Introduction to microarrays

Overview

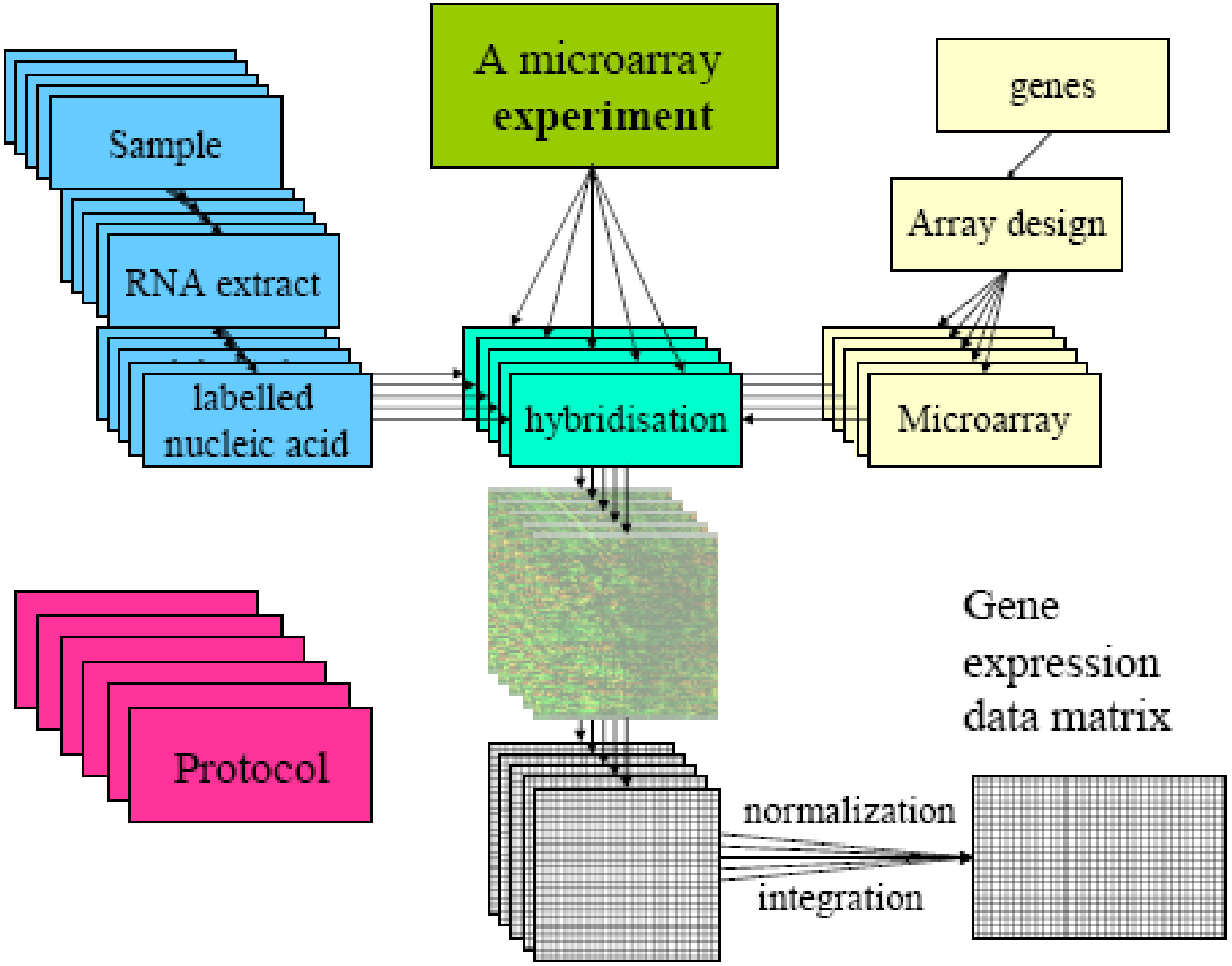
The analysis process

Limitations

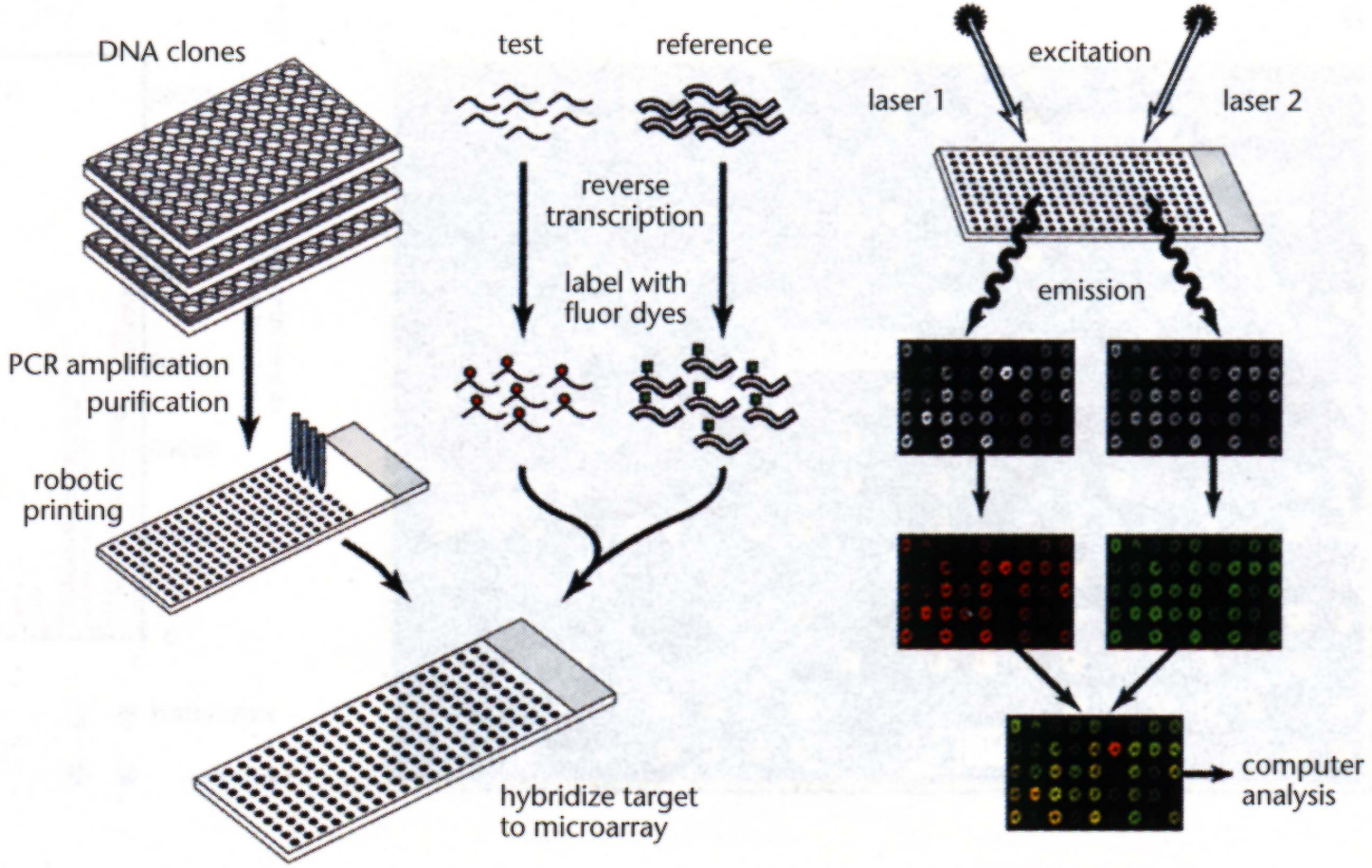
Extensions (NGS)

Outline

- An overview (a review) of microarrays
- Experiments with microarrays
- The data analysis process
- Microarray limitations
- From microarrays to Next generation sequencing

