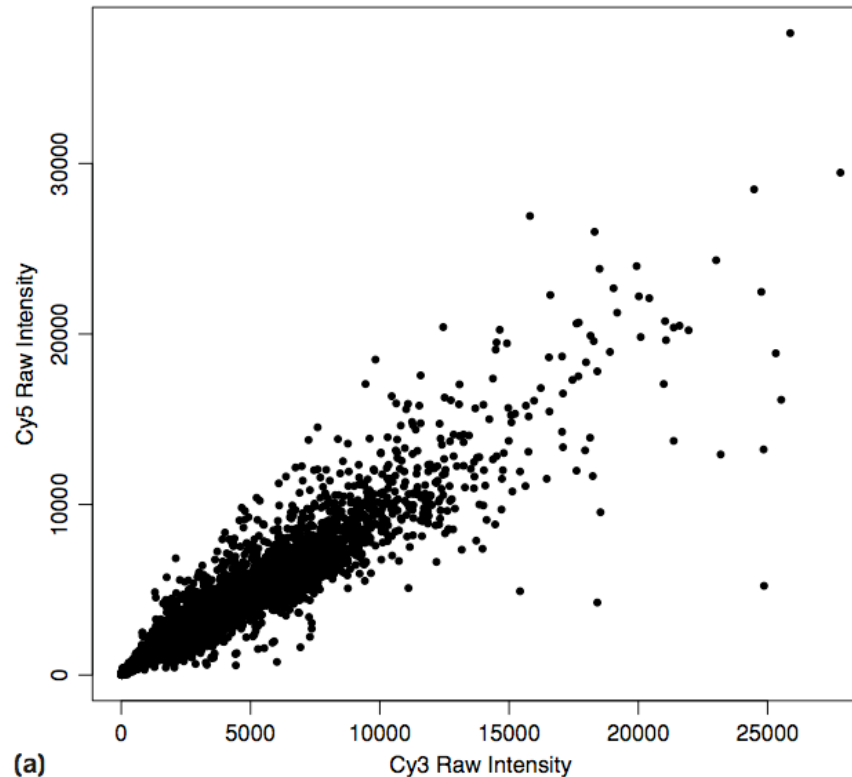
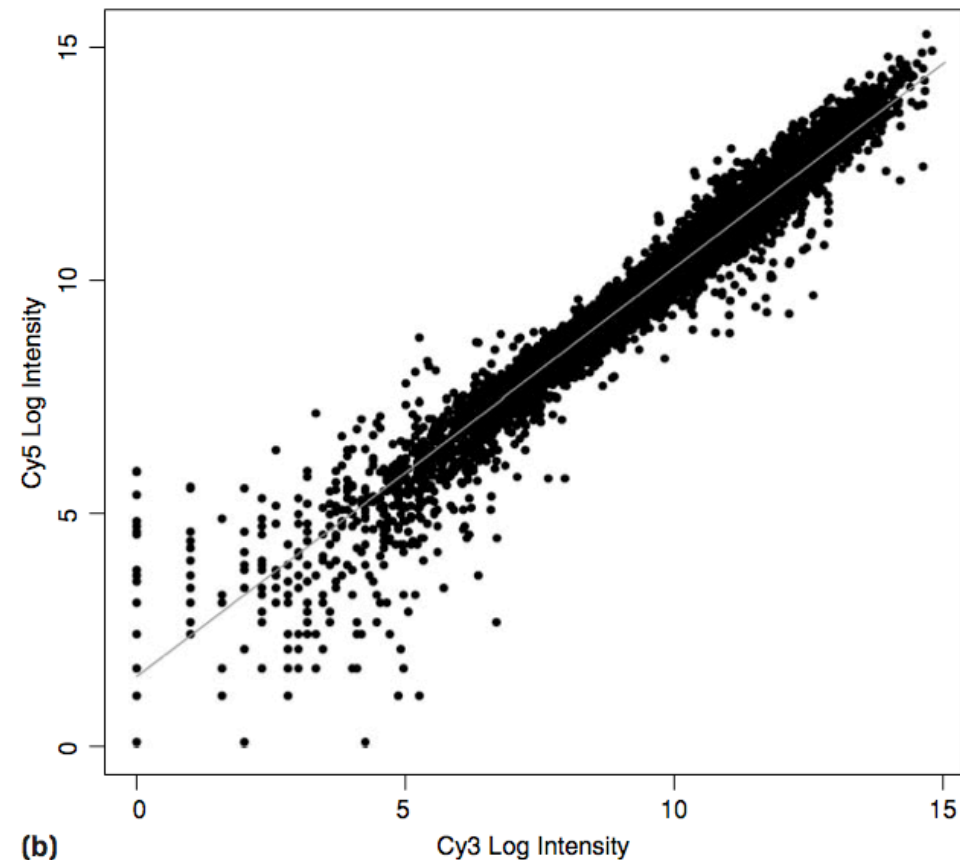


