The microarray data analysis process





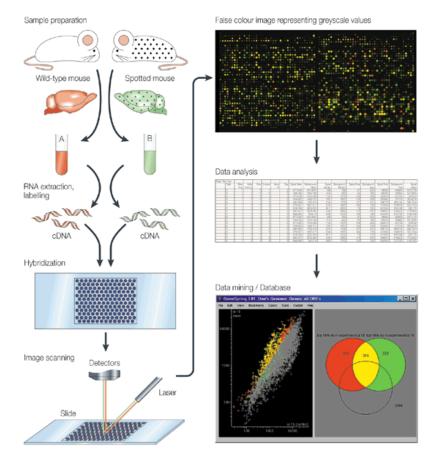
Goals

- Overview
 - □ Types of studies
 - □ The basic "pipe-line"
- Get familiar with the process
 - □ Input/Output for each step,
 - □ Problems and solutions
 - Interpretations
- Examples.





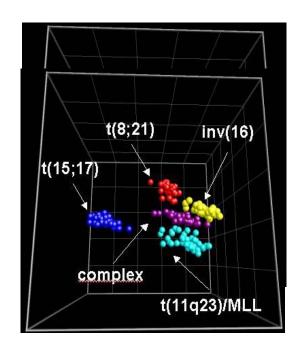
(1): Class comparison

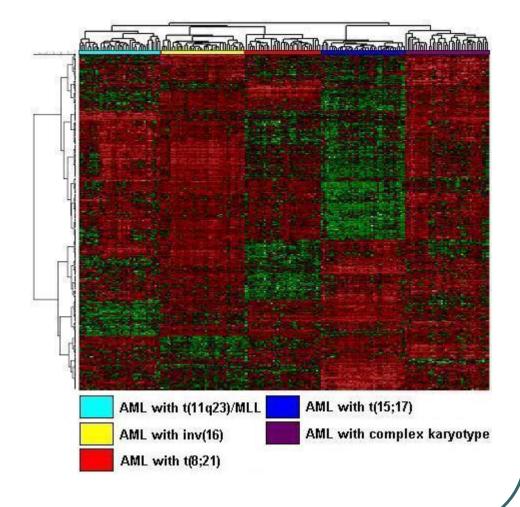






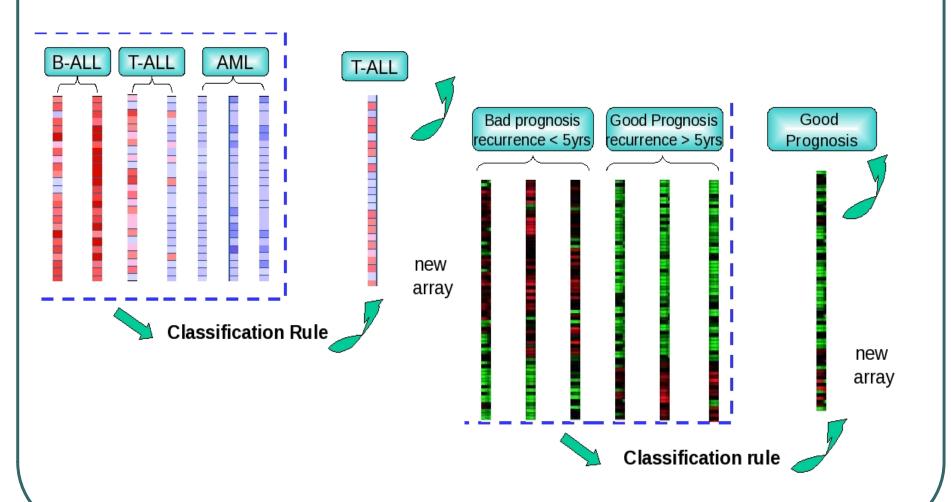
(2): Class discovery







(3): Class prediction





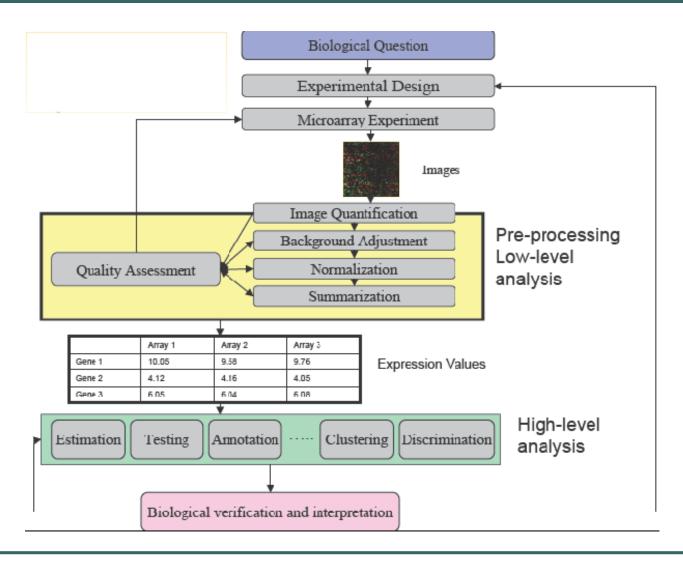


And many, many more ...