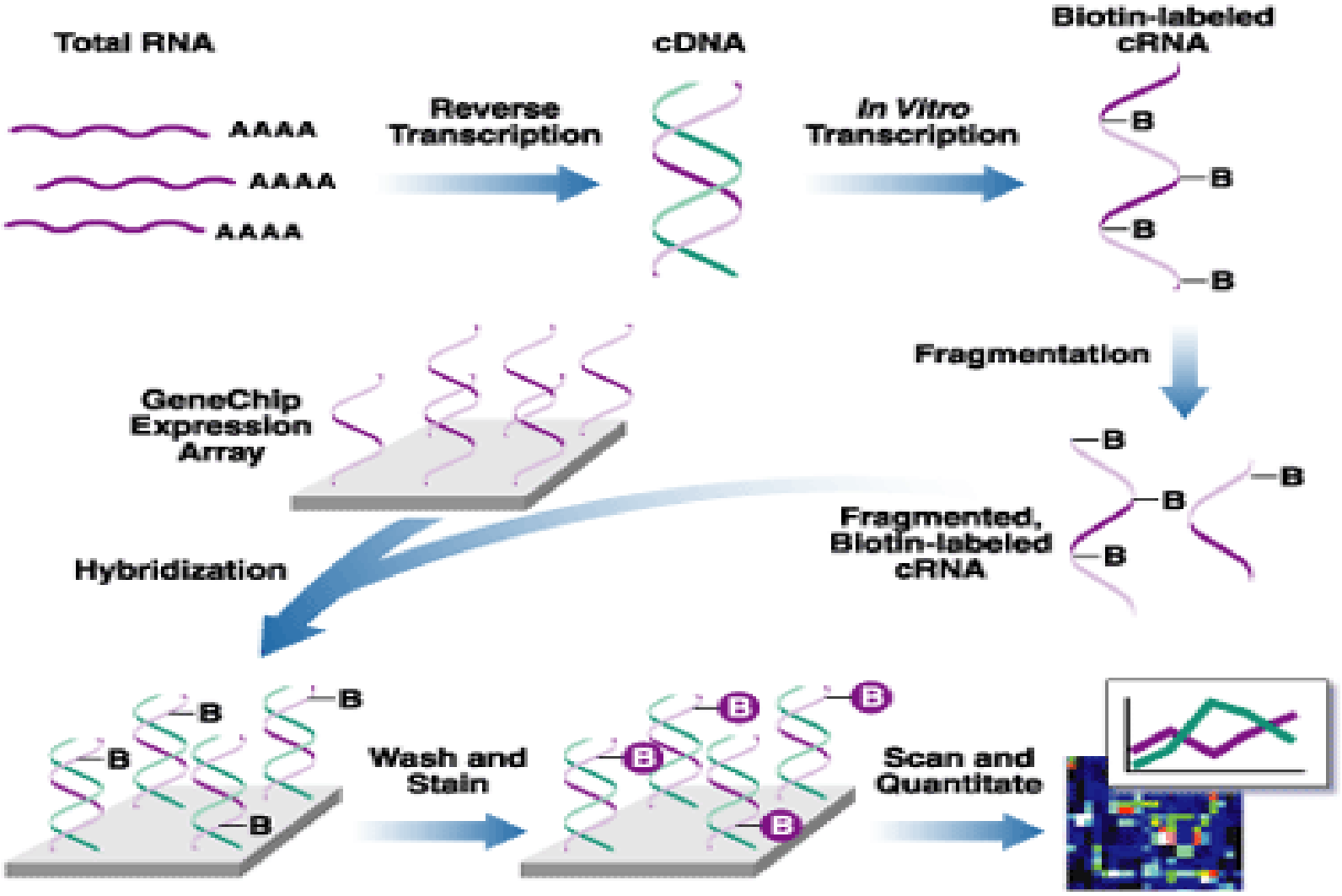

A microarray
experiment
genes
Array design
Sample
RNA extract
labelled
nucleic acid
hybridisation
Microarray
Protocol
Gene
expression
data matrix
normalization
integration

cDNA microarrays: Overview



To visualize an animation go to:
http://www.bio.davidson.edu/courses/genomics/chip/chip.html

Oligo (Affy) microarrays: overview



@Affymetrix

Comparison between two types

cDNA Microarrays

ADVANTAGES

Cheaper (not anymore)
Flexibility (customizable)
High signal intensity (long sequences)

DISADVANTAGES

Lower reproducibility
Cross-hybridization (low specificity)
Need more manual handling (possibility of contamination)

Oligonucleotide Microarrays

ADVANTAGES

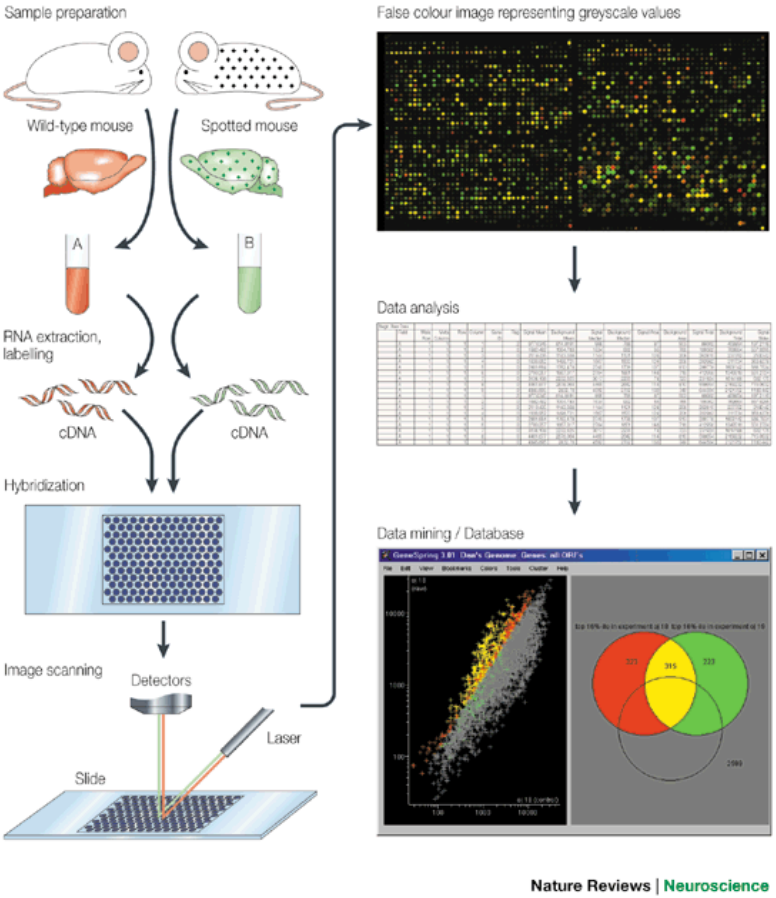
Quick and robotic manufacturing
Higher Reproducibility
High specificity (short sequences)
Use many probes / gene

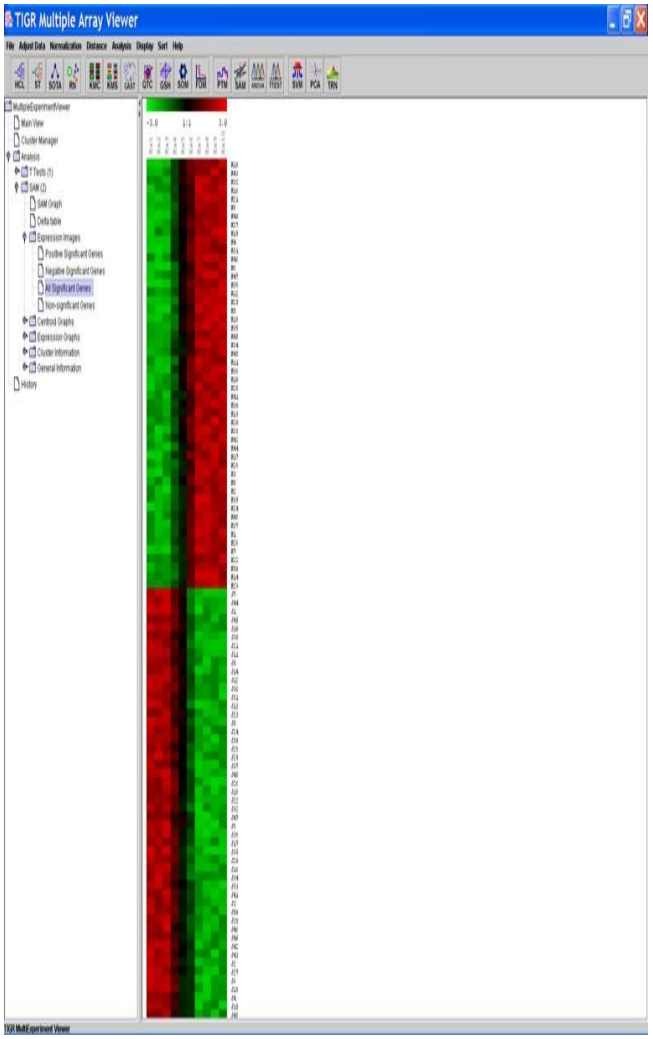
DISADVANTAGES

Requires more specialized equipment
Expensive
Less flexible (genes on the chip cannot be selected)