Figure 5.1: Plots of Cy3 vs. Cy5 for data set 5A. Human foreskin fibroblasts have been infected with *Toxoplasma gondii* for a period of 1 hour. A sample has been prepared, labelled with Cy5 (red), and hybridised to a microarray with approximately 23,000 features. The Cy3 (green) channel is a sample prepared from uninfected fibroblasts. Because the infectious period is short, most genes in this experiment are not differentially expressed. **(a)** Scatterplot of the (background-subtracted) raw intensities; each point on the graph represents a feature on the array, with the x coordinate representing the Cy3 intensity, and the y coordinate representing the Cy5 intensity. The graph shows two weaknesses of the raw data that would have a negative impact on further data analysis:

1. Most of the data is bunched in the bottom-left-hand corner, with very little data in the majority of the plot.
2. The variability of the data increases with intensity, so that it is very small when the intensity is small and very large when the intensity is large.

Μετατροπή έντασης σήματος σε \log_2

◇	A	B
1	Cy5 raw intensity	Cy5 log2 transformed intensity
2	1	0
3	2	1
4	4	2
5	8	3
6	16	4
7	32	5
8	64	6
9	128	7
10	256	8
11	512	9
12	1024	10
13	2048	11
14	4096	12
15	8192	13
16	16384	14
17	32768	15



(b) Figure 5.1: (continued)

(b) Scatterplot of the log (to base 2) intensities. This plot is better than (a). The data is spread evenly across the intensity range, and the variability of the data is the same at most intensities. The genes with log intensity less than 5 have slightly higher variability, but these genes are very low expressed and are below the detection level of microarray technology.

The straight line is a linear regression through the data. The linear regression is not perfect (the data appears to bend upwards away from the line at high intensities), but is approximately right. The intercept is 1.4, and the gradient is 0.88. If the two channels were behaving identically, the intercept would be 0 and the gradient would be 1. We conclude that the two Cy dyes behave differently at different intensities; this could result from differential dye incorporation or different responses of the dyes to the lasers.

