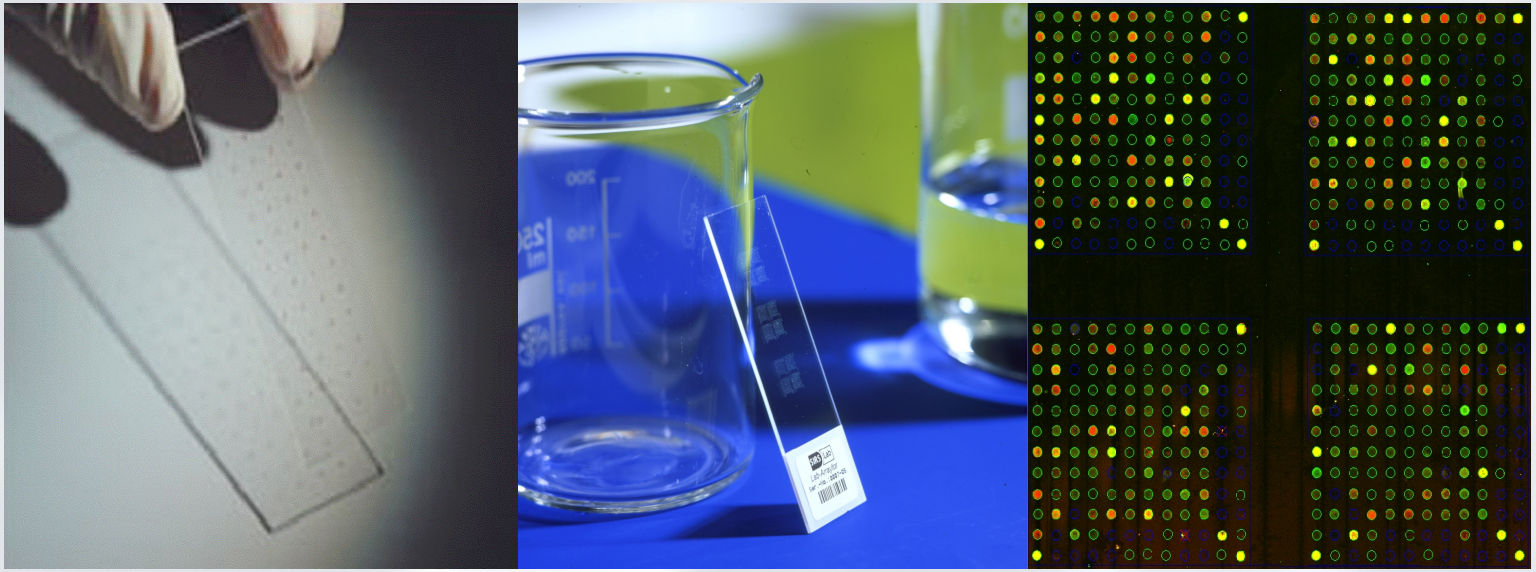
Analyse von Genexpressionsdaten bei Sepsispatienten



Dr. Matthias Kohl

SIRS-Lab GmbH

© SIRS-Lab GmbH, 2001 – 2006

SIRS Lab

SIRS-Lab GmbH

Business focus

Molecular Diagnostics

Indication

Novel diagnostic systems for sepsis and critical care medicine

Our company

Privately held research driven biotech company located in Jena, Germany, Started in April 2001

© SIRS-Lab GmbH, 2001 – 2006

SIRS-Lab's clinical network

European Society for Intensive Care and Emergency Medicine

Access to European wide epidemiologic data base on the influence of genetic factors
GenOSept Network

Clinic for Anesthesiology and Intensive Therapy, Jena

Continuously expanding the Jena Sepsis Register
Consultancy
Close scientific collaboration

SIRS Lab

Organization of 23 clinical centers in Germany
Rapid implementation of multi-center trials

German Sepsis Competence Network (SepNet)