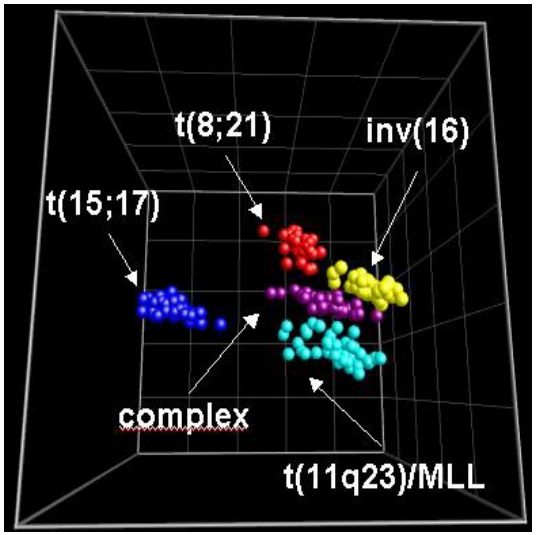
Experiments with microarrays

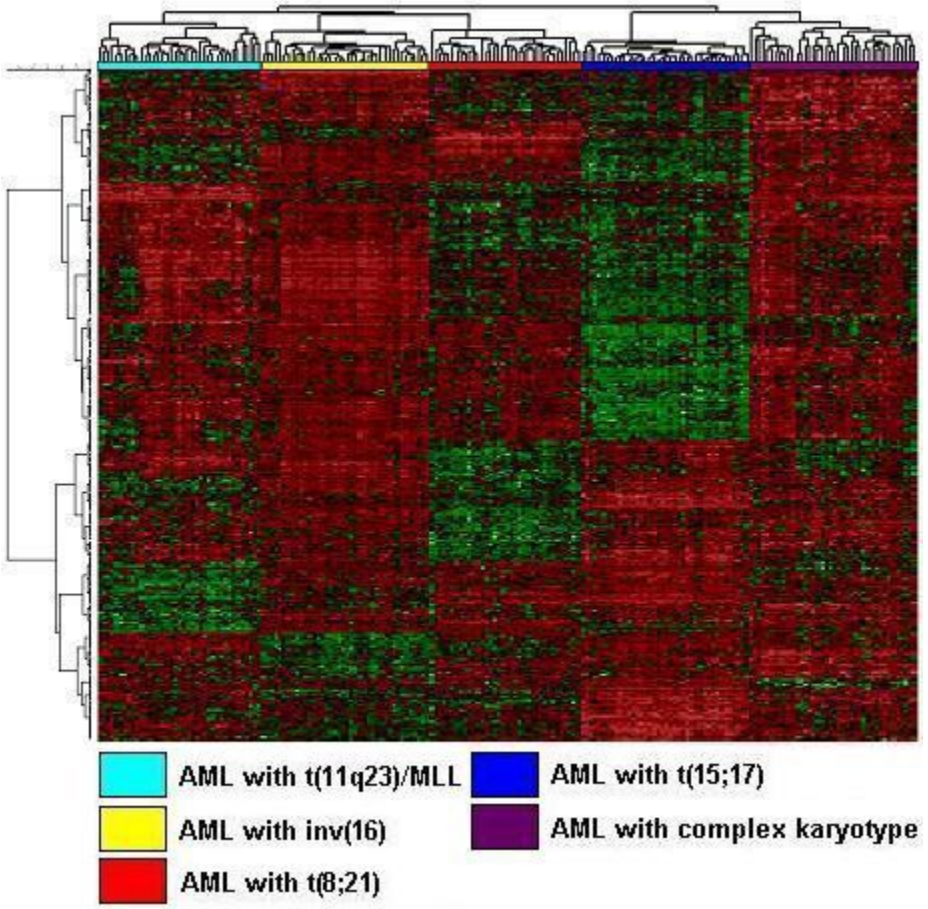
Types of studies (1): *Class comparison*