Μετατροπή έντασης σήματος σε \log_2

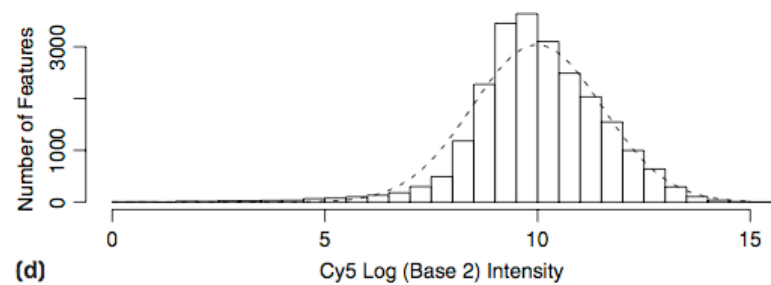
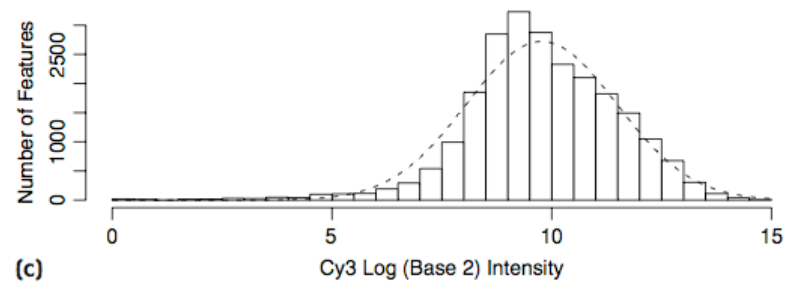
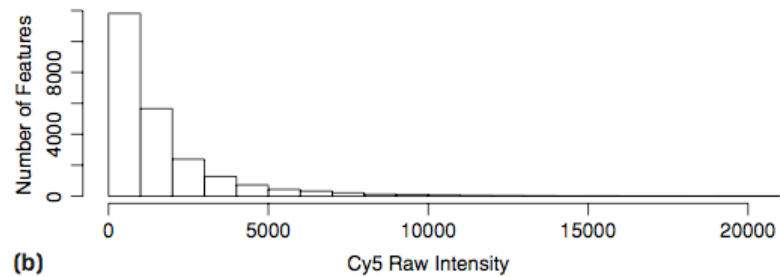
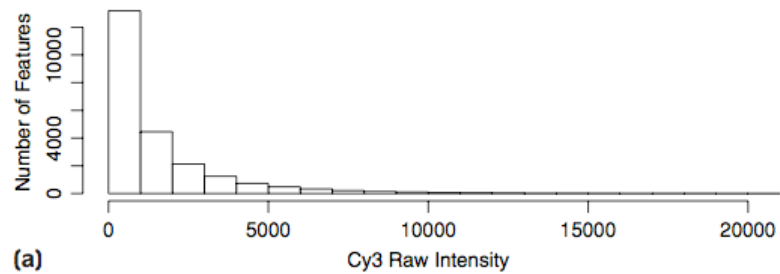
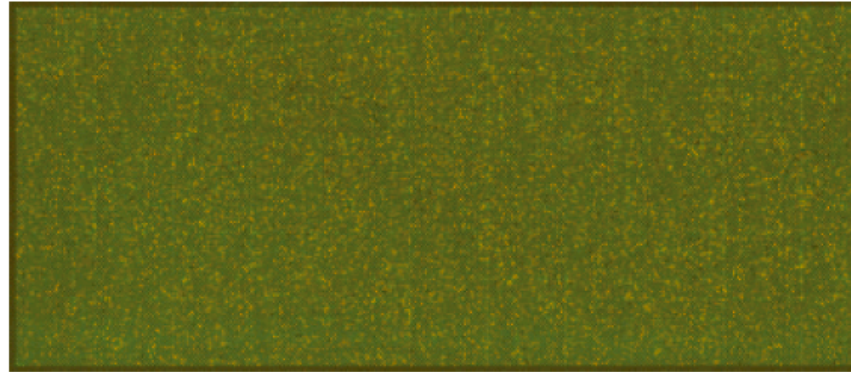
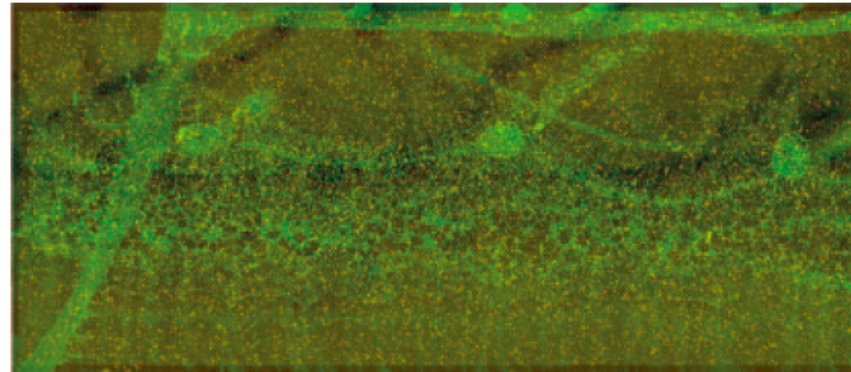


Figure 5.2: Histograms of the raw and log Cy3 and Cy5 intensities. Histograms of the intensities of the features for the human fibroblast data. **(a)** The raw intensities for the Cy3 channel; the data is right-skewed, with the majority of features having low intensity and decreasing numbers of features having higher intensity. **(b)** The raw intensities for the Cy5 channel; the pattern is the same as (a). **(c)** The log intensities for the Cy3 channel; the intensities are closer to a bell-shaped normal curve (shown as a dashed line). There is still a slight right skew, but the logged data is better for data analysis than the raw data. **(d)** The log intensities for the Cy5 channel, along with a normal curve (dashed line). As with (c), the intensities are approximately normal, with a slight right skew.