Discussion platform
Information exchange
Providing expert opinion

German Sepsis Society (DSG)

© SIRS-Lab GmbH, 2001 – 2006

Jena Sepsis Register - patient documentation

Unique documentation:

- Currently 14,676 fully documented ICU days from 1,583 ICU patients available
- Data obtained by an independent team of residents and study nurses
- Extensive, daily documentation of clinical and lab findings (incl. CDC criteria, scores, organ dysfunction parameter, etc.) including also patient history, primary indication, ICU & hospital outcome, etc.
- Clinical data is quality & plausibility checked
- Independent ICU committee available

© SIRS-Lab GmbH, 2001 – 2006

Patients DB – data exchange

ICU Jena

Data entry using SIRS-Lab Patients DB Front-end

Anonymized Patients DB

Personal patients data

- Full access to all patients data
- Data plausibility tests

Secure unidirectional data transmission

SIRS-Lab

Samples

Patents

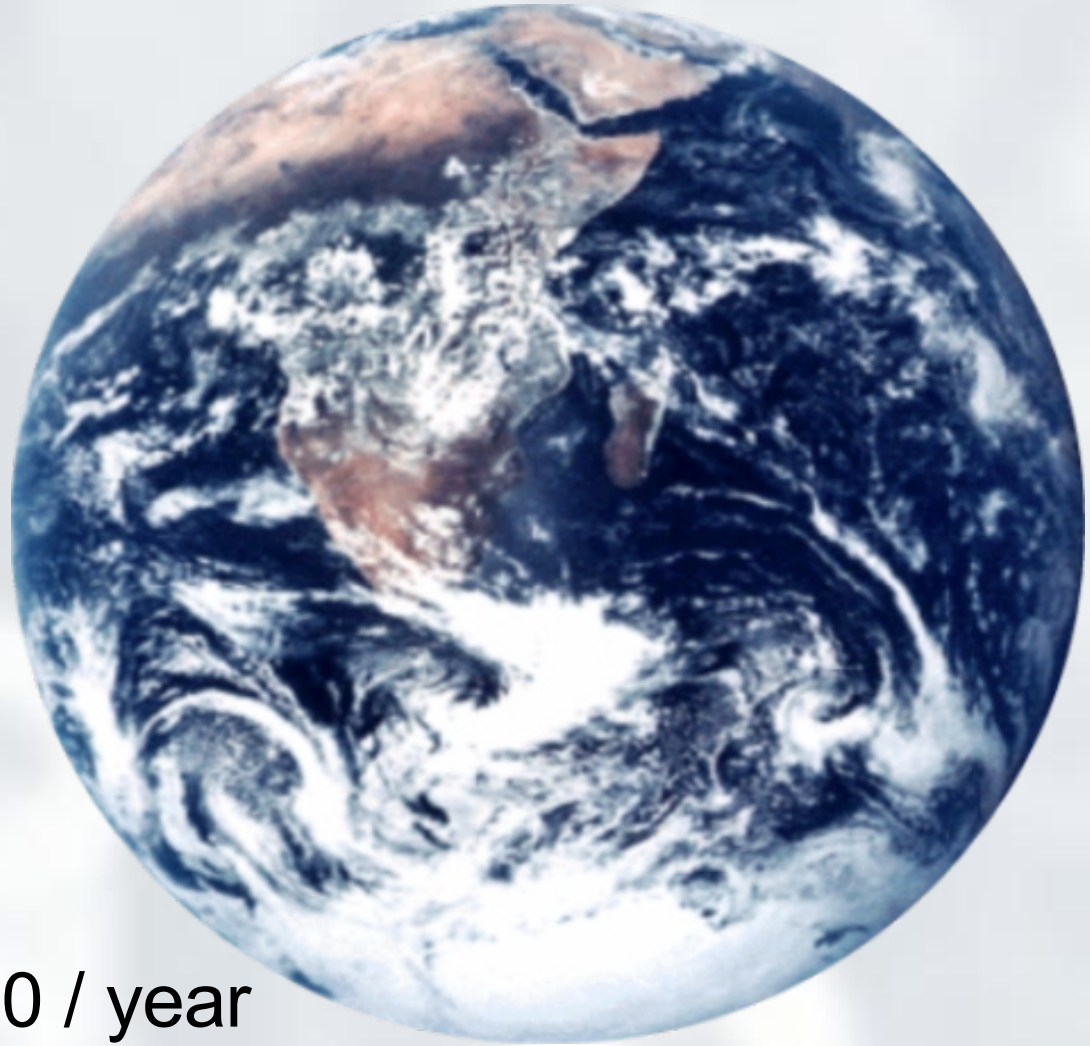
Probes

Patients

Expression

© SIRS-Lab GmbH, 2001 – 2006

Sepsis



Global:
~1,500,000 / year

Rangel-Frausto et al. JAMA 1995
Linde-Zwirble et al. Crit Care Med 1999

U.S.:
~750,000 / year

Martin GS et al. N Engl J Med 2003;348:1546-54
Angus DC et al. Crit Care Med 2001; 29: 1303-1310

Germany:
~125,000 – 300,000 / year

Rychlik R et al. Gesundh Ökon Qual Manag 2000; 5:67-72

© SIRS-Lab GmbH, 2001 – 2006

Sepsis-Incidence - Comparison with other diseases

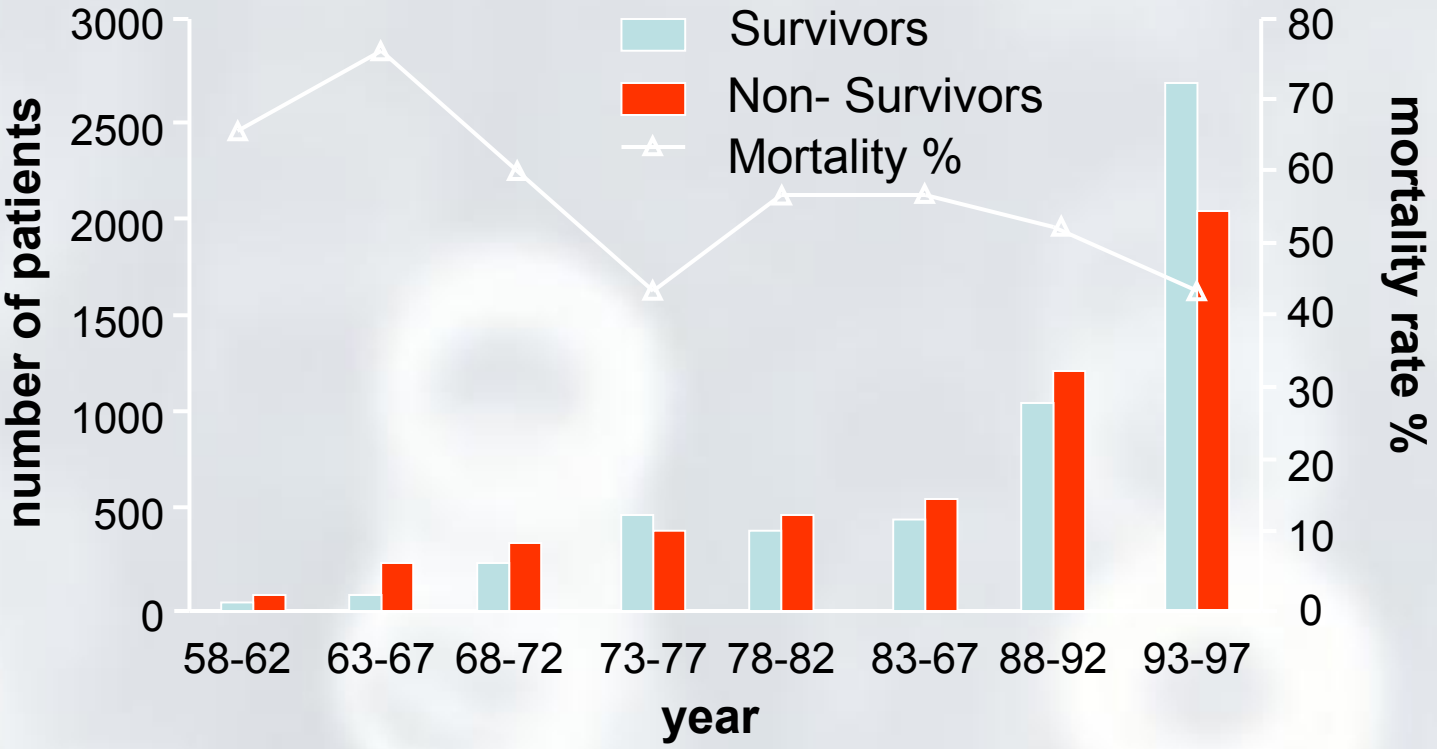
Breast cancer	**110 / 100,000 inhab.**
Colon carcinoma	**50 / 100,000 inhab.**
AIDS	**17 / 100,000 inhab.**
Severe sepsis	**300 / 100,000 inhab.**