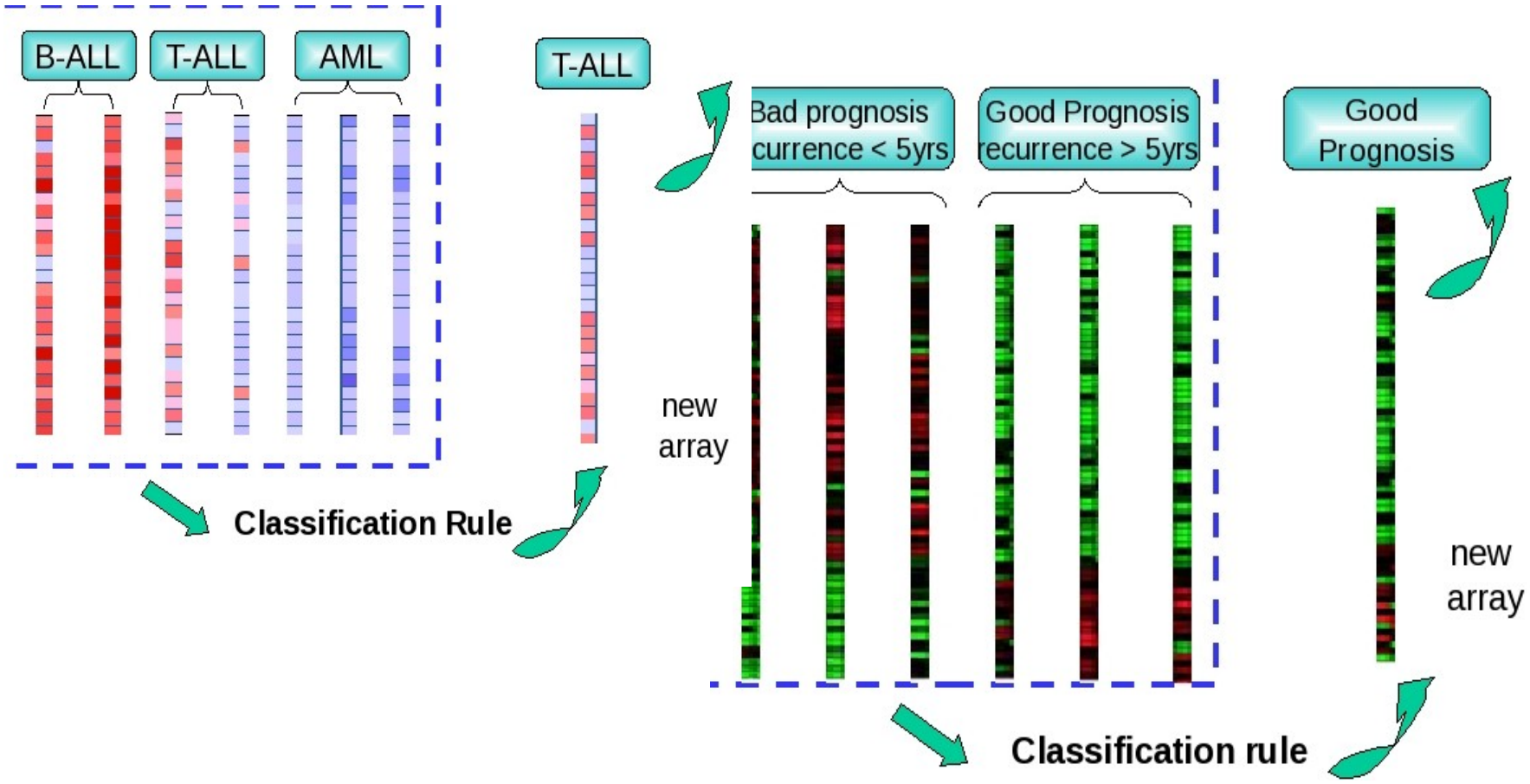


Types of studies (2): *Class discovery*

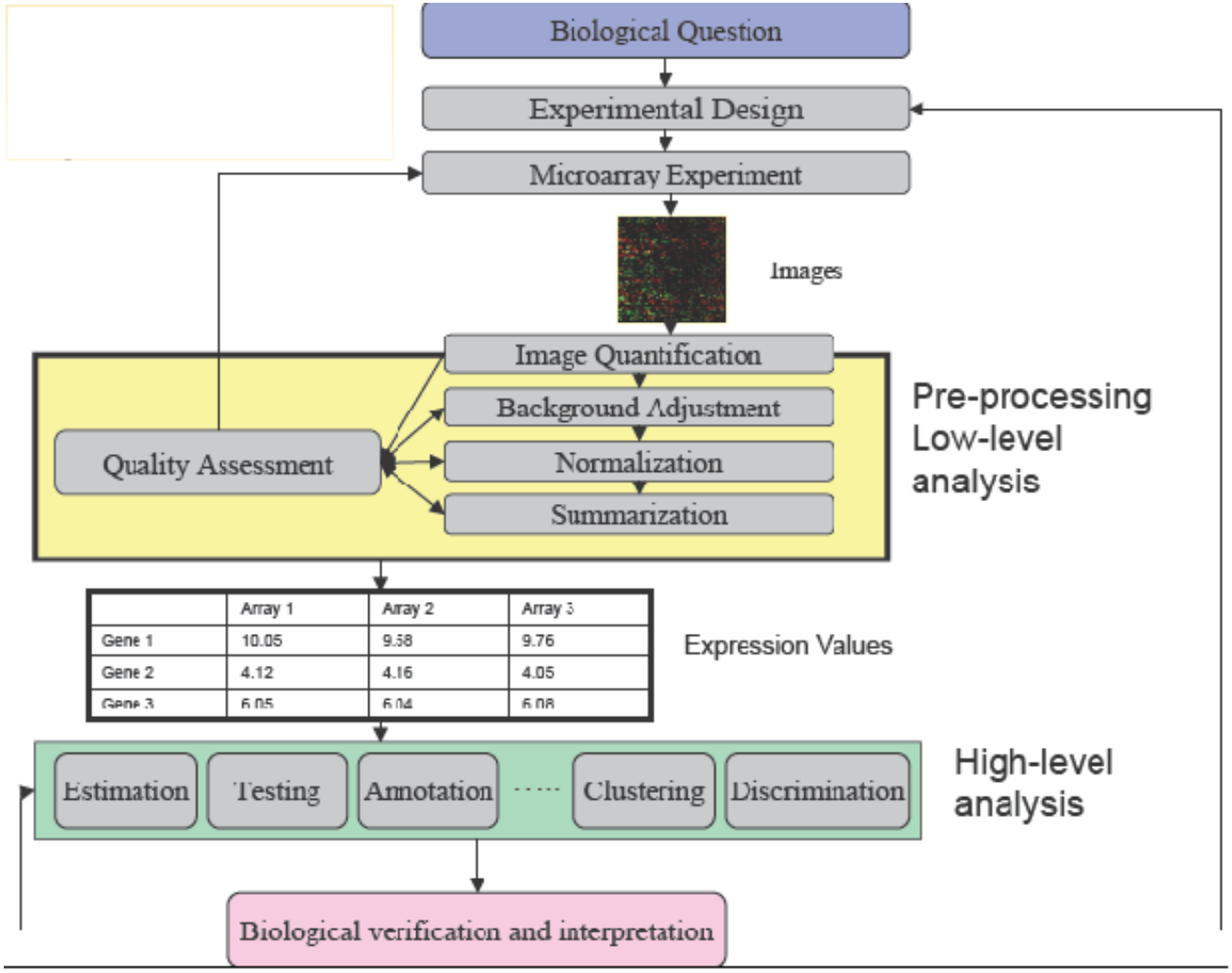




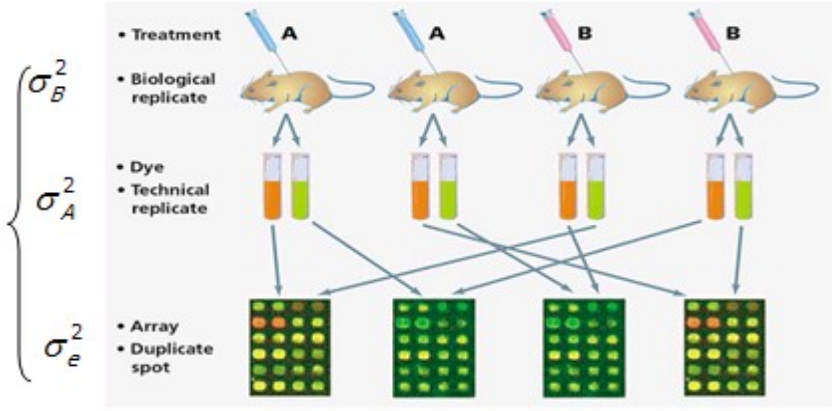
Types of studies (3): *Class prediction*



The microarray analysis process



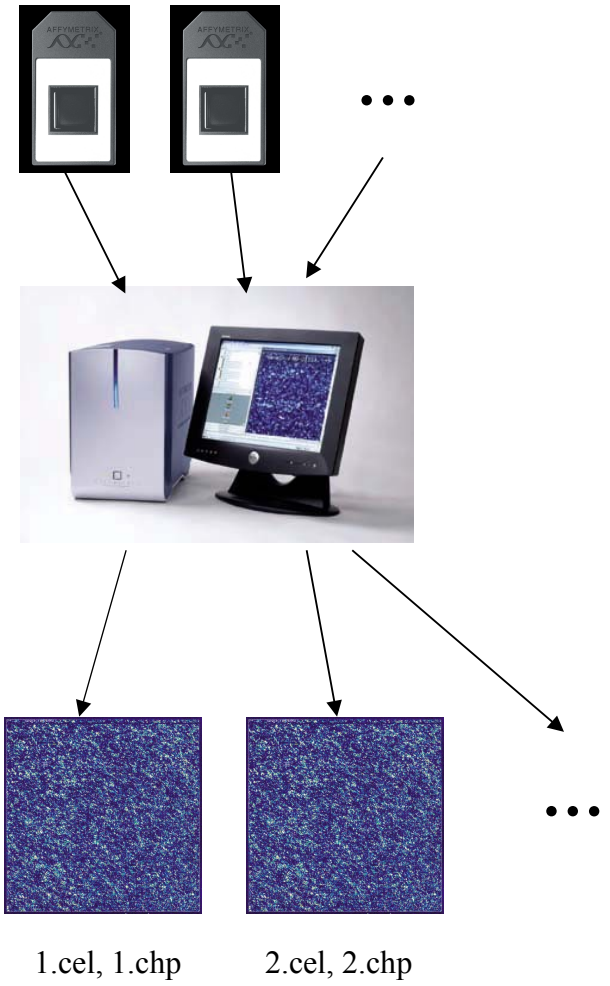
(0) Experimental design



	Awful design :-(				Balanced design :-)		
Sample	Treatment	Sex	Batch	Sample	Treatment	Sex	Batch
1	A	Male	1	1	A	Male	1
2	A	Male	1	2	A	Female	2
3	A	Male	1	3	A	Male	2
4	A	Male	1	4	A	Female	1
5	B	Female	2	5	B	Male	2
6	B	Female	2	6	B	Female	1
7	B	Female	2	7	B	Male	1
8	B	Female	2	8	B	Female	2

