Who is this new face in the classroom?

Valentina Perissi Assistant professor of Biochemistry at Boston University

Perissi Lab website: <u>http://blogs.bu.edu/vperissi/</u>

Questions: vperissi@bu.edu

BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.

Follow this format for each person. DO NOT EXCEED FIVE PAGES.

NAME: PERISSI, VALENTINA									
eRA COMMONS USER NAME (agency login): VPERISSI									
POSITION TITLE: Assistant Professor									
EDUCATION/TRAINING (Begin with baccalaureate or other initial professional education, such as nursing,									
include postdoctoral training and residency training if applicable.)									
INSTITUTION AND LOCATION	DEGREE (if applicable)	Completion Date MM/YYYY	FIELD OF STUDY						
University of Torino	Laurea (equivalent to BSc/Master)	07/1997	Molecular Biology						
University of California San Diego	PHD	06/2004	Molecular Pathology						
University of California San Diego	Postdoctoral Fellow	06/2010	Endocrinology and Metabolism						

A. PERSONAL STATEMENT

My current research efforts focus on investigating the role of non-proteolytic ubiquitination in the regulation of cell growth, inflammation and metabolism in cellular homeostasis and disease. After postdoctoral training in Endocrinology and Metabolism in Dr. Rosenfeld's laboratory at UCSD, I have established a fully funded and independent research group in the Department of Biochemistry at Boston University. This transition was supported by a K99-R00 career development award from NIH and a Peter Paul Career Development Professorship from Boston University. In 2013, I secured R01 funding from NIDDK to investigate the crosstalk between ubiquitin signaling and inflammation in the adipose tissue and this year I have received a Breast Cancer Breakthrough Award from the Department of Defense to investigate the regulation of breast cancer cells growth and metabolism through non-proteolytic ubiquitination. Since joining Boston University I have mentored three postdoctoral fellows, one PhD, one master and numerous undergraduate students, the last often funded by the Boston University Undergraduate Research program (UROP Awards). Current lab members consist of one research faculty, one postdoctoral fellow, one research technician, two graduate students.

B. POSITIONS AND HONORS

Positions and Employment

2005-2006	Fellow in Endocrinology and Metabolism, UCSD, NIH T32 Training Grant
2010-2011	Adjunct Assistant Professor of Medicine, University of California San Diego
2011- present	Assistant Professor of Biochemistry, Boston University
2018- present	Co-Director BNORC Adipocyte Biology Core

C. CONTRIBUTIONS TO SCIENCE

.....

1. My most recent, independent research efforts have focused on investigating the **role of the nuclear** receptor cofactor G Protein Suppressor 2 (GPS2), in regulating non-proteolytic ubiquitin signaling

3. My initial work as a graduate student focused on *understanding the basic molecular mechanisms of nuclear receptors' cofactors actions in modulating chromatin remodeling*. My contribution to the field was particularly significant in dissecting the molecular basis of the interaction between nuclear receptors and corepressors as it led to the identification of a previously unknown signature structural motif that differentiate corepressor from coactivators in their ability to bind to unliganded or liganded nuclear

4. I became interested in transcription early on during her undergraduate research thesis and along the years I have developed an extensive knowledge and vast technical expertise in studying transcriptional regulation in mammalian cells and tissues. As a result, I significantly contributed to the basic understanding of the molecular mechanisms underlying the regulation of gene expression in a number of cell models, including human breast and prostate cancer, murine retina and macrophage cells, and contributed in a significant manner to the development of Chromatin Immunoprecipitation (ChIP) techniques (first in vivo ChIP in mouse early embryonic retina).

••••

.....

 ${f V}~$ Who is this new face in the classroom?

What is she going to teach?
 Module 3 and 4: - transcriptome

- transcriptional regulation

She needs help – never taught this class before and wasn't here when Michele introduced the previous chapters

 $\mathbf V$ How to help - Interaction: i) questions

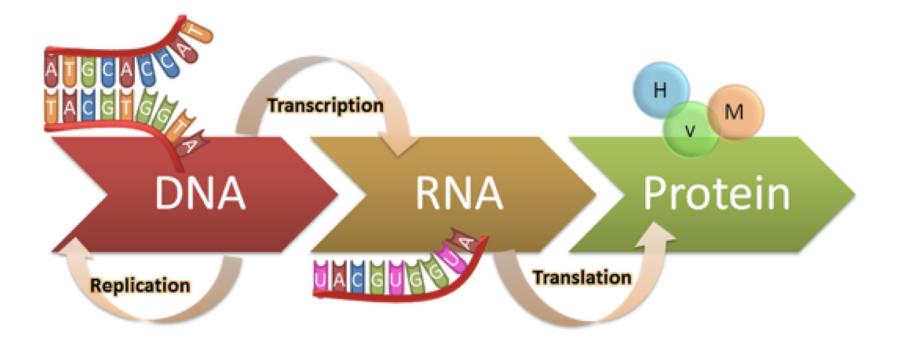
ii) feedback

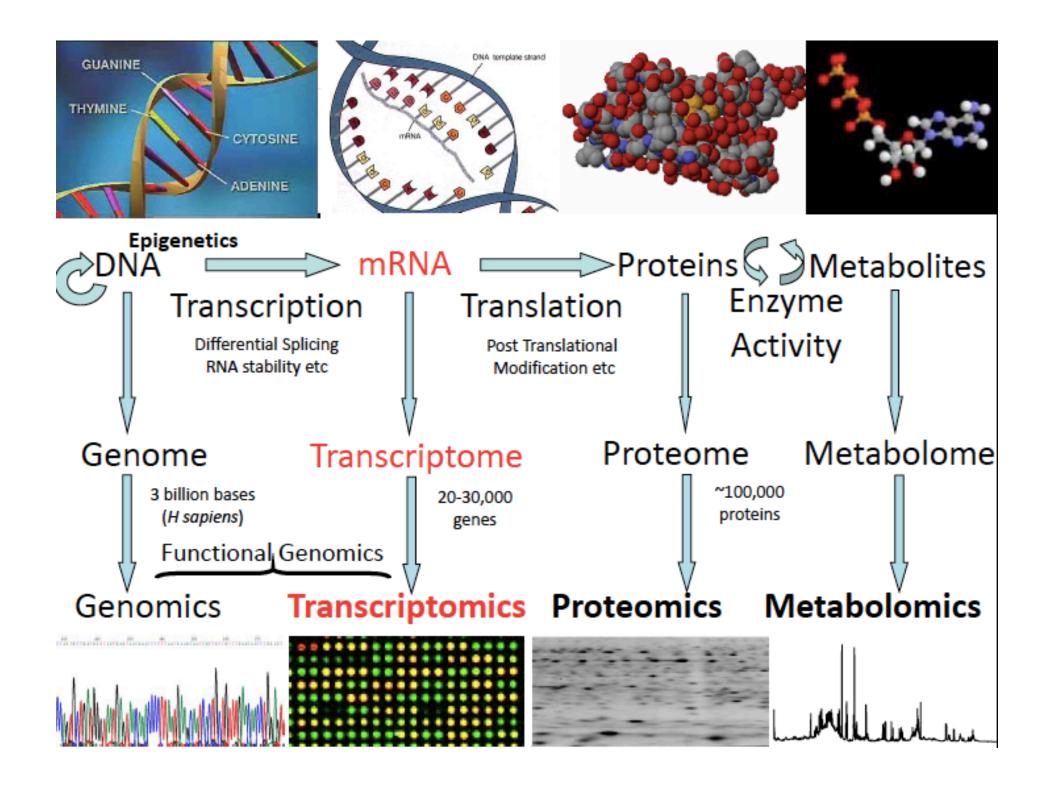
ii) papers discussion

L3.1 – Transcriptomes (Pre-NGS)

What is the "transcriptome"?

What does "transcriptomics" refer to?