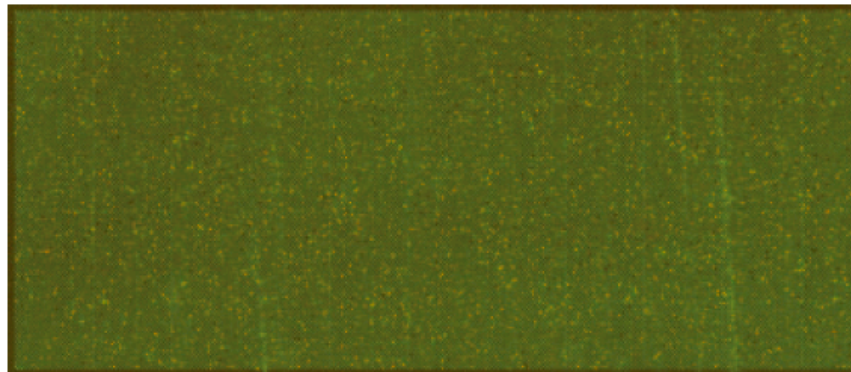
Microarray Quality control



Control, normal microarray



Microarray with wash artifacts



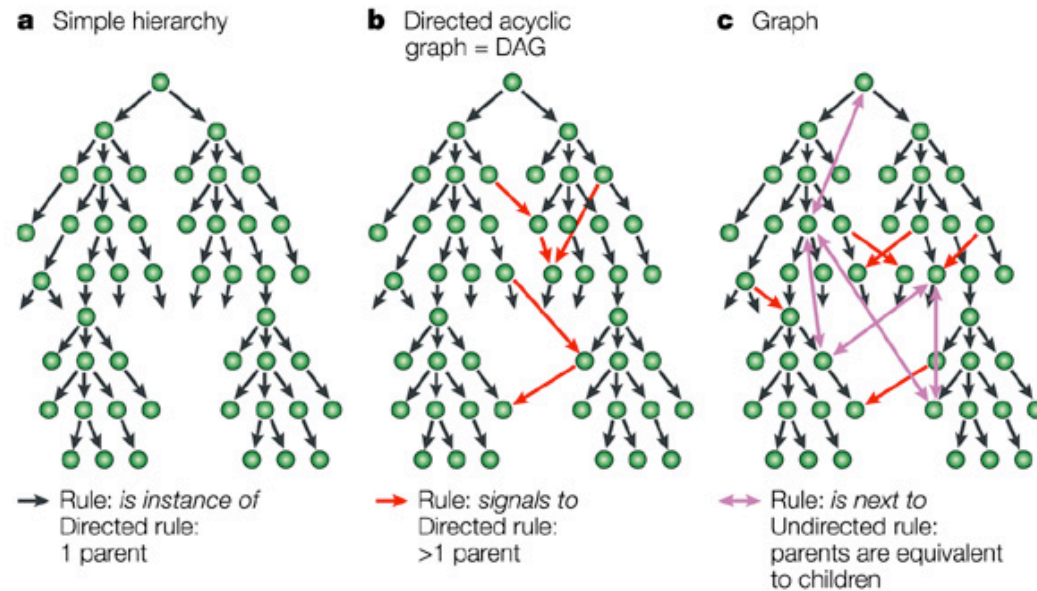
Microarray with degraded total RNA

Οντολογίες

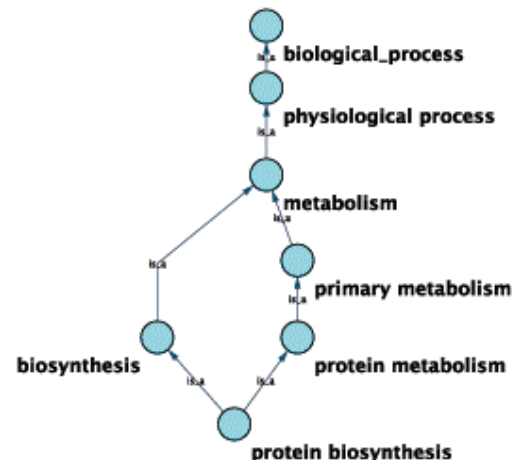
- www.geneontology.org
- Ελεγχόμενο λεξιλόγιο για την περιγραφή των ιδιοτήτων των γονιδίων και των πρωτεϊνών.
- Περιγράφουν:
 - Μοριακές λειτουργίες του βιομορίου (1 ή περισσότερες).
 - Βιολογικές διαδικασίες στις οποίες εμπλέκεται το βιομόριο (1 ή περισσότερες).
 - Κυτταρικό διαμέρισμα στο οποίο συναντάται το βιομόριο (1 ή περισσότερα).

Οντολογίες: Η δομή τους

- Δείχνει τις σχέσεις μεταξύ των διαφορετικών όρων.
- Ένας όρος μπορεί να αποτελεί πιο εξειδικευμένη περιγραφή ενός άλλου όρου.
- Είναι κατευθυνόμενα ακυκλικά γραφήματα (DAG).
- Παρόμοια με ιεραρχίες.
- Η διαφορά είναι ότι ένας κόμβος-απόγονος μπορεί να έχει περισσότερους από έναν προγόνους.