- Variability
  - Sistematic
    - Calibrate/Normalize
  - Random
    - Experimental design
    - Statistical inference
- Must decide about:
  - Replicates
  - Batches (“Batch effect”)
  - Pools …

(1) Image obtention



Input: Microarrays

Output:

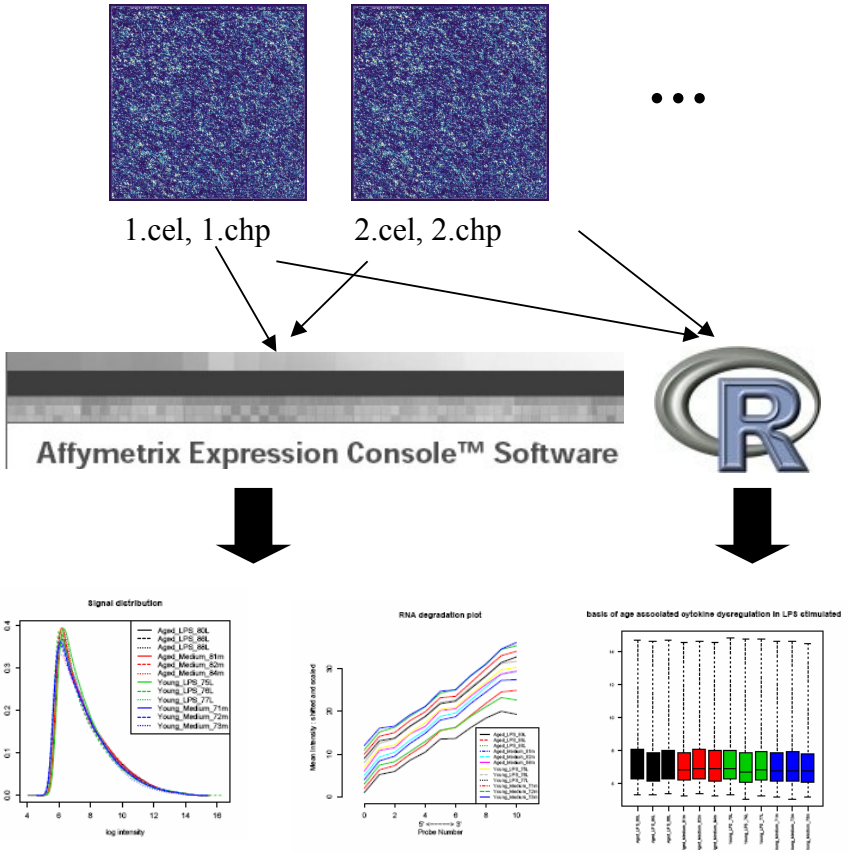
- Images (1/chip)

Information for each individual probe

Data to be used for

- Quality control
- Preprocess
- Summarization

(2) Quality control



Input:

- Images (.CEL, ...)

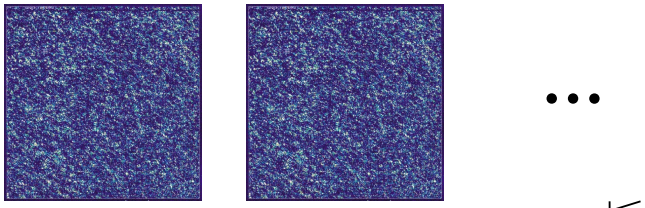
Process

- Diagnostsics
- Quality checks

Output:

- Plots
- Quality indexes

(3) Preprocess



BRB-ArrayTools

Partek®
turning data into discovery®

GEPAS

	C01-001.CEL	C02-001.CEL	C03-001.CEL
1415670_at	8.954387	9.088924	8.833863
1415671_at	10.700876	10.639307	10.610953
1415672_at	10.377266	10.510106	10.461701
1415673_at	7.320335	7.252635	7.112313
1415674_a_at	8.381129	8.332256	8.393718
1415675_at	8.120937	8.082713	8.051514
1415676_a_at	10.322229	10.287371	10.282812
1415677_at	9.038344	8.979641	8.905711

Input:

- Images

Process

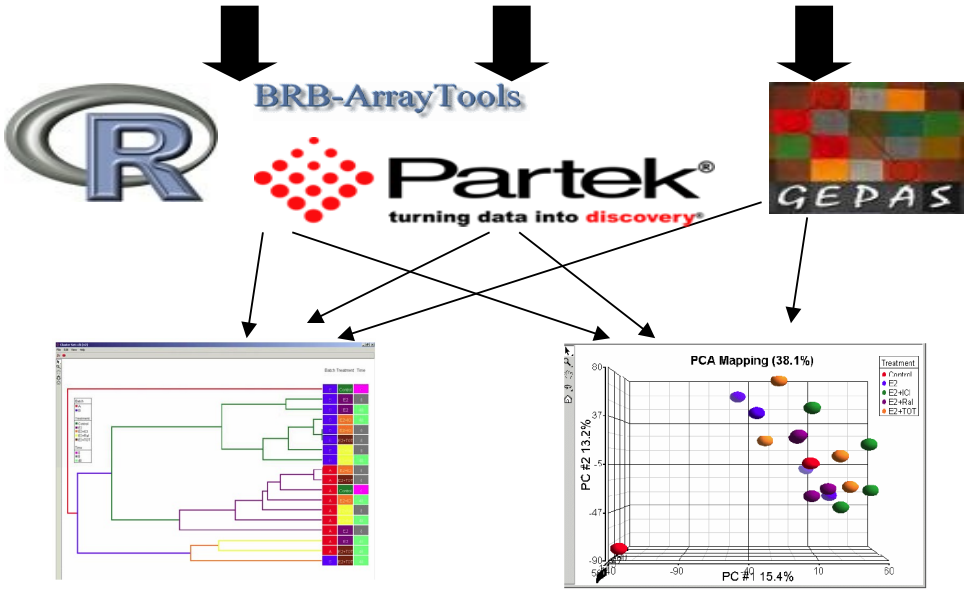
- Noise filtering
- Normalization
- Summarization
- Filtering

Output:

- Expression marix

(4) Exploration

	C01-001.CEL	C02-001.CEL	C03-001.CEL
1415670_at	8.954387	9.088924	8.833863
1415671_at	10.700876	10.639307	10.610953
1415672_at	10.377266	10.510106	10.461701
1415673_at	7.320335	7.252635	7.112313
1415674_a_at	8.381129	8.332256	8.393718
1415675_at	8.120937	8.082713	8.051514
1415676_a_at	10.322229	10.287371	10.282812
1415677_at	9.038344	8.979641	8.905711



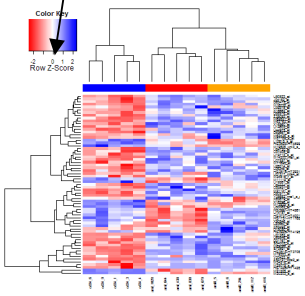
- Input
  - Expression matrix
- Process
  - PCA, Cluster, MDS
  - 2D/3D plots
- Output
  - Visualizations
  - Possible unexplained groupings (e..g. batch effect) detected

(5) Statistical analysis (i):
Selecting differentially expressed genes

	C01-001.CEL	C02-001.CEL	C03-001.CEL
1415670_at	8.954387	9.088924	8.833863
1415671_at	10.700876	10.639307	10.610953
1415672_at	10.377266	10.510106	10.461701
1415673_at	7.320335	7.252635	7.112313
1415674_a_at	8.381129	8.332256	8.393718
1415675_at	8.120937	8.082713	8.051514
1415676_a_at	10.322229	10.287371	10.282812
1415677_at	9.038344	8.979641	8.905711