Angus DC et al; Crit Care Med 2001; 29: 1303-1310

© SIRS-Lab GmbH, 2001 – 2006

Mortality in Sepsis Trials

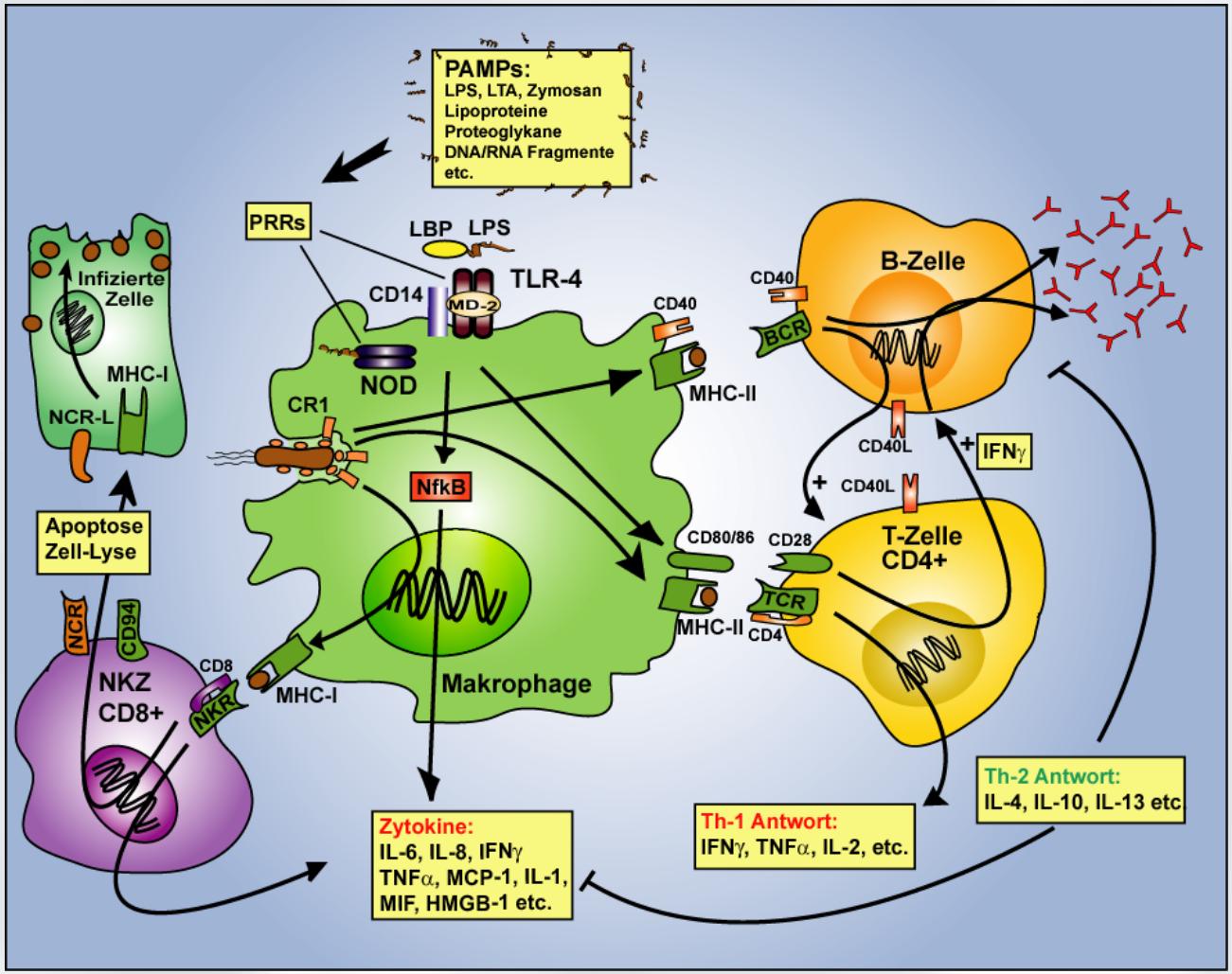
- Meta-analysis of 131 studies
- Mortality - 49.7% (range: 40 - 80%)