- Time Course
- Pathway Analysis-(Systems Biology)
- Whole Genome,
- CGH,
- Alternative Splicing



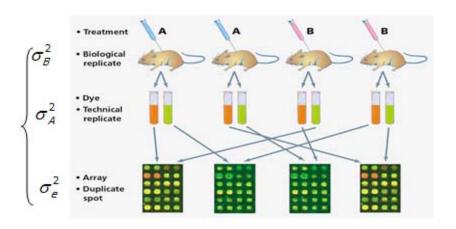
The microarray analysis process







(0) Experimental design



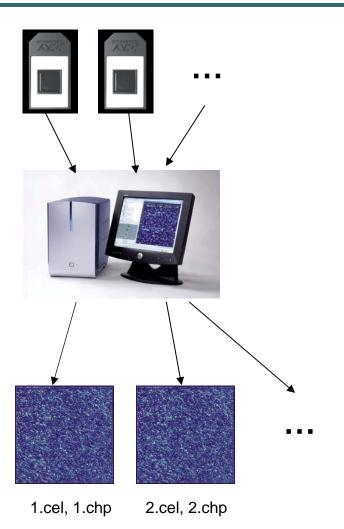
	Awful design :-(Balanced design :-)		า :-)
Sample	Treatment	Sex	Batch	Sample	Treatment	Sex	Batch
1	Α	Male	1	1	Α	Male	1
2	Α	Male	1	2	Α	Female	2
3	Α	Male	1	3	Α	Male	2
4	Α	Male	1	4	Α	Female	1
5	В	Female	2	5	В	Male	2
6	В	Female	2	6	В	Female	1
7	В	Female	2	7	В	Male	1
8	В	Female	2	8	В	Female	2

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

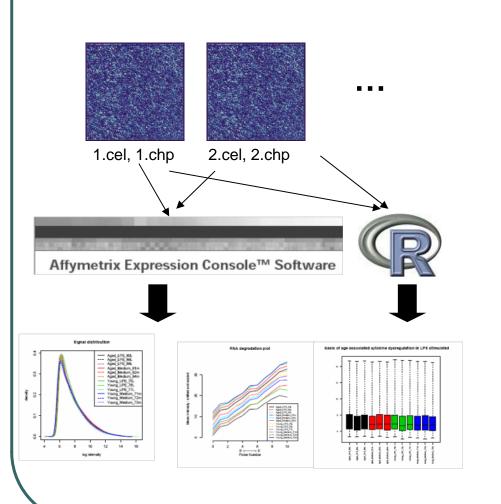
Quality control

Preprocess

Summarization



(2) Quality control



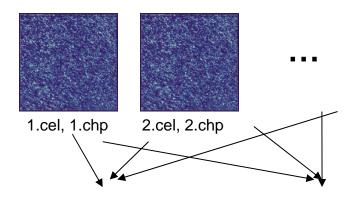
Input Images (.CEL, ...)

Process
Diagnostsics
Quality checks

Output
Plots
Quality indexes



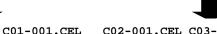
(3) Preprocess











	COT-OOT.CED	C02-001.C	.uu coo-
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

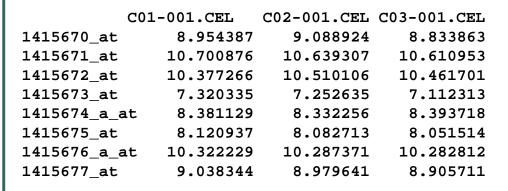
Output

Expression marix





(4) Exploration



Input

Expression matrix

Process

PCA, Cluster, MDS 2D/3D plots

BRB-Array Tools turning data into discovery PCA Mapping (38.1%) PCA M

Output

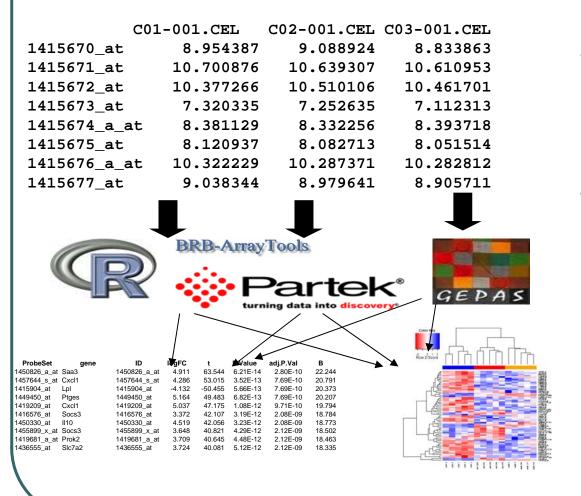
Visualizations

Possible unexplained groupings (e..g. batch effect) detected





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



Input

Expression matrix Analysis models

Process

t-tests, ANOVA

P-value adjustment

Output

Gene lists

Fold change, p.values

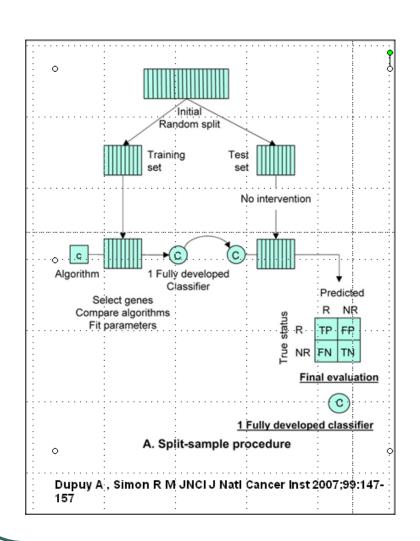
Plots

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	II10	1450330_at	4.519
1455899_x_at	Socs3	1455899_x_at	3.648
1419681_a_at	Prok2	1419681_a_at	3.709
14365 <u>55</u> at	Slc7a2	1436555_at	3. <u>724</u>
		_	

Input

Gene lists



GEA, GSEA,

Network analysis

Output

Relevant GO or KEGG terms

Relevant pathways

Networks