BRB-ArrayTools

ProbeSet	gene	ID	logFC	t	P.Value	adj.P.Val	B
1450826_a_at	Saa3	1450826_a_at	4.911	63.544	6.21E-14	2.80E-10	22.244
1457644_s_at	Cxcl1	1457644_s_at	4.286	53.015	3.52E-13	7.69E-10	20.791
1415904_at	Lpl	1415904_at	-4.132	-50.455	5.66E-13	7.69E-10	20.373
1449450_at	Ptges	1449450_at	5.164	49.483	6.82E-13	7.69E-10	20.207
1419209_at	Cxcl1	1419209_at	5.037	47.175	1.08E-12	9.71E-10	19.794
1416576_at	Socs3	1416576_at	3.372	42.107	3.19E-12	2.08E-09	18.784
1450330_at	Il10	1450330_at	4.519	42.056	3.23E-12	2.08E-09	18.773
1455899_x_at	Socs3	1455899_x_at	3.648	40.821	4.29E-12	2.12E-09	18.502
1419681_a_at	Prok2	1419681_a_at	3.709	40.645	4.48E-12	2.12E-09	18.463
1436555_at	Slc7a2	1436555_at	3.724	40.081	5.12E-12	2.12E-09	18.335



Input:

- Expression matrix
- Analysis models

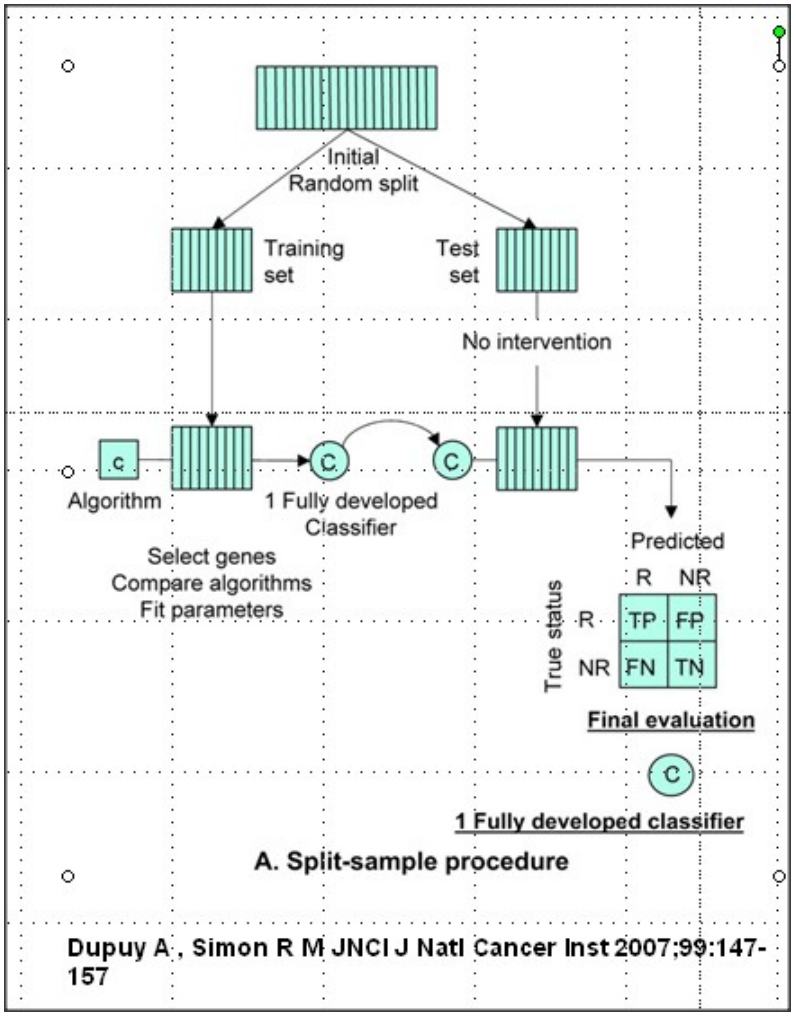
Process

- t-tests, ANOVA
  - P-value adjustment

Output

- Gene lists
  - Fold change, p.values
- Plots
- Expression profiles

(5) Statistical analysis (ii):
Building and validating a predictor



Input:

- Expression matrix

Process

- Variable selection
- Model fitting
- Validation

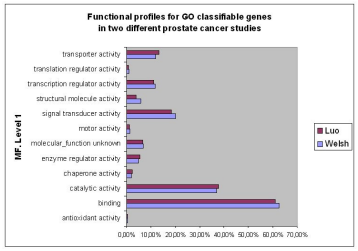
Output

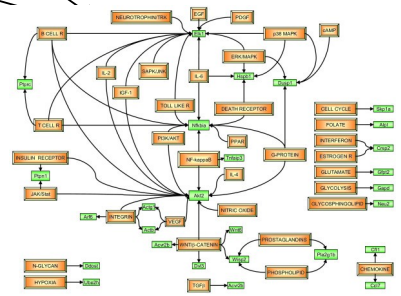
- Predictive models
- Measures of sensitivity and reproducibility

(6) Biological significance

ProbeSet	gene	ID	logFC
1450826_a_at	Saa3	1450826_a_at	4.911
1457644_s_at	Cxcl1	1457644_s_at	4.286
1415904_at	Lpl	1415904_at	-4.132
1449450_at	Ptges	1449450_at	5.164
1419209_at	Cxcl1	1419209_at	5.037
1416576_at	Socs3	1416576_at	3.372
1450330_at	Il10	1450330_at	4.519
1455899_x_at	Socs3	1455899_x_at	3.648
1419681_a_at	Prok2	1419681_a_at	3.709
1436555_at	Slc7a2	1436555_at	3.724

BRB-ArrayTools





Input

- Gene lists

Process

- GEA, GSEA,
- Network analysis

Output:

- Relevant GO or KEGG terms
- Relevant pathways
- Networks

Microarray limitations

All that glitters is not gold

THE LANCET

Search for in All Fields GO Advance

Home | Journals | Collections | Audio | Conferences | Education | Resource Centres | For Authors

The Lancet, Volume 365, Issue 9458, Pages 454 - 455, 5 February 2005
doi:10.1016/S0140-6736(05)17878-7 Cite or Link Using DOI

< Previous Article | Next Article >

Microarrays and molecular research: noise discovery?

John PA Ioannidis a b c

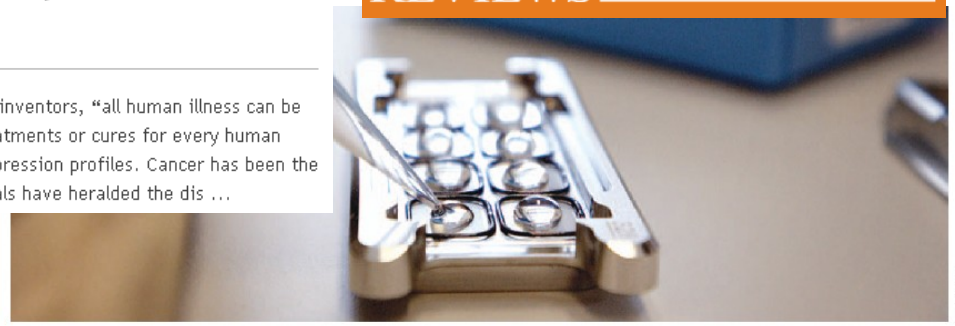
The promise of microarrays has been of apocalyptic dimensions. As put forth by one of their inventors, "all human illness can be studied by microarray analysis, and the ultimate goal of this work is to develop effective treatments or cures for every human disease by 2050". 1 All diseases are to be redefined, all human suffering reduced to gene-expression profiles. Cancer has been the most common early target of this revolution 2 and publications in the most prestigious journals have heralded the dis ...

PLoS MEDICINE

a peer-reviewed open-access journal published by the Public Library of Sci

Home Browse Articles About For Readers For Authors and

nature REVIEWS DRUG DISCOVERY



An array of problems

Despite the huge amount of published microarray data in cancer, little is being converted into clinical practice. Validating initial data is proving to be a key challenge, reports SIMON FRANTZ.