The key aims of **transcriptomics** are:

- 1) to catalogue all species of transcript, including mRNAs, noncoding RNAs and small RNAs;
- 2) to determine the transcriptional structure of genes, in terms of their start sites, 5' and 3' ends, splicing patterns and other post-transcriptional modifications;
- to quantify the changing expression levels of each transcript during development and under different experimental or pathological conditions



Two main aspects:

1. Qualitative:

Assemble a full catalogue of all transcribed sequences Describe their function (i.e. are they coding or not)

2. Quantitative:

Perform absolute or comparative measurement of individual RNA transcripts in different samples

AGENDA

- 1. Pre-genomic qualitative approaches (Sanger sequencing of cDNA/EST libraries)
- 2. Pre-genomic quantitative approaches (Northern blotting, RT-PCR..)
- 3. Microarrays

General

Approaches for accessing to RNA:

- 1. Hybridization based methods
- 2. Sequencing based methods
 - 1. The RNA sequence is not observed directly, but it is inferred since it hybridizes with probes or primers.
 - 2. The RNA sequence is converted to DNA (cDNA) and the DNA sequenced (*)

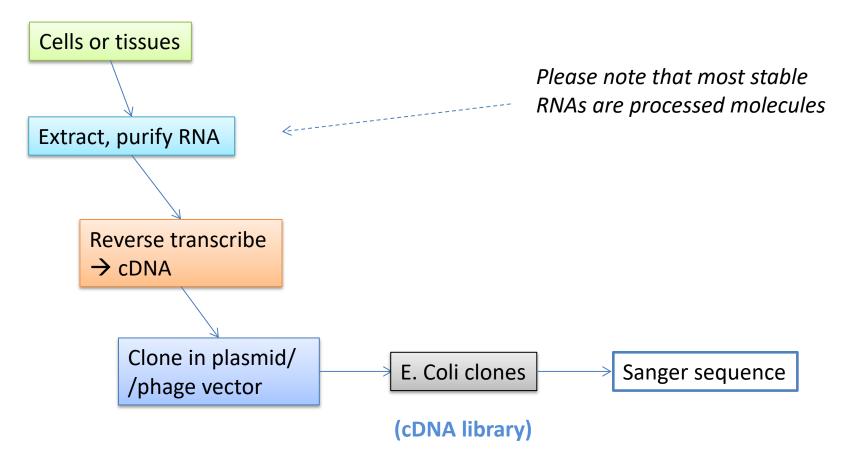
(*) Technologies are being developed to sequence RNA directly without cDNA sysnthesis (Helicos/Nanopore – discussed at the end).

In the **pre-genomic** years, transcriptome was accessed only using single-transcript measurement (or few in parallel)

Pre-genomic

Qualitative

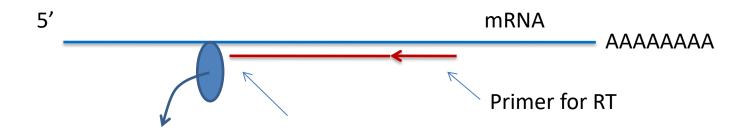
Individual **clones** of cDNA (complementary DNA) were cloned in plasmid/phage vectors



Limitations:

One common problem is that Reverse Transcriptase is not very "processive" and often terminates before reaching the 5' end of RNA.

This has led to 3'-end biased RNA databases and difficulty of mapping the true 5' (and as a consequence, promoters)



Reconstruction of full-lenght cDNA (RNA) required cumbersome analysis.

- RACE
- Primer extension
- Other...

How to complete cDNA structure

RACE or primer extension methods

Database is gradually populated with sequences of various kinds:

- full-lenght cDNA	(usually recostructed by multiple clones, contain all the
	RNA sequence, verified)
- Partial cDNA	(verified, but not containing all the RNA sequence, due to
	partial cloning and/or sequencing)
- EST	(expressed sequence tags, often without annotation)

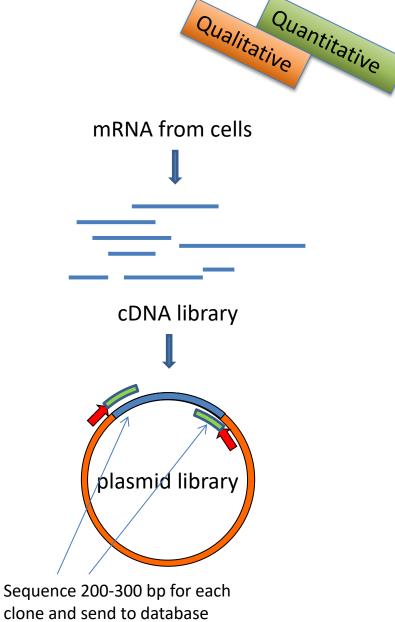
Sequences in database are contributed both by individual scientists and labs, and by Sequencing Centers participating in Genome projects.

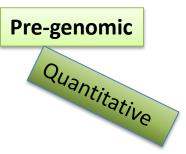
Pre-genomic

EST = expressed sequence tags

- 1. mRNA extraction from cells or tissues
- 2. cDNA synthesis (oligo-dT or random-primed)
- 3. cloning into plasmid vectors
- 4. sequencing from vector primers (200-300nt)
- 5. Estimate expression from frequency

[Current EST Databases contain millions of EST]





Quantitative measurements of single genes:

Northern blotting RNase Protection Assay (RPA) RT-PCR qRT-PCR

Gene-by-gene methods to measure gene expression (mRNA)

Post-genomic approaches: goal to simultaneously analyze a large number of transcripts in any particular cell or tissue (Highthroughput)

Making qualitative and quantitative analysis of RNA highly parallel

Hybridization methods

DNA microarrays, oligonucleotide microarrays.

Spotted arrays (1996) In situ synthesized oligo arrays (1999) Bead-arrays[®] (2001)

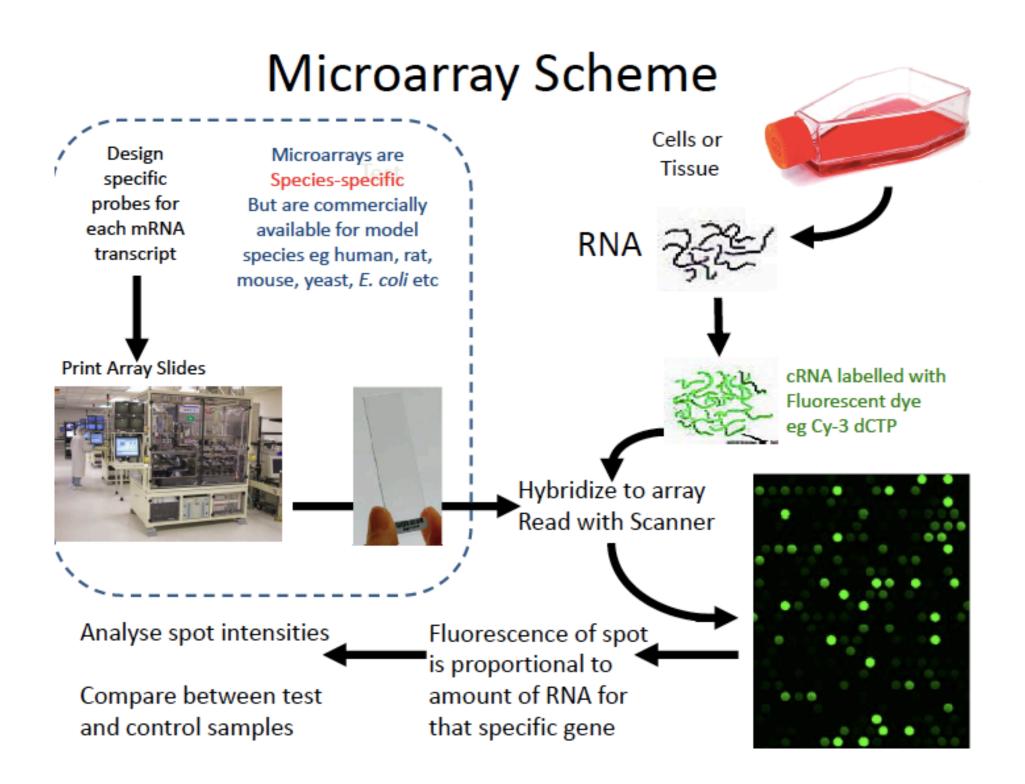
Tiling microarrays (2004)