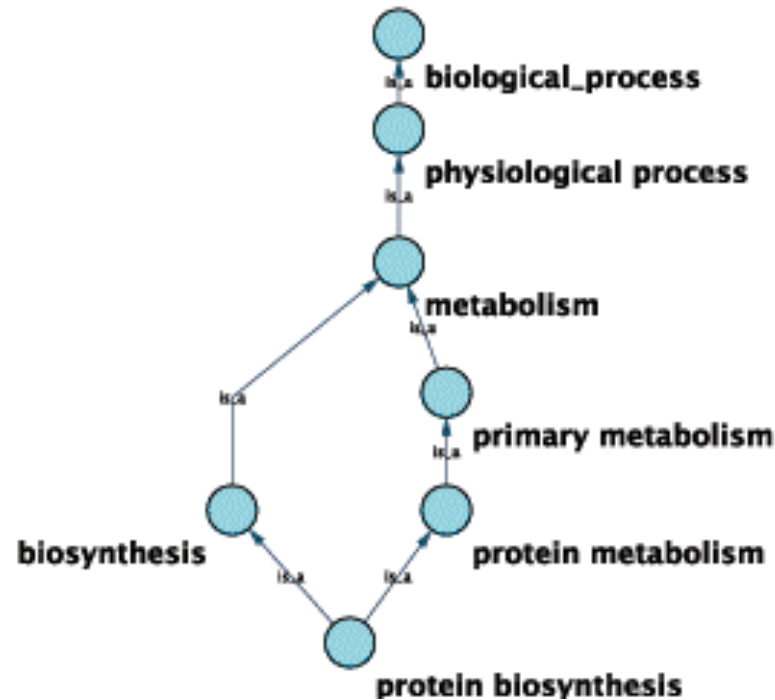


Nature Reviews | **Genetics**



Οντολογίες: Η δομή τους

- Θεωρούμε ότι αν σε ένα βιομόριο αντιστοιχεί ένα όρος-οντολογία, τότε σε αυτό το βιομόριο ανήκουν και όλοι οι πρόγονοι του όρου-οντολογίας.



Οντολογίες: στατιστική ανάλυση

- Παράδειγμα:
 - 1 γονιδίωμα με 10.000 γονίδια.
 - 1.000 γονίδια εμπλέκονται στον κυτταρικό κύκλο (GO_term: cell-cycle). (10% του γονιδιώματος).
 - Αν επιλέξουμε τυχαία έναν αριθμό X γονιδίων, θα περιμέναμε (από τύχη) περίπου το 10% (με κάποιες διακυμάνσεις) να έχουν τον όρο “κυτταρικός κύκλος”.
 - Η τυχαία διακύμανση εξαρτάται από τον αριθμό των γονιδίων.
 - Έστω ότι με τα microarrays σε ένα πείραμα βρήκαμε ότι X αριθμός γονιδίων υπερεκφράζονται.
 - Σε αυτό τον X αριθμό, βρήκαμε ότι 20% των γονιδίων ανήκουν στον κυτταρικό κύκλο.
 - Αυτή η απόκλιση (20% παρατηρούμενο - 10% αναμενόμενο) είναι στα όρια των τυχαίων διακυμάνσεων, ή είναι στατιστικά σημαντική?
 - Στατιστικά σημαντική, σημαίνει ότι τα υπερεκφρασμένα γονίδια είναι εμπλουτισμένα για την κατηγορία “κυτταρικός κύκλος”. Δηλαδή, ο κυτταρικός κύκλος εμπλέκεται στην διαδικασία που μελετάμε.