Friedman et al. Crit Care Med 1998;26:2078-2086

© SIRS-Lab GmbH, 2001 – 2006

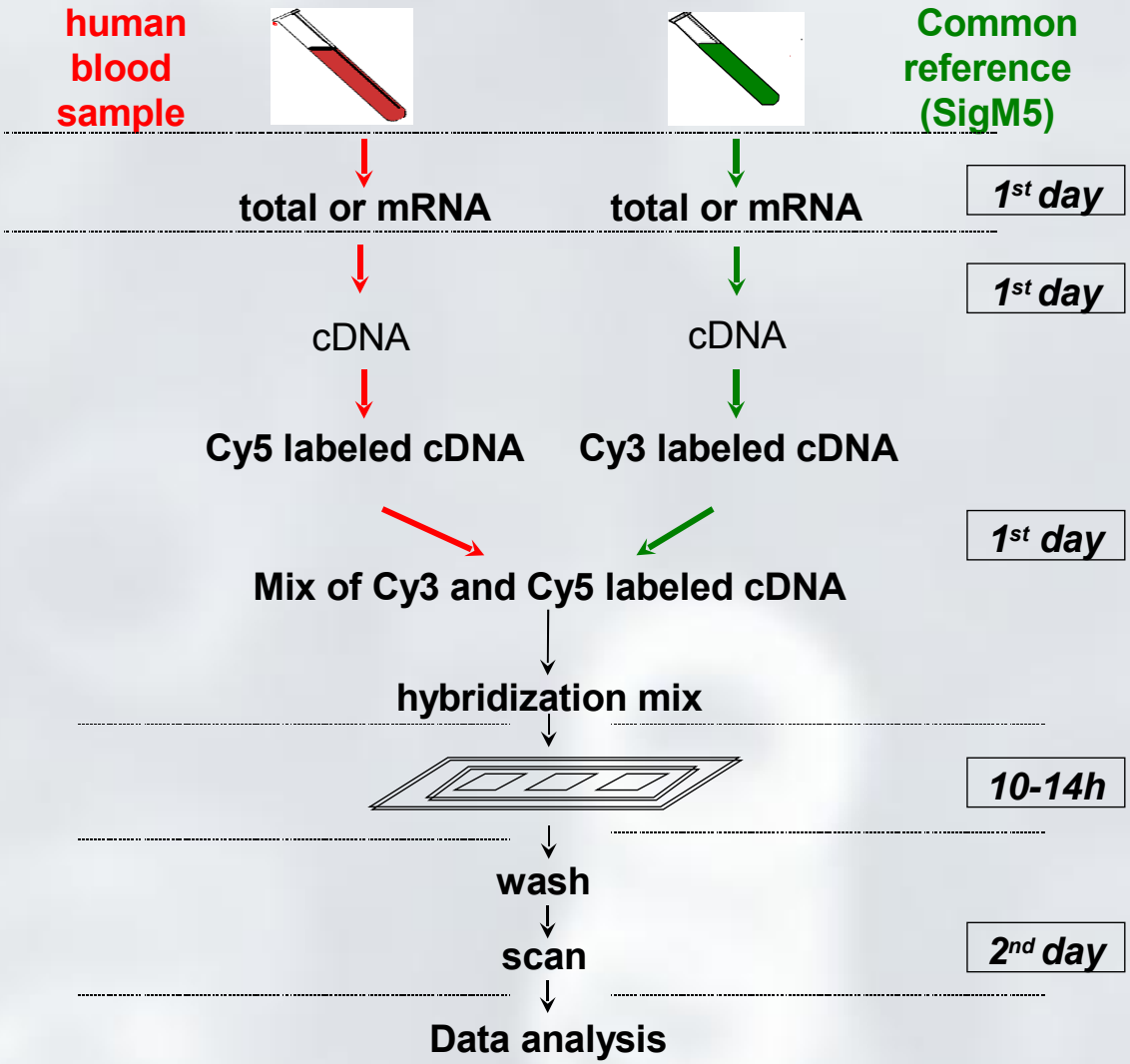
The problem: complex interacting molecular mechanisms