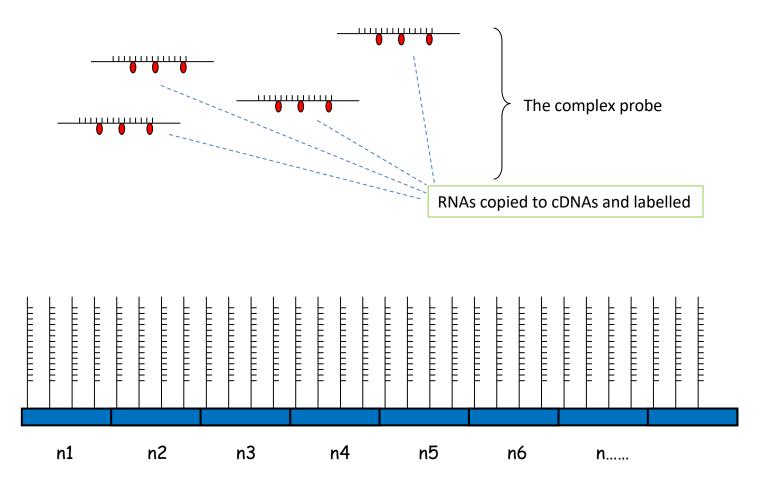
Sequencing methods

EST (1980)

SAGE (1995)(LongSAGE, CAGE)

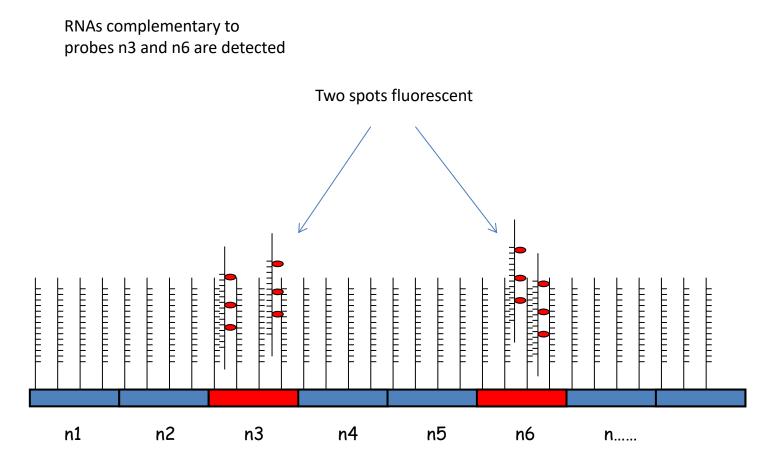
NGS sequencing (2006) (RNA-Seq)





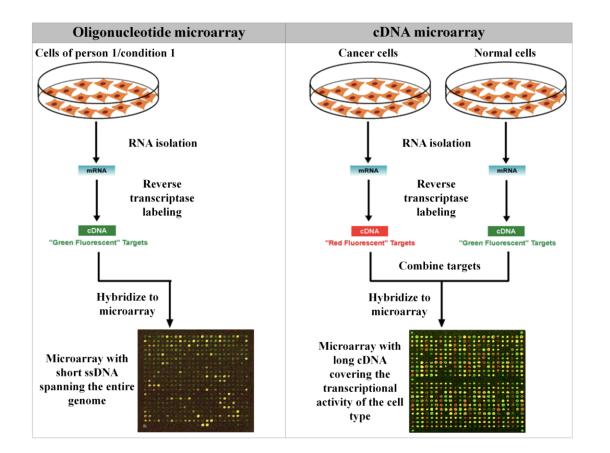
a section of a oligonucleotide microarray at row "n"

microarrays



a section of a oligonucleotide microarray at row "n"

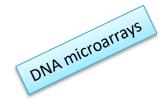




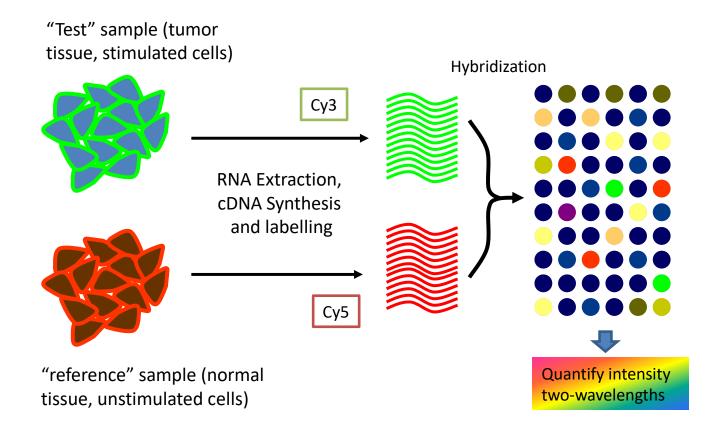
TYPES of DNA MICROARRAYS

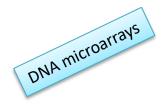
- 1. Oligo Microarray (Short oligo probes, multiple probes per gene usually synthesized on chip)
- 2. cDNA Microarray (Longer cDNA probes usually amplified by PCR, spotted or printed on nylon/glass slides)

- a) Spotted (pens, ink-jet, other technologies)
- b) Printed (ink-jet technology)
- c) In situ-synthesized



Spotted cDNA microarray are best used using double-color, making comparative measures

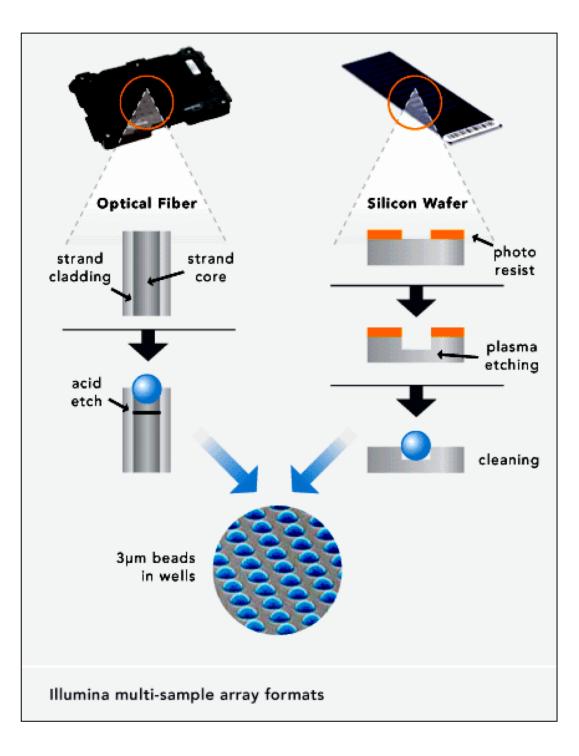


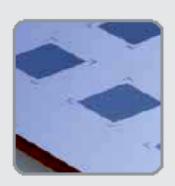


In situ synthesized

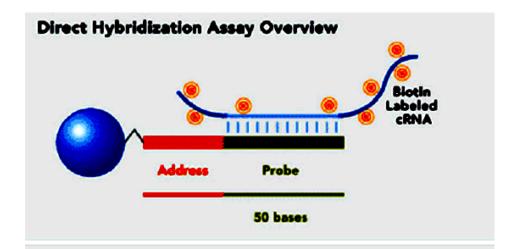
Oligonucleotide arrays (Affychip®) Affymetrix	Over 1,000,000 20-25nt long oligonucleotides / cm ² are sythesized directly on the chip surface, using a photolitographic technique. Each gene is represented by a "probeset" of 12-13 probes.
Bead Arrays [®] Illumina	Oligonucleotide probes (30-50 nt) are synthesized on beads, with an identification address. Beads are randomly arrayed on surfaces and position of each oligonucleotide determined using addresses.
Agilent sure-print	Probes are synthesize on the microarrays using a proprietary DNA synthesis method based upon inkjet printing technology.

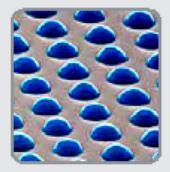
Bead-arrays[®] are patented by Illumina





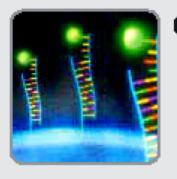
1 Each array cluster contains about 50,000 3-micron beads, or features, assembled in dense geometries.



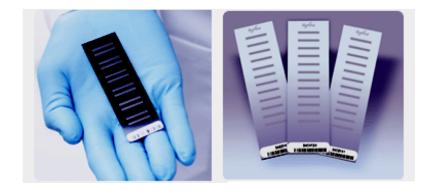


2 Over 1500 probes, or bead types, at >30x average feature redundancy, are represented in each array cluster.

A 50-base gene-specific probe linked to short address. This probe is hybridized to labeled nucleic acid derived from total RNA.