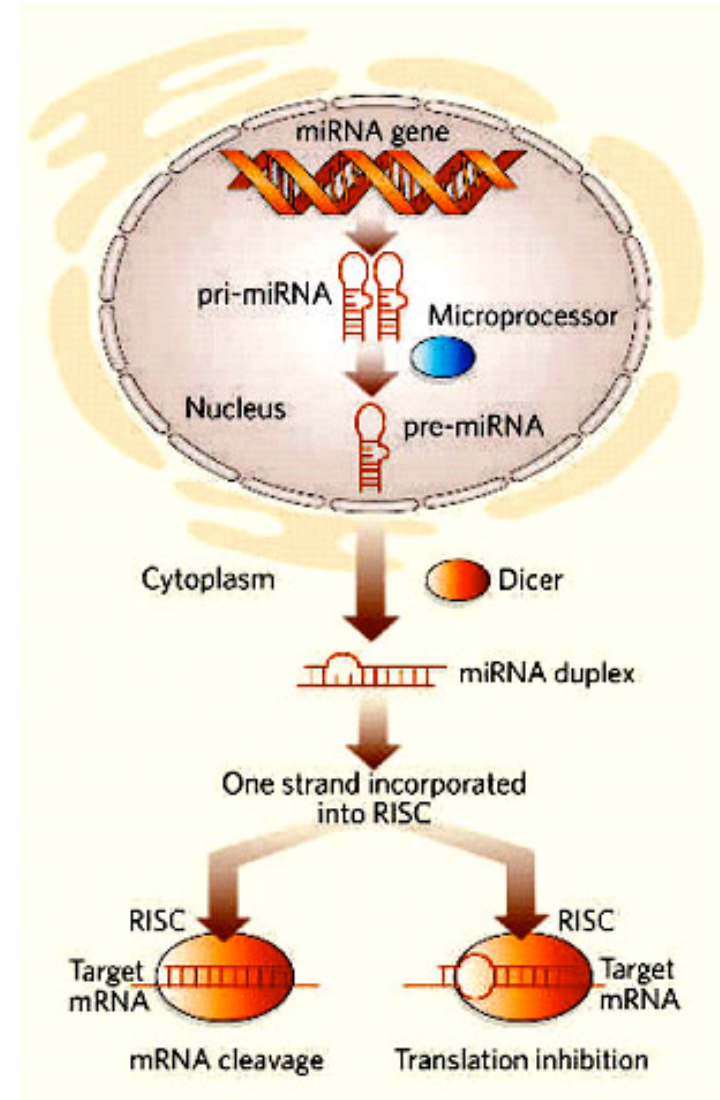
Οντολογίες: στατιστική ανάλυση

- Η στατιστική ανάλυση γίνεται με το υπεργεωμετρικό τεστ.
- Παίρνουμε ένα p-value.
- Αν $p\text{-value} < 0.05$, τότε είναι στατιστικά σημαντικό.

- Αν στις οντολογίες μας είχαμε 100 όρους, θα επαναλαμβάναμε τα παραπάνω τεστ για τον κάθε όρο.
- Όμως, όσο περισσότερα τεστ κάνουμε για το πείραμά μας, τόσο αυξάνει ή πιθανότητα να βρούμε κάτι στατιστικά σημαντικό ($p\text{-value} < 0.05$) καθαρά από λάθος.
- Άρα, πρέπει να λάβουμε υπόψην μας πόσα τεστ διενεργούμε και να διορθώσουμε τα p-values (multiple testing correction).
 - False discovery rate (Benjamini-Hochberger)
 - Bonferroni correction

Micro-RNAs


- Μικρά RNAs που ρυθμίζουν (κυρίως καταστέλουν) την μεταγραφή των mRNAs
 - Αποδόμιση του mRNA.
 - Αναστολή της μεταγραφής του mRNA.
- Μέγεθος ~ 22 nt.
- ~1000 στον άνθρωπο.
- Ρυθμίζουν πολλά mRNAs (30%-100%).
- Ένα mi-RNA -> στοχεύει πολλά mRNAs.
- Πολλά mi-RNAs -> στοχεύουν το ίδιο mRNA (hunting pack).
- Έχουν εξειδικευμένο προφίλ έκφρασης για τον κάθε ιστό (και καρκίνο) (συνήθως σχετίζονται με συγκεκριμένες αναπτυξιακές διαδικασίες).
- Μπορούν να λειτουργήσουν ως διαγνωστικοί ή προγνωστικοί μοριακοί δείκτες.



Toxicology Applications

Circulating microRNAs, potential biomarkers for drug-induced liver injury

Kai Wang,¹ Shile Zhang, Bruz Marzolf, Pamela Troisch, Amy Brightman, Zhiyuan Hu, Leroy E. Hood,¹ and David J. Galas

 Author Affiliations

Contributed by Leroy E. Hood, January 16, 2009 (received for review December 9, 2008)

Abstract

Drug-induced liver injury is a frequent side effect of many drugs, constitutes a significant threat to patient health and has an enormous economic impact on health care expenditures. Numerous efforts have been made to identify reliable and predictive markers to detect the early signs of drug-induced injury to the liver, one of the most vulnerable organs in the body. These studies have, however, not delivered any more informative candidates than the serum aminotransferase markers that have been available for ≈ 30 years. Using acetaminophen overdose-induced liver injury in the mouse as a model system, we have observed highly significant differences in the spectrum and levels of microRNAs in both liver tissues and in plasma between control and overdosed animals. Based on our survey of microRNA expression among normal tissues, some of the microRNAs, like messenger RNAs, display restricted tissue distributions. A number of elevated circulating microRNAs in plasma collected from acetaminophen-overdosed animals are highly expressed in the liver. We have demonstrated that specific microRNA species, such as mir-122 and mir-192, both are enriched in the liver tissue and exhibit dose- and exposure duration-dependent changes in the plasma that parallel serum aminotransferase levels and the histopathology of liver degeneration, but their changes can be detected significantly earlier. These findings suggest the potential of using specific circulating microRNAs as sensitive and informative biomarkers for drug-induced liver injury.