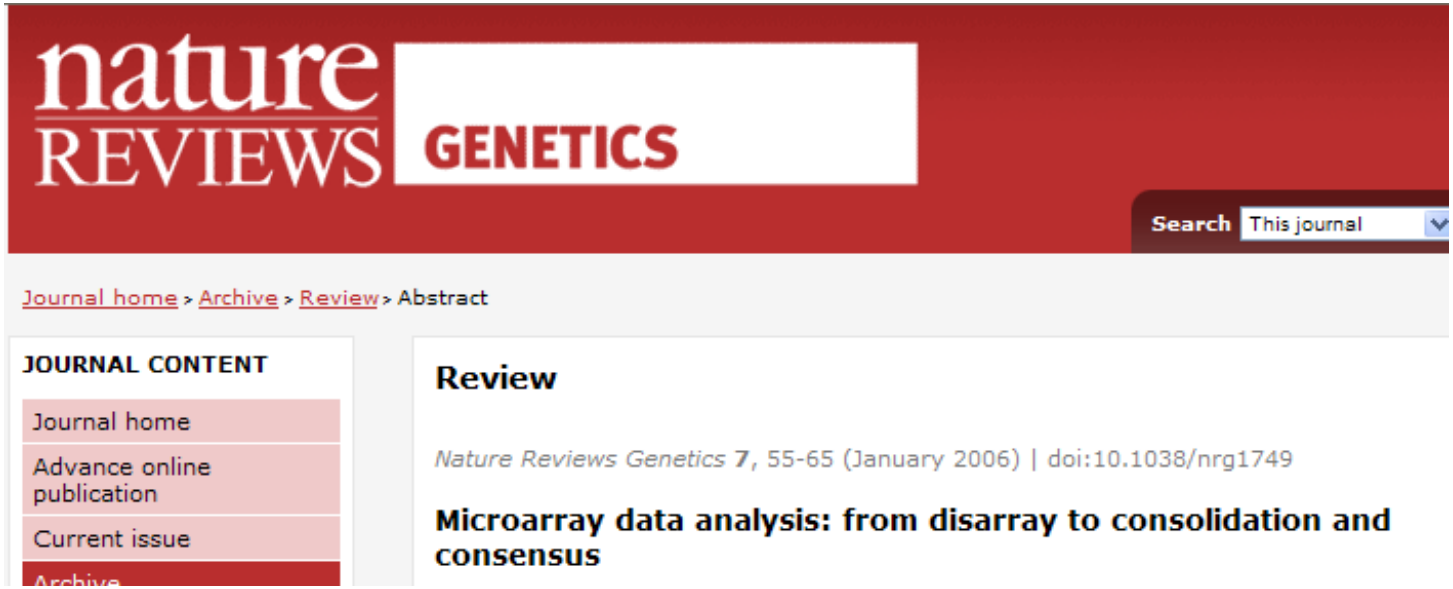
EDITORIAL OPEN

Why Bigger Is Not Yet Better: The Problems with Huge Datasets

An array of problems?

- By the middle of the decade some claims against microarrays were raised.
  - Lack of reproducibility between studies.
  - Few coincidences between gene lists.
  - Predictions on new test data did not reproduce those in training data.
  - …
  - The step to the clinic always waiting.

Que no estamos tan mal...



Critical Review of Published Microarray Studies for Cancer Outcome and Guidelines on Statistical Analysis and Reporting

Alain Dupuy, Richard M. Simon

So what?

- Sistematic reviews showed that the main problem was not the technology but unappropriate application of (statistical) methodology.

- Large quality control studies (MAQC) were promoted to investigate reliability and applicability of the technique.

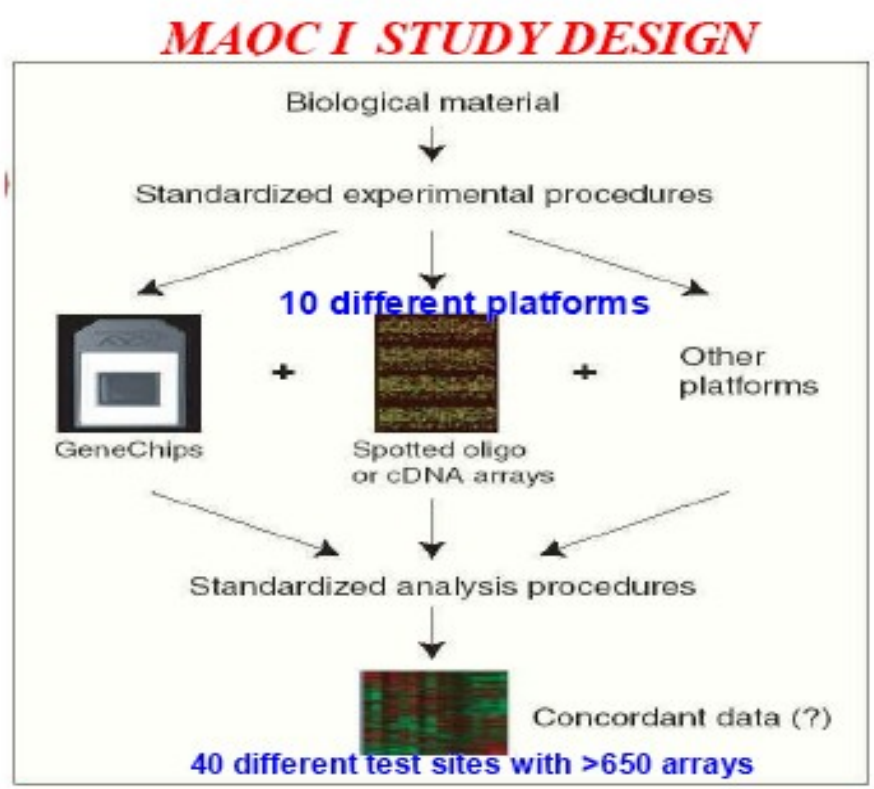
Some consensus (Allison 2006)

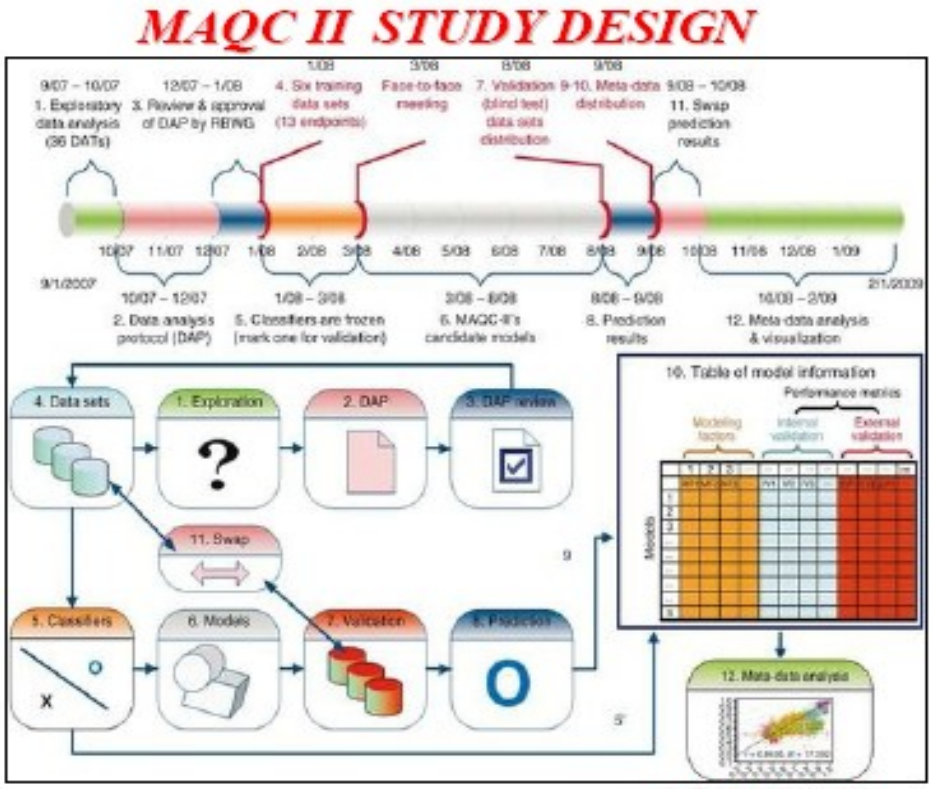
- Design
  - Biological replication is essential
  - There is strength in numbers: power & sample size
  - Pooling biological samples can be useful