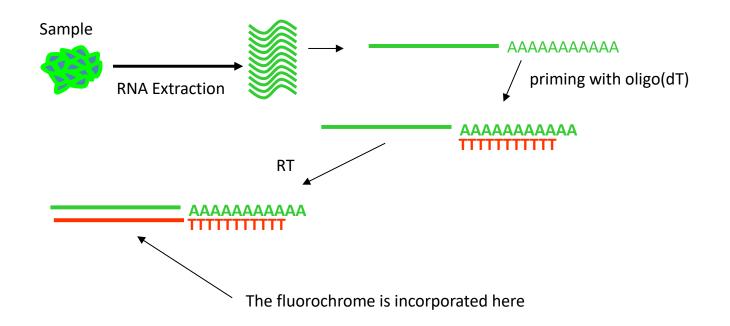
3 Labeled sample targets hybridize to capture probes immobilized on the beads.



PROBE PREPARATION

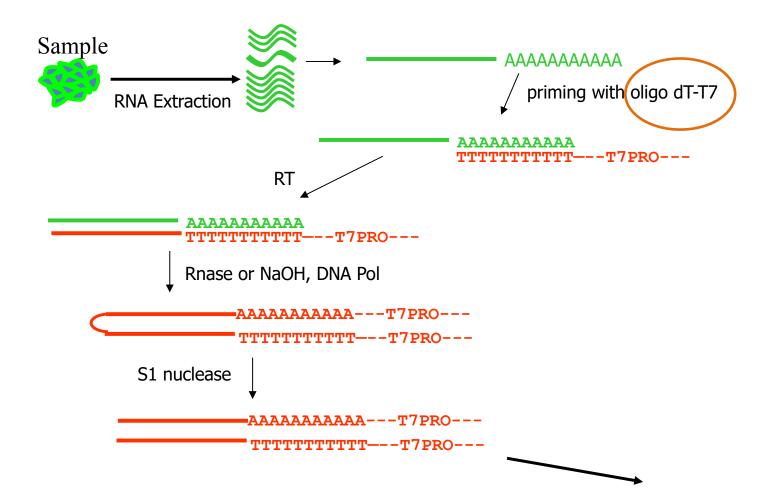
Nonamplified complex probe preparation (very low sensitivity):

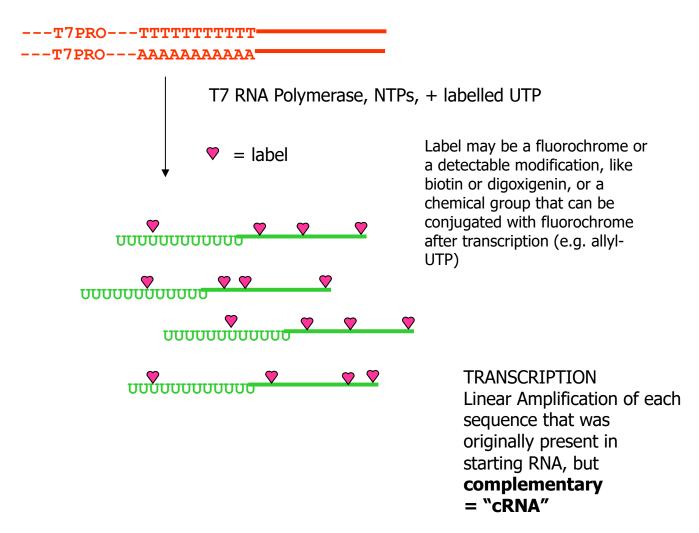
RNAs are labelled using fluorochrome-conjugated ribonucleotides (NTP) directly by reverse transcription (RT), priming the synthesis either with oligo-dT or with random primers.



Linearly amplified complex probe preparation using T7 RNA polymerase

Why not using PCR amplification?





Read-out of a microarray experiment for analysis of gene expression



Primary read out is a table of **Fluorescence intensities**

Gene name	Probe ID	Fluorescence Normalized				
ABCD	Tor123456 346.78	1				
CDH1	Tor123457 111.23	0.324				
HRAS	Tor123458 222.46	0.648				
TFF1	Tor123459 11.12	0.032				
•••••						

•••••

(continues for N lines, from few hundreds up to 50-60,000 for genomewide microarrays)



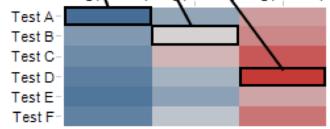
Secondary read-out is a **comparative** expression analysis

In the case of **double-color** assay: For each gene (probe) the ratio between red and green fluorescence is measured. This is called «ratio» and is usually expressed as a Log in base 2 (Log2ratio)

In the case of **single-color** assay, each gene (probe) is associated to a Normalized value, i.e. the ratio of his normalized fluorescence to a **standard reference** gene(s).