Τοξική δράση acetaminophen

- Έλεγχος με microarrays αν τα mi-RNA μπορούν να καταδείξουν την ηπατική καταστροφή μετά από χορήγηση φαρμάκου (acetaminophen overdose).
- Σε κατάσταση υπερχορήγησης, τα μεταβολικά μονοπάτια που καταβολίζουν το φάρμακο είναι σε κορεσμό.
- Το πλεονάζον φάρμακο υπόκειται σε καταβολισμό από ένζυμα των P450 -> NAPQI (N-acetyl-p-benzoquinone imine) (τοξικό).
- NAPQI εξουδετερώνεται από ενδοκυτταρική γλουταθιόνη (GSH).
- Ότι NAPQI περισσεύει, προκαλεί ηπατική καταστροφή.
- Σε περίπτωση υπερβολικής δόσης acetaminophen (-> υπερβολικό NAPQI), χορηγούμε N-acetylcysteine (NAC) -> ενδοκυτταρικό GSH.

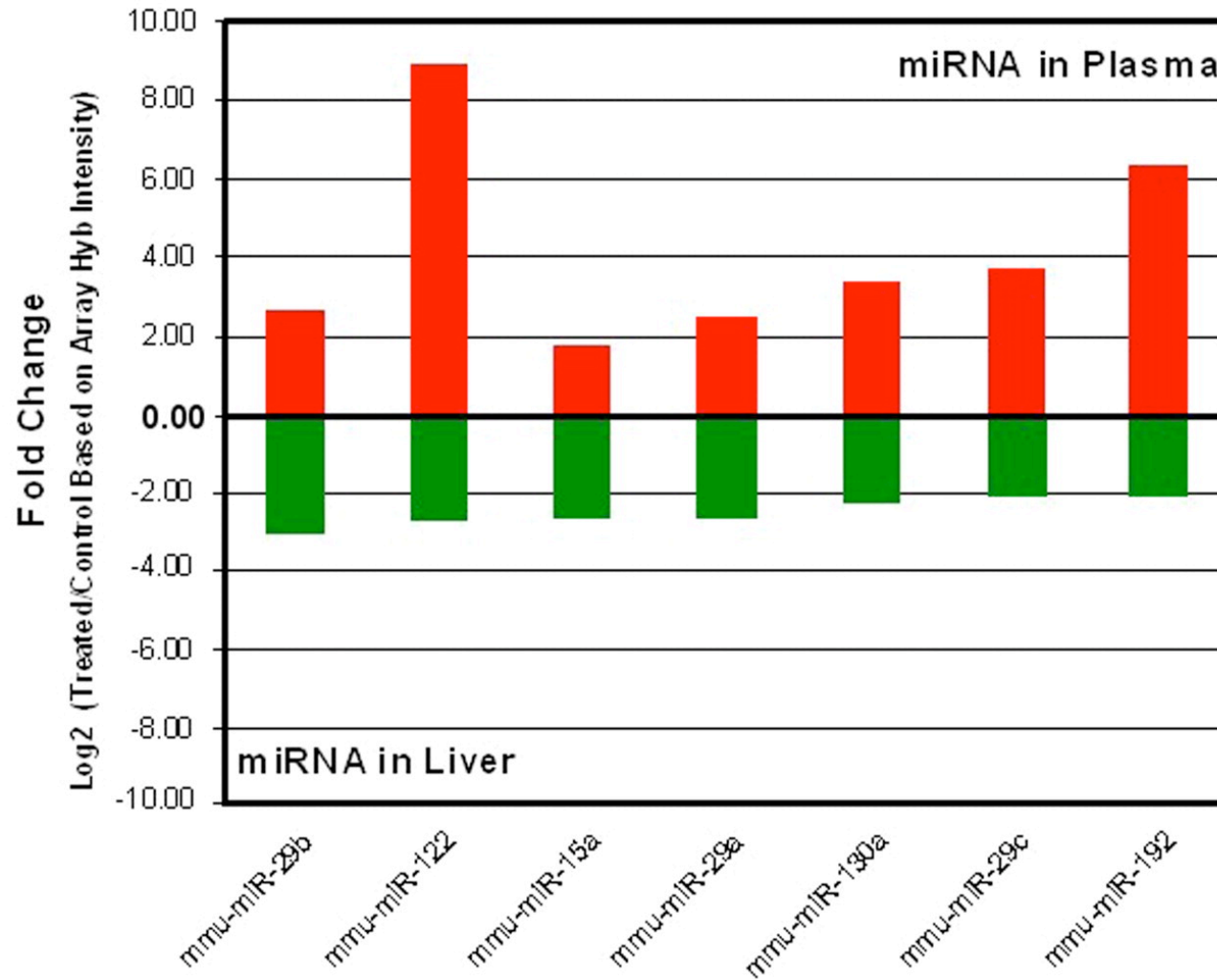
Μέτρηση της τοξικής δράσης του acetaminophen

- Η τοξικότητα δεν εμφανίζεται αμέσως.
- Χρειάζεται γρήγορη και αξιόπιστη μέτρηση της τοξικότητας με αιματολογικό έλεγχο (ορό).
- Μέτρηση της αμινοτρανσφεράσης του ορού (κλασσική μέθοδος).