- Selecting diferentially expressed genes
  - Using FC alone as a differential expression test is not valid
  - 'Shrinkage' is a good thing
  - **FDR is a good alternative to conventional multiple-testing approaches**

- Classification and Prediction
  - Unsupervised classification is overused
  - Unsupervised classification should be validated using resampling-
  - Supervised-classification requires independent cross-validation

MicroArray Quality Control (MAQC)





ARTICLES
nature biotechnology

The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements

MAQC Consortium*

Sept. 2006

nature biotechnology
ARTICLES

The MicroArray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models

MAQC Consortium*

August 2010

In summary

- Microarrays have some reasonable limitations.
  - Some, such as noise or restriction to known sequences, are intrinsec and cannot be removed.
  - Other issues can be solved using appropriate data analysis methods.
- Studies suggest that, ***if well used***, it is a reliable technique that yields reliable reproducible results.
- Statistics has made important contributions and has benefited from the problems raised by microarrays.
- ***However*** the step from beadside to clinic is taking much longer than expected.

Is microarray time over?

InSightu

a nature network blog by Anthony Fejes

HOME | ARCHIVE | ABOUT | POLICY

Announcing the death of the Micro-array

Bookmark in Connotea

Ok, here it is: Micro-arrays are dead. I
the fish head someone tossed into the co
compost won't be picked up till this Tue
waiting to be done.

nature methods

Techniques for life scientists and chemists

Access

To read this story in full you will need to login or make a payme

nature.com > Journal home > Table of Contents

NEWS AND VIEWS

Nature Methods **5**, 585 - 587 (2008)
doi:10.1038/nmeth0708-585

The beginning of the end for microarrays?

Jay Shendure[1]

1. Jay Shendure is in the Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA.
e-mail: shendure@u.washington.edu

Two complementary approaches, both using next-generation sequencing, have successfully tackled the scale and the complexity of mammalian transcriptomes, at once revealing unprecedented detail and allowing better quantification.

ARTICLE LINKS

- Figures and tables

SEE ALSO

- Article by Cloonan *ET AL.*
- Article by Mortazavi *ET AL.*

ARTICLE TOOLS

- Send to a friend
- Export citation

nature news

nature news home | news archive | specials | opinion | features | news blog

comments on this story

Published online 15 October 2008 | *Nature* **455**, 847 (2008) | doi:10.1038/455847a

News

The death of microarrays?

High-throughput gene sequencing seems to be stealing a march on microarrays. Heidi Ledford looks at a genome technology facing intense competition.

Heidi Ledford

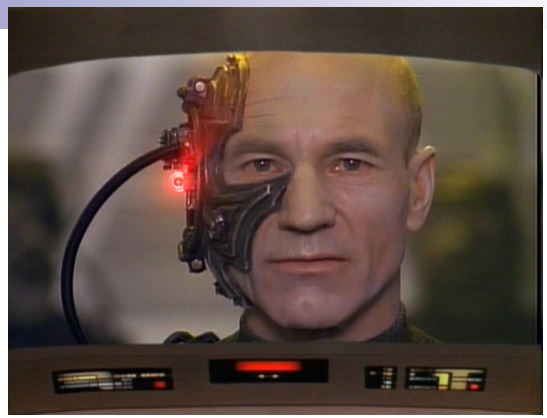
Stories by subject

- Genetics
- Business
- Biotechnology

Stories by keywords

- Gene chips
- Next-generation sequencing
- Genomics

Faster, cheaper DNA sequencing technology is revolutionizing the burgeoning field of personal



Next generation sequencing

The future is here, now

Next generation Sequencing

- By the middle decade new technologies consolidated allowing the massive production of tens of millions of short sequencing fragments.
- These techniques could be used to
  - Deal with similar problems than microarrays,
  - But also with many other.
- “Again” they raised the promise of personalized medicine..

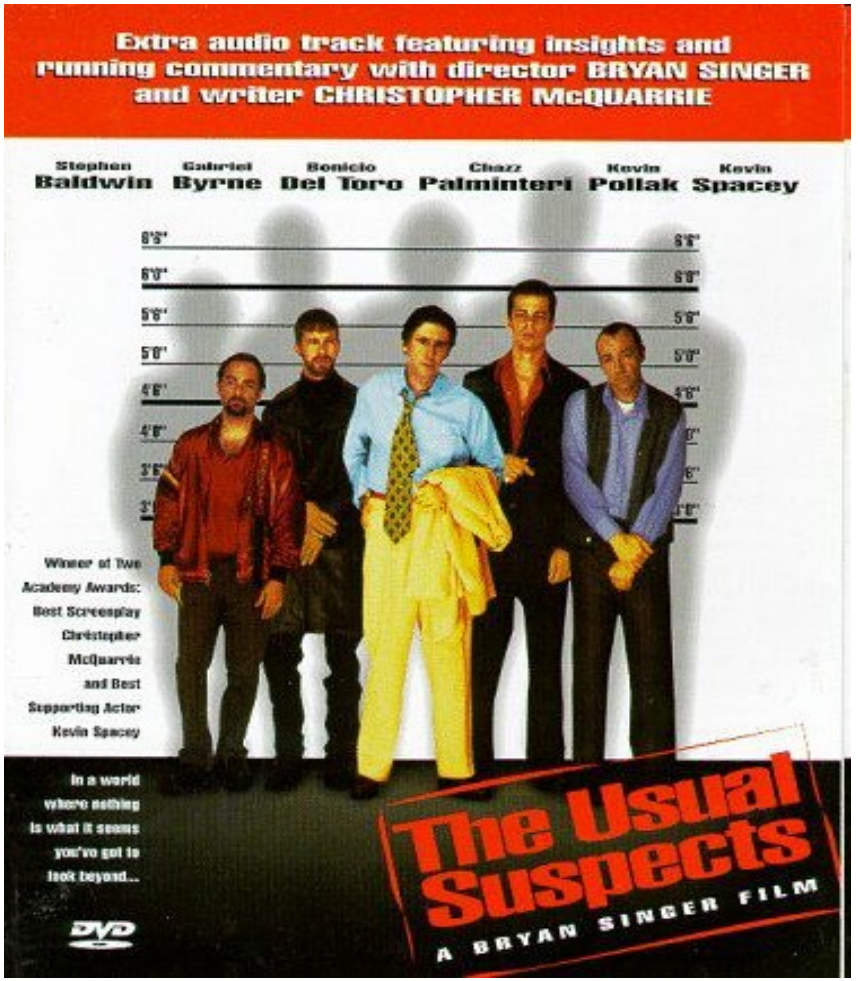
A CSI approach to gene selection (again)

- A crime has been commited (immune response)
- You're CSI –Horatio Fisher- and want to find who's responsible for this.
- Let's see how you would act …
  - In the old times,
  - In the microarray age,
  - In the next generation age.

In the old times …

- You would chase the “Usual Suspects” and make an in deep interrogation.
  - If guilty you might make them talk,
  - But if not you might miss the bad guy.

- *That is looking at specific genes may yield great or awful results.*



In the microarray age…

- You have the census of most people and their fingerprints.
  - If you find a fingerprint in your database that is clean enough you may find the bad guy.
    - What about bad prints?
    - What about those who are not censed.
    - And those no-fingerprints?
- *That is you may look at all known genes but you*
  - *do it Indirectly and noisly*
  - *miss genes/forms that are uncensed.*

Why is sequencing different?

- If the crime scene had had cameras you would have directly known who the criminal was.

- *Sequencing allows you to access **everything***
  - *Known and unknown forms are sequenced.*
  - *The technique is less noisy and the resolution higher.*

To be continued ...