Quantitative

24 P478940THC26688151592.824997.1755873.5041607.86456829.5325040.509937.32011259.797590.8834642.92854210.6665307.47732 P8133AK0577091014.571872.695224.9987256.10892096.175213.531888.37971007.321359.0979268.07682749.3212277.64132 P8139PGR367.4936431.2933268.8744256.610892096.175213.531878.37971007.321359.0979251.6292050.6031702.06523 P216225EGR3188.7889117.2561137.351895.94292849.5961378.4002117.642166.3025104.9777127.9327439.5834667.725623 P13293PGR202.7657223.1941165.6498178.8664336.991218.5055122.4987106.5085570.5793502.136523 P13293PGR202.7657223.1941159.4434257.212175.69961108.057421.208287.875276.8222171.06511070.959898.767623 P329768GREB1387.7765397.571458.2524240.04725.6029173.3951792.4646795.0189133.4958500.7196896.31624 P17601UGT2817225.9956308.0241155.428240.0428451.852297.9674204.1438185.0093143.4956676.1984573.002123 P20245RET1163.705693.7436614.1019302.315248.58641702.553696.5198														
United OLGE OLGE <tholge< th=""> OLGE OLGE <t< td=""><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></t<></tholge<>														
vy 5 5 1 2 5 5 7 7 5 5 6 7														
Nome vetrino 1 2 3 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 3 5 6 1 <t< td=""><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></t<>														
Nome vetrino 1 <t< td=""><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></t<>														
Numero vertifica 1 2 3 4 5 6 1 2 3 4 5 6 Sendo Tancoffen Tancoffen Tancoffen Tancoffen Tancoffen Tancoffen Tancoffen Sendo		Nome vetrino												
Colore Red Red Red Red Green Green<														
Tratsmento siCri sTab2 siCri sTab2 stCri stStab2 stStab2 <t< td=""><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></t<>														
Tatkamento siCri sTab2 siCri sTab2							_							
OPENING GPE_1AB2_25 <							•	•					•	•
Intervision Intervision <thintervision< th=""> <thintervision< th=""></thintervision<></thintervision<>		Trattamento	siCtrl	siTab2	siCtrl	siTab2	siCtrl	siTab2	siTab2	siCtrl	siTab2	siCtrl	siTab2	siCtrl
Intervision Intervision <thintervision< th=""> <thintervision< th=""></thintervision<></thintervision<>			CED TADO DE	CED TAD2 25	CED TAR2 25	CED TAR2 25								
U_022- 01_022- <th0_0222-< th=""> <th0_0222-< th=""> <th0_< td=""><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></th0_<></th0_0222-<></th0_0222-<>														
vigos fetor, vigos fe														
Oben Name Gene Name 1.3 1.4 1.2 1.4 1.4 1.2 1.2 1.3 1.4 2.4 P478960 HTC2666815 1592.84 997.755 873.501 100.831 1029.797 500.8814 642.282 500.0824 642.282 500.8834 642.282 500.8834 642.282 500.8834 642.282 500.8834 642.282 500.8834 642.282 500.8837 100.321 350.697 521.620 500.606 707.641 327.216 128.797 500.8834 647.7256 37.643 339.791 351.862 500.777 127.922 439.584 667.7256 37.643.83 238.9917 128.505 121.8055 121.8055 120.9967 51.600 500.776 37.376 448.257 448.257 448.257 448.257 448.257 448.257 421.206 218.505 121.8055 120.9957 114.8007 500.719 698.376 32.929760 OREB1 387.716 388.241 157.506 100.808.16 150.200.11 139.751 321.449 <td></td>														
24 P278940 THC:068815 1522.824 997.1755 873.5041 607.8455 6827.532 215.3065 973.301 1229.797 500.838 642.2925 4210.666 530.7477 32 P38163 Ax057709 104.571 877.656 284.9991 256.108 1716.041 451.8824 339.7921 353.0977 251.639 2207.661 32 P18225 E683 118.8788 117.2661 137.318 95.4922 849.3661 374.402 117.6472 166.3057 104.9777 127.3927 439.884 667.7256 32 P18267 CITED1 256.6044 242.6861 100.6541 106.6672 397.4761 127.8685 512.8685 507.5758 502.1365 502.136 502.136 502.136 502.136 502.136 502.137 502.1466 502.662.137 50			2_NO_Spikein	2_NO_Spikein	2_NO_Spikein	2_NO_Spikein	2_NO_Spikein	2_NO_Spikein	2_NO_Spikein	2_NO_Spikein	2_NO_Spikein	2_NO_Spikein	2_NO_Spikein	2_NO_Spikein
$\frac{32}{32} passis}{32, passis} = \frac{1014 4571}{32, passis} = \frac{224, 987}{225, 2233} = \frac{226, 1029}{206, 226, 127} = \frac{2133, 331}{227, 641} = \frac{883, 3797}{1007, 321} = \frac{355, 0579}{350, 0579} = \frac{260, 758}{251, 629} = \frac{277, 643}{277, 643} = \frac{277, 644}{277, 643} = \frac{117, 273}{277, 644} = \frac{277, 644}{277, 643} = \frac{117, 273}{277, 644} = \frac{117, 273}{277, 757} = \frac{117, 273, 273}{272, 223, 224, 5576} = \frac{117, 273, 273}{272, 223, 224, 5576} = \frac{117, 273, 273}{272, 223, 224, 5577} = \frac{117, 273, 273}{272, 223, 224, 5577} = 112, 273, 273, 273, 273, 273, 273, 273, 27$	robeName	Gene Name	_1_3	_1_4	_1_1	_1_2	_1_3	_1_4	_1_3	_1_4	_1_1	_1_2	_1_3	_1_4
32. Padrogs 967.4939 968.41.2933 296.8744 286.6231 195.634 171.6041 451.8224 333.0971 251.629 295.06693 1702.05 23. P16225 EGR3 188.7889 117.2661 173.518 95.4029 843.5661 378.0021 117.6414 451.8224 333.0971 251.629 259.06693 1702.05 23. P15256 CIED1 256.0484 424.8688 100.6554 106.6672 586.6913 574.1368 239.8019 218.5055 122.4987 106.5085 570.5793 502.1365 23. P38407 Ust7815 220.189 327.214 194.4447 277.2121 755.6966 1106.0574 22.864 107.055 109.779 524.827 276.822 177.059 887.767 39.87.751 488.445 206.348 173.5951 172.4482 290.148 380.097 31.43.8466 675.032 31.43.986 671.094 97.0021 32.87.744 20.7488 31.6461 105.5763 223.889 31.0120 39.6255 165.5763 223.889 31.0120 31.0120 31.0201 32.986.253 155.0783 223.889 31.0120 <td>_24_P478940</td> <td>THC2668815</td> <td>1592.824</td> <td>997.1755</td> <td>873.504</td> <td>607.8645</td> <td>6829.53</td> <td>2 5040.509</td> <td>937.320</td> <td>1 1259.79</td> <td>7 590.8834</td> <td>642.928</td> <td>5 4210.66</td> <td>6 5307.477</td>	_24_P478940	THC2668815	1592.824	997.1755	873.504	607.8645	6829.53	2 5040.509	937.320	1 1259.79	7 590.8834	642.928	5 4210.66	6 5307.477
212 2	_32_P38153	AK057709	1014.571	1 872.695	284.998	7 256.1089	2096.17	5 2135.35	888.379	7 1007.32	1 359.0979	268.0768	8 2749.32	1 2277.641
22 22 23 24 25 25 24 25 25 24 25 25 24 25 25 24 25 25 25 25 25 25 25 25 25 25 25 27 26 26 <td< td=""><td>_32_P49199</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></td<>	_32_P49199													
22.9 138938 PGR 202.0567 223.1904 166.2488 178.8664 85.921 950.671 279.9407 197.215 201.922 154.4607 911.349 882.6516 23.9 58407 UGT2815 290.318 392.5214 159.4444 257.211 755.696 108.057 421.2808 287.787 276.822 171.0651 1070.597 6665.519 139.733 141.534 500.719 6966.316 32.9 29669 NPYIR 1125.268 151.697 077.773 122.4469 250.027 947.734 257.744 244.892 256.092 207.6822 10.080.51 697.609 207.976 214.482 256.092 207.618 108.051 697.609 207.976 214.882 259.092 207.751 102.092 127.986 77.6223 100.013 943.625 105.9763 223.5989 315.400 315.400 315.400 315.400 315.400 315.400 315.400 315.400 315.400 315.406 316.41 315.736 105.9763 123.586 110.9867 472.9865 452.9623 315.400 315.400 315.400 315.400 315.400	_23_P216225	EGR3	188.7889			8 95.94292	2 849.596	1 378.4002	117.6472					4 667.7256
22.98807 UCT2N15 29.0318 397.551 159.444 257.2121 755.6996 1100.057 421.208 227.755 276.822 171.0651 1070.959 898.767 23.923768 GRE11 387.7765 397.7517 458.2526 438.8445 205.436 173.951 792.4646 795.0189 139.793 1415.548 5009.719 6896.316 23.920246 UGT2B17 125.593 151.1697 1073.773 1224.392 4468.823 451.35 2907.967 2214.882 2500.699 2207.618 1008524 673.3021 23.920245 NFT 1163.705 693.7486 614.1019 302.24657 143.0269 92.293 157.3606 157.2323 120.7246 148.6814 135.7534 609.369 151.409 23.923541 CBorde 117.3481 130.8704 123.666 157.3233 120.7246 148.6814 135.7534 609.366 673.329 23.91557 MWBL 1035.348 538.7544 102.305 50.977 114.0966 117.0981 116.7034 676.53.329 23.9155258 114.9933 32.012.265 </td <td>_23_P73526</td> <td>CITED1</td> <td>256.0494</td> <td>4 242.8608</td> <td>100.6554</td> <td>4 106.6672</td> <td>2 586.6913</td> <td>3 574.1368</td> <td>3 239.8019</td> <td>218.505</td> <td>5 122.498</td> <td>106.5085</td> <td>5 570.579</td> <td>3 502.1365</td>	_23_P73526	CITED1	256.0494	4 242.8608	100.6554	4 106.6672	2 586.6913	3 574.1368	3 239.8019	218.505	5 122.498	106.5085	5 570.579	3 502.1365
22 923768 GEB1 387.765 397.7517 458.2526 438.8445 2063.366 1733.951 792.4666 795.0189 139.973 1415.548 5009.719 6696.316 24 P17691 UG72B17 225.9956 308.0241 155.5428 244.09 725.0027 934.7345 297.7644 204.143 185.0093 143.4956 676.1984 573.0021 32 P020245 RET 1163.705 693.7436 614.1019 302.3615 248.864 1702.553 696.5159 810.0036 319.0016 433.6257 1255.428 248.964 137.775 1105.754 603.366 115.1439 125.7641 130.269 92.993 157.303 124.832 130.266 319.0016 433.027 125.667 676.198 513.449 430.625 165.3764 133.857 403.9266 427.056.518 133.74663 270.6644 163.2549 854.666 675.329 32.912579 H998 319.048 537.044 102.336 501.996 374.4663 126.844 163.2549 8554.666 675.329 32.91253 M91.16 105.3537 114.133 104.554<	_23_P138938	PGR	202.7657	7 223.1904	166.2498	8 178.8664	4 836.92	1 950.673	279.940	7 197.231	5 201.9929	154.460	7 911.34	9 842.6516