- Βρήκαν mi-RNAs που εκφράζονται σημαντικά στο ήπαρ.
- Παρακολούθησαν την έκφρασή τους στο ήπαρ μετά τη χορήγηση acetaminophen και επίσης παρακολούθησαν την ανίχνευσή τους στον ορό (λόγω πιθανής ηπατικής κυτταρικής καταστροφής).

Toxicology Applications

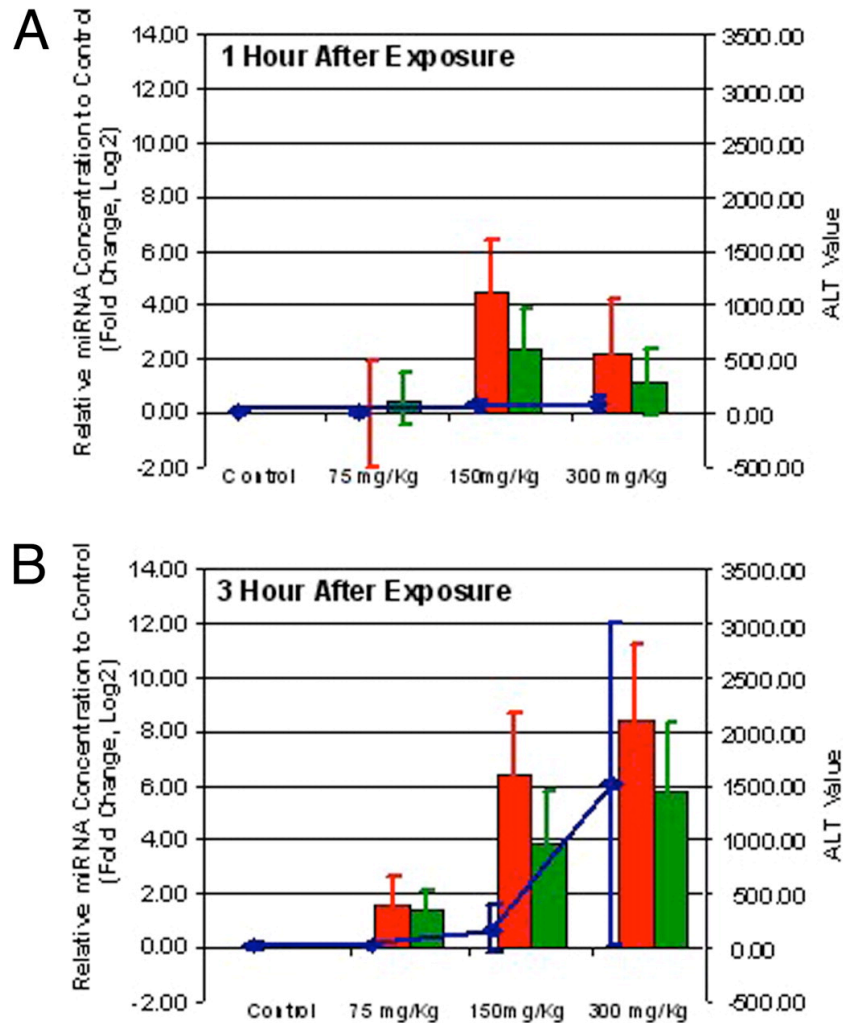
Selected miRNA shows opposite changes between liver and plasma samples based on microarray results.



Wang K et al. PNAS 2009;106:4402-4407

PNAS

Τα mi-RNAs αποδείχθηκαν πιο αξιόπιστοι και πιο γρήγοροι δείκτες ηπατικής καταστροφής



MicroRNAs are more sensitive markers than ALT for liver injury. The comparison among the levels of mir-122 (red bars), mir-192 (green bars), and ALT (blue line) in plasma samples collected from mice at 1 (A) and 3 (B) h after exposure to different doses of acetaminophen (indicated on x axis). The relative change of miRNA expression levels (ratio in log 2 compared with control) is indicated on the left side of the figure and the scale of ALT level is on the right. The relative change of miRNA levels is expressed in log 2 ratio of each treatment condition compared with the corresponding control. The values of miRNA fold change and ALT levels are the average of 4 independent samples from each time point, and the standard deviations are shown as error bars.