2/2 2	_23_P58407	UGT2B15	290.318	392.5214	159.4434	4 257.2121	1 755.6996	6 1108.05	7 421.2808	8 287.787	5 276.8222	171.065	1 1070.95	9 898.7676
23-966699 NPY1R 1252.663 1511.697 1073.773 1224.392 4469.823 4513.5 2907.967 2214.892 2500.699 2207.618 10689.16 9776.069 23-97877 C0365380 143.910 149.429 125.7661 143.020 92.929 157.308 151.409 125.7661 127.838 696.519 891.0366 139.0103 493.6255 1659.763 2235.899 32.7578.77 1058518 131.403 100.897.763 2235.899 32.7578.77 105.951 144.1019 302.361 123.206 126.1182 131.7437 110.9967 472.9365 492.9662 32.916257 HSPB8 3140.044 548.744 1683.745 2566.708 677.1538 6117.95 533.493 3374.66 126.259 8554.665 169.2567 429.9662 149.2861 149.286 140.988 101.608 101.608 101.608 101.608 101.608 101.608 101.608 101.608 102.569 8554.666 151.409 341.987 Test 42.022 3.21 5.577 32.9103345 CKL12 697.8454 128.649 136.633	_23_P329768													
23-P202245 RET 1163.705 693.7436 614.1019 302.3615 2485.864 1702.553 699.1036 319.0103 493.6255 1659.763 2235.899 32.P57877 C0365380 143.9107 149.4294 125.7661 143.029 92.993 157.3608 157.2323 120.7246 148.6814 135.7534 609.369 151.313 131.71471 110.9076 472.3956 492.9662 23.P152579 H\$P88 3419.048 5547.424 1683.745 2566.708 6719.538 8617.95 533.431 337.4663 2706.644 1632.549 8554.665 6753.329 24.P520201 TH (2553558 114.9983 107.5519 110.133 120.1143 796.6127 114.0366 121.6014 870.427 116.0014 870.427 116.003 2706.644 153.2549 8554.665 6753.329 23.P15255 MYBL 1053.348 538.7504 1012.306 501.590 376.47 Test Z.022 3.21 5.577 23.P15255 MYBL 13673.57 1378.152 10014.45 1063.03 36387.2 7.19 13.628	_24_P17691													
32 P57877 C0365380 143.9107 149.4294 125.7661 143.0269 922.993 157.3608 157.2323 120.7246 148.6814 135.7534 609.369 151.409 22.9535364 C80r46 117.3481 130.8704 124.6863 124.9957 468.4099 442.0125 133.374.663 270.6644 135.7534 609.369 149.9662 24.9520201 THC2553558 114.9983 107.5519 114.1133 120.1143 796.6127 114.0966 115.0374.663 270.6644 135.753 609.359 157.353.29 23.P102579 MYBL1 1035.348 558.7504 1012.306 501.9996 374.44 Test Time 1 Time 2 Time 3 23.P108351 HEV2 222.4111 362.0376 151.434 228.0122 50.177 Test A 2.02 3.21 5.577 24.910022 XBP1 1367357 1378135 100.554 91.3075 341.982 Test A 2.02 4.377 6.02 24.910028 KBP1 135.628 152.637 101.554 91.3075 341.982 Test D	_23_P69699													
23-9353614 C8ort46 117.3481 130.8704 124.6863 124.9572 468.4099 442.0125 123.2606 126.1182 131.7437 110.9967 472.9365 492.9662 23-p152579 HsPB8 3419.048 5547.424 1683.745 256.708 671.953 8617.25 533.391 337.4663 2706.644 153.2549 8554.665 6753.329 23-p4157 MVBL1 1035.348 538.7504 1012.306 501.9906 374.4 Test Time 1 Time 2 Time 3 23-p202448 CXL112 679.2454 710.6996 156.1434 228.012 550.17 Test A 2.02 3.21 5.57 23.21 557.37 1378.152 1001.485 10653.04 3687.7 Test B 2.02 4.37 6.02 24.P100228 X8P1 136.757 1378.152 1001.485 10563.04 3687.7 Test B 2.02 4.37 6.02 24.P100228 X8P1 1367.57 1378.152 1001.485 10563.04 3687.7 Test B 2.02 4.37 6.02 24.P510666	_23_P202245													
22 122 123 1452579 145988 3419.048 5547.424 1683.745 2566.708 6719.538 8617.95 5334.391 3374.663 2706.644 1632.549 8554.665 6753.329 23 P202001 THC2553558 114.9983 107.5519 114.1133 120.1143 796.6127 114.0986 114.0986 115.0024 679.4726 126.908 573.329 23 P10488 CXL12 679.2454 719.6995 560.4945 597.2649 2245.06 Test Time 1 Time 2 Time 3 23 P10288 HV2 222.4111 362.0376 151.4344 2280.022 550.77 Test A 2.02 3.21 5.57 24 P102028 XBP1 13673.57 13781.52 10014.85 10653.44 36387.2 Test B 2.02 4.37 6.02 24 P102028 XBP1 13673.57 13781.52 10014.85 10653.44 36387.2 Test B 2.02 4.37 6.02 24 P11666 G616845 918.5523 663.974 512.1214 4														
24 P520201 THC2553558 114,9983 107,5519 114,1133 120,1143 796,6127 114,036 127,0388 114,038 121,005 115,0024 Cond (726) 126,0034 Cond (726) Cond	_23_P353614													
22 P43157 MYBL1 1035.348 538.7504 1012.306 501.9906 3764 Test Time 1 Time 2 Time 3 22 p433157 MYBL1 679.2454 719.6995 560.4945 597.2649 224.500 7 <td></td>														
12, 2/2/2/449 OXLI12 0/9, 4/94 12/9, 6/93 300, 4/94 300, 4/94 300, 2/94 2/2, 4/30, 2/2, 4/30, 2/3, 2/1 32, P168351 HY2 222, 4/11 362, 0/36 151, 34/4 228, 0/2 2/4, 30/3 3.2.1 5.5.7 24, P100228 X8P1 13673, 57 1378, 152 1001, 48 10664, 30/3 36387, 3 341, 98 Test A 2.02 3.21 5.5.7 24, P100228 X8P1 13673, 57 1378, 152 1001, 48 1063, 0/4 36387, 3 341, 98 Test B 2.02 4.37 6.02 24, P01417 PTGES 128, 3569 115, 2637 101, 5564 91, 301, 75 341, 98 Test B 2.092 4.37 6.02 24, P511686 CR616845 918, 5523 963, 9754 521, 2184 499, 116 1763, 44 584, 15 23, P148609 PLAC1 212, 4161 231, 6628 172, 3428 151, 8443 584, 15 Test C 2.04 5.02 7.19 Test E 2.21 3.12 5.38 152.37 3.44 8.21 5.43 Test F 2.4								7 11/1 9204	177 9385					6 126 1093
12/202449 OLLI2 6/9/2494 12/0995 500/4945 22/24/05 23/P16851 HY2 22/2111 36/2076 15/14342 22/8012 25/017 23/P16851 HY2 22/2111 36/2076 15/14342 22/8012 16/57.4 Test A 2.02 3.21 5.57 24/P100228 X8P1 13673.57 13781.52 1001485 10/63.04 36387.7 Test B 2.02 4.37 6.02 24/P010228 X8P1 13673.57 13781.52 1001485 10/63.04 36387.7 Test B 2.02 4.37 6.02 24/P01028 X8P1 13673.57 13781.5247 101.1564 1763.44 Test C 2.04 5.02 7.19 24/P0104 212.4161 231.6628 172.3428 1518443 584.14 Test C 2.04 5.02 7.19 Test D 2.33 3.44 8.21 Test E 2.21 3.12 5.38 70 best visualize these data? Test F 2.43 3.84 6.47 Avg((Time 1) Avg(Time 2) Avg(T								Test	Tim	e1 1	fime 2) Tim	ie 3	
23-932535 PPPIRI6B 257.9374 228.1916 253.6183 312.2321 1657.4(Test A 2.02 3.21 5.57 24-P00228 WB1 13673.57 13781.52 1001485 10063.04 33897.2 Test A 2.02 3.21 5.57 24-P00347 Proces 128.3509 155.237 1015.850 91.30175 314.982 Test B 2.02 4.37 6.02 24-P01347 Proces 918.5523 863.9754 521.2184 499.116 1763.4(Test B 2.02 4.37 6.02 23-P148609 PLAC1 212.4161 231.6628 172.3428 151.843 584.15 Test C 2.04 5.02 7.19 Test D 2.33 3.44 8.21 Test E 2.21 3.12 5.38 Test F 2.43 3.84 6.47 4.47 4.49 4.49 4.49 4.43 4.47 400 constructure Test F 2.43 3.84 6.47 4.49 4.49 4.49 4.43 4.47 4.49 4.49 4.49 4.43 4.47														
24_P10028 X8P1 13673.57 13781.52 1001485 10653.04 36387.7 24_P403417 PTGES 128.3569 115.2637 101.5564 91.0175 341.98 Test B 2.02 4.37 6.02 24_P511666 CR616845 918.5523 863.9774 521.2184 499.116 1763.47 Test B 2.02 4.37 6.02 23_P148609 PLAC1 212.4161 231.6628 172.3428 151.8443 584.15 Test C 2.04 5.12 7.19 Test D 2.33 3.448 8.21 172.3428 151.8443 584.15 Test E 2.21 3.12 5.38 Test E 2.21 3.12 5.38 172.3428 151.8443 584.15 Test F 2.43 3.84 6.47 Avg((ime 1)) Avg((ime 1)) Avg((ime 1)) Avg(Time 2) Avg(Time 2) Avg(Time 2) Avg(Time 2)								Toot A	21	0.2	2.24	6	67	
24-9403417 PTGES 128.3569 115.2637 101.5564 91.3015 341.981 Test B 2.92 4.37 6.02 24.9511686 CR616845 918.5523 863.9754 521.2144 499.116 1763.47 Test B 2.92 4.37 6.02 23.9148609 PLAC1 212.4161 231.6628 172.3428 151.843 584.15 Test C 2.04 5.02 7.19 Test D 2.37 3.48 8.21 Test E 2.21 3.12 5.38 Test F 2.43 3.84 6.47 Avg(Time 1) Avg(Time 2) Avg(Time 2) Avg(Time 2)								ICSLA	Z.	02	3.21	э.	.07	
23.p148609 PLACI 212.4161 231.6628 172.3428 1951.10 1954.1 Test C 2.64 5.12 7.19 Test D 2.37 3.48 8.21 Test E 2.21 3.12 5.38 Test F 2.43 3.84 6.47 Avg(time 1) Avg(Time 2) Avg(T											4.05	-	~~	
23.p148609 PLACI 212.4161 231.6628 172.3428 1951.10 1954.1 Test C 2.64 5.12 7.19 Test D 2.37 3.48 8.21 Test E 2.21 3.12 5.38 Test F 2.43 3.84 6.47 Avg(time 1) Avg(Time 2) Avg(T								Test B	2)	92	4.37	6.	.02	
Test C 2.64 5.02 7.19 Test D 2.37 3.48 8.21 Test E 2.21 3.12 5.38 Test F 2.43 3.84 6.47 Avg(time 1) Avg(Time 2) Avg(T											1			
o best visualize these data? Test D 2.3 3.48 8.21 Test E 2.21 3.12 5.38 Test F 2.43 3.84 6.47 Avg(time 1) Avg(Time C) Avg(T	23_P148609	PLACI	212.416	1 231.6628	5 172.3428	5 151.844:	5 584.15	Test C	21	614	5 02	7	19	
o best visualize these data? Test E 2.21 3.12 5.38 Test F 2.43 3.84 6.47 Avg((ime 1) Avg(Time 2) Avg(T										~~	- Y-			
o best visualize these data? Test E 2.21 3.12 5.38 Test F 2.43 3.84 6.47 Avg(Time 1) Avg(Time 2) Avg(Time 2)								Teet D	2.	হ 🖈 👘	3 / A	8	21	
o best visualize these data? Test F 2.43 3.84 6.47 Avg(Time 1) Avg(Time 2) Avg(Time 2)								reaco	2.	<i>31</i>	J.40		2	
o best visualize these data? Test F 2.43 3.84 6.47 Avg(Time 1) Avg(Time 2) Avg(Time 2)								Test F	2 1	21	3.12	۲ ۱	X .	
O best visualize these data? Avg(Time 1) Avg(Time 2) Avg(T								TGALE	2.	<u> </u>	0.12	1 .	~	
O best visualize these data? Avg(Time 1) Avg(Time 2) Avg(T						_		Test F	2	43 \	3.84	16	47	
Avg(lime 1) Avg(lime X) Avg(l	ta hea	st vicual	ize th <i>i</i>	ace r	lataî	2		10at I	2.	<u>۱</u>	0.04	_ \ °.	· · ·	
					auta	•				Ava	Time 1) Ava(Time	Ava(1)
Test A													1	X
								т.	ect A -					

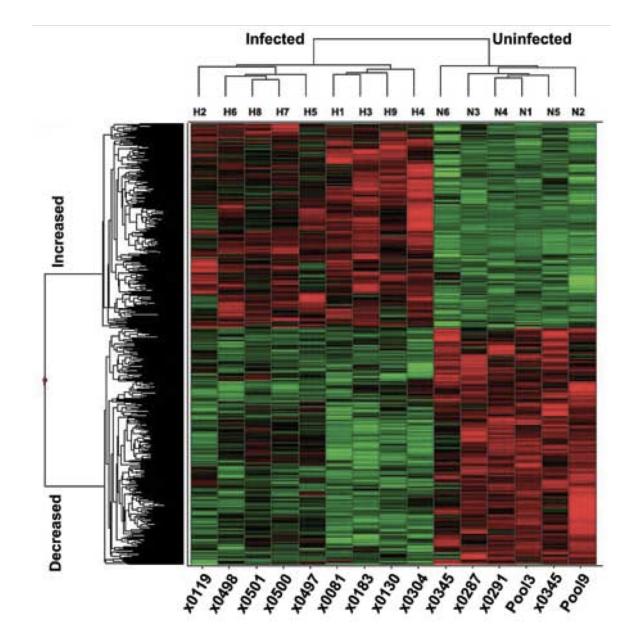


How

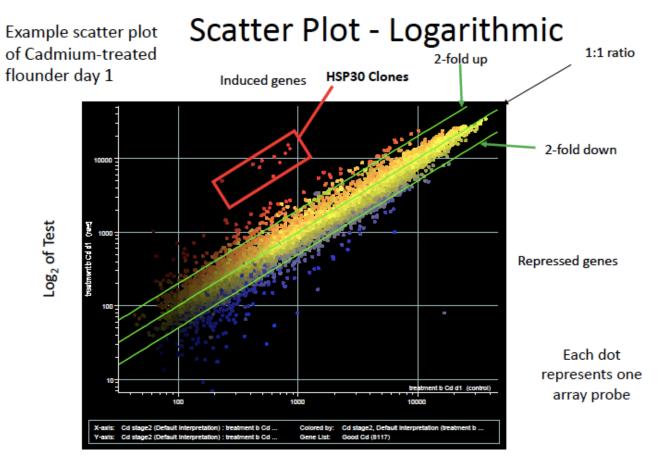
Gene expression data is often displayed as a heatmap. In heat maps the data is displayed in a grid where each row represents a gene and each column represents a sample.

The heatmap can be combined with clustering methods which group genes and/or samples together based on the similarity of their gene expression pattern (supervised or unsupervised hierarchical clustering).

This can be useful for identifying genes that are commonly regulated, or biological signatures associated with a particular condition (GO analysis).



Alternative representation: SCATTER PLOTS





An Example Application

- · 72 leukemia patients
 - 47 ALL
 - 25 AML
- 1 chip per patient
- 7132 human genes per chip

Golub, et al., Science 286:531-537 (1999).



REPORTS

Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring

T. R. Golub, ^{1,2*}[†] D. K. Slonim, ¹[†] P. Tamayo, ¹ C. Huard, ¹
 M. Gaasenbeek, ¹ J. P. Mesirov, ¹ H. Coller, ¹ M. L. Loh, ²
 J. R. Downing, ³ M. A. Caligiuri, ⁴ C. D. Bloomfield, ⁴
 E. S. Lander^{1,5*}

Key Issue: What's Different?

- · What genes are behaving differently between ALL & AML (or other disease/normal states)?
- Potential uses:
 - Diagnosis
 - Prognosis
 - Insight into underlying biology/biologies
 - Treatment

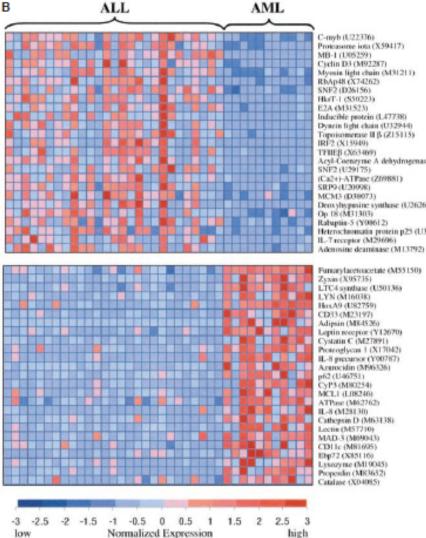


Fig.2: 50 genes with highest correlation with the two diseases

Myosia light chain (M31211) Inducible protein (L47738) Dynein light chain (U32944) Topoisomerase II B (Z15115) Acyl-Coenzyme A dehydrogenase (M91432) (Ca2+)-ATPase (Z09881) Deoxyhypusine synthuse (U26266) Heterochromatin protein p25 (U35451) IL-7 receptor (M29696)

Fumarylacetoacetate (M55150) LTC4 synthase (US0136) Leptin receptor (Y12670) Cystatin C (M27891) Proteoglycan 1 (X17042) IL-8 precursor (Y90787) Azurocidin (M96326) Cathepsin D (M63138) Lysecyne (MI 9045) Properatin (M83652)

www.nature.com/leu

ORIGINAL ARTICLE

Molecular heterogeneity in chronic lymphocytic leukemia is dependent on BCR signaling: clinical correlation

A Rodríguez¹, R Villuendas¹, L Yáñez², ME Gómez³, R Díaz⁴, M Pollán⁵, N Hernández³, P de la Cueva¹, MC Marín¹, A Swat¹, E Ruiz⁶, MA Cuadrado², E Conde², L Lombardía⁷, F Cifuentes⁸, M Gonzalez⁹, JA García-Marco¹⁰ and MA Piris¹ for Spanish National Cancer Centre (CNIO)

Chronic lymphocytic leukemia (CLL), the most frequent form of adult leukemia in Western countries, is characterized by a highly variable clinical course. Expression profiling of a series of 160 CLL patients allowed interrogating the genes presumably playing a role in pathogenesis, relating the expression of functionally relevant signatures with the time to treatment.

First, we identified genes relevant to the biology and prognosis of CLL to build a CLL disease-specific oligonucleotide microarray. Second, we hybridized a training series on the CLLspecific chip, generating a biology-based predictive model. Finally, this model was validated in a new CLL series. Clinical

identifies three risk-score groups with treatment-free survival probabilities at 5 years of 83, 50 and 17%. This molecular predictor can be applied to early clinical stages of CLL. This

Genes associated with CLL and progression: gene selection

We used a multistep approach to develop a predictive model for progression in CLL. First, a highthroughout analysis was performed in a previous series of 23 cases of CLL21 using the Agilent Human 1A 22K Oligonucleotide Microarray, identifying 88 genes statistically associated with survival (adjusted P-valueo0.2).

After a comprehensive literature search, we also selected all the genes previously described as being associated with pathogenesis and prognosis of CLL, 409 genes derived from 23 publications. In total, 497 genes (Supplementary Table 5) were obtained from the first step and the literature search and were included in the CLL-specific oligonucleotide microarray.

For control purposes, 58 of these selected genes were printed in duplicate or triplicate on the microarray.

The 526 genes with a low variability of expression in the 23 cases hybridized in the first step were included as normalization genes. Oligonucleotide sequences as internal controls of hybridization (323 sequences) were also incorporated.

Finally, 1900 sequences were printed in known positions on the microarray using an eight-pack microarray format. The array description has been submitted to ArrayExpress (accession number: A-MEXP-328).

We performed genomic-scale gene-expression profiling of a series of 160 untreated CLL patients using the original CLLspecific microarray. This was divided between a training set (98 patients) to create a predictive model and a validation set (62 patients) from a different Institution to corroborate the results.

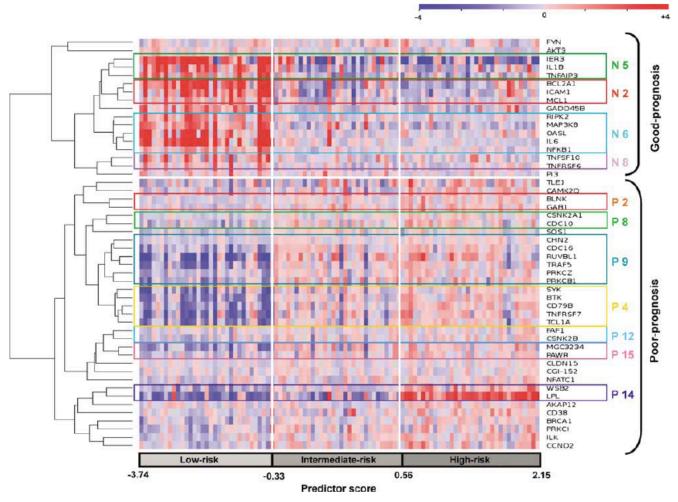


Figure 2 Heat map of prognosis genes. Identification of the signatures associated with variation in time to progression in training series. The series has been divided into terciles. Cases with increased expression of the genes associated with BCR signaling have a shorter time to progression, while cases with greater expression of MAPK/NF- κ B genes have a more favorable outcome. Gene expressions are represented as median center values.

Rodriguez et al., Leukemia 2007

Table 3 KEGG pathway annotation for 50 genes associated with TFS

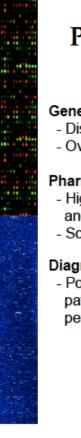
KEGG pathways	Genes	No. of genes
Genes associated with shorter time to progression	n	
B-cell receptor signaling pathway	BTK, CD79B, BLNK, NFATC1 and SYK	5
Wnt signaling pathway	CCND2, CSNK2B, RUVBL1, CAMK2D and NFATC1	5
Tight junction	PRKCZ, CSNK2B, PRKCI and CLDN15	4
Focal adhesion	CCND2, SOS1 and ILK	3
Insulin signaling pathway	SOS1, PRKCZ and PRKCI	3
Calcium signaling pathway	CD38, CAMK2D and NFATC1	3
T-cell receptor signaling pathway	SOS1 and NFATC1	2
Jak-STAT signaling pathway	CCND2 and SOS1	2
Genes associated with longer time to progressio	n	
MAPK signaling pathway	GADD45B, MAP3K8, IL-1B, TNFRSF6, NF-κB1 and AKT3	6
Apoptosis	IL-1B, TNFRSF6, TNFSF10, NF-kB1 and AKT3	5
Toll-like receptor signaling pathway	IL-6, IL-1B, NF-kB1 and AKT3	4
Cytokine-cytokine receptor interaction	IL-6, IL-1B, TNFRSF6 and TNFSF10	4
T-cell receptor signaling pathway	FYN, MAP3K8, NF- <i>k</i> B1 and AKT3	4
Focal adhesion	FYN and AKT3	2
Hematopoietic cell lineage	IL-6 and IL-1B	2
Adipocytokine signaling pathway	NF-KB1 and AKT3	2
B-cell receptor signaling pathway	NF-kB1 and AKT3	2
Jak-STAT signaling pathway	IL-6 and AKT3	2

Abbreviations: KEGG, Kyoto Encyclopedia of Genes and Genomes; TFS, treatment-free survival.

Tomorrow: How to perform pathway analyses

Rodriguez et al., Leukemia 2007

DNA microarrays have been extensively used in the last decade and <u>are still largely used</u> to gather qualitative and quantitative information on known transcribed sequences.



Practical Applications of Microarrays

Gene Target Discovery

- Diseased vs normal cell comparison suggests sets of genes having key roles.
- Over/underexpressed genes in the diseased cells can suggest drug targets

Pharmacology and Toxicology

- Highly sensitive indicator of a drug's activity (pharmacology) and toxicity (toxicology) in cell culture or test animals.
- Screen or optimize drug candidates prior to costly clinical trials.

Diagnostics

 Potential to diagnose clinical conditions by detecting gene expression patterns associated with disease states in either biopsy samples or peripheral blood cells.

CONCLUSIONS on MICROARRAYS

Microarrays were the first, widely used method to address the transcriptome in many experimental and pathological sistuations

Very important for pathological characterization: guide to prognosis and treatment

Very important experimentally: DE gene have functional significance (G.O.), belong to defined functional pathways and can also be explored to identify regulatory motifs.

Pros:

- 1. Easy to use and analyze
- 2. Fast readout
- 3. Relatively cheap (possibility to limit to specific transcriptomes)

<u>Cons:</u>

- 1. Probes describe only «part» of the transcript
- 2. Alternatively spliced exons can be missed
- 3. Novel transcripts are missed
- 4. Quantitation is indirect
- 5. Comparison among different platforms problematic