© SIRS-Lab GmbH, 2001 – 2006

SIRS Lab

Microarray work flow



Advantages:

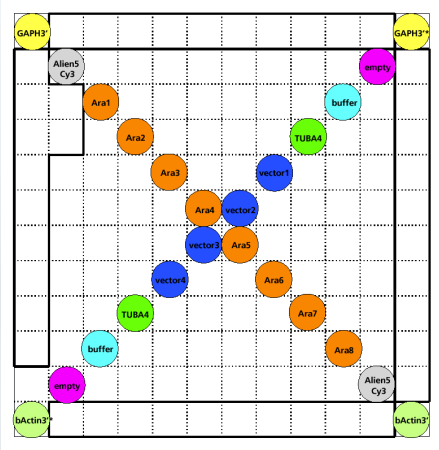
- Miniaturisation and high parallel handling
- Massive time savings in comparison with monometric tests
- Efficient utilisation of available samples and reagents
- High data output
- All together: much cheaper and rapid than the classical monometric measurements

© SIRS-Lab GmbH, 2001 – 2006

SIRS Lab

The microarray platform

- Genome-wide approach for gene selections
- Pre-screen of 74.000 human EST clones
- Screen with RNA samples from SIRS (w/o OD) and Sepsis (with OD) patients
- Selection of differentially regulated genes
- Design of a medium density microarray with 5226 80mer polynucleotide probes and additional QC features

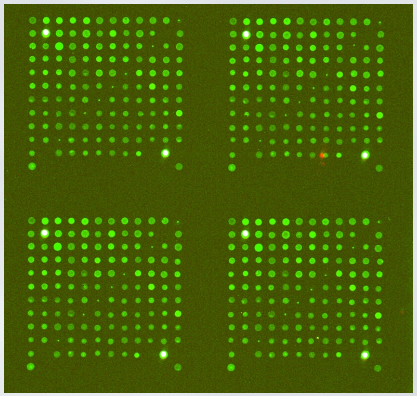


© SIRS-Lab GmbH, 2001 – 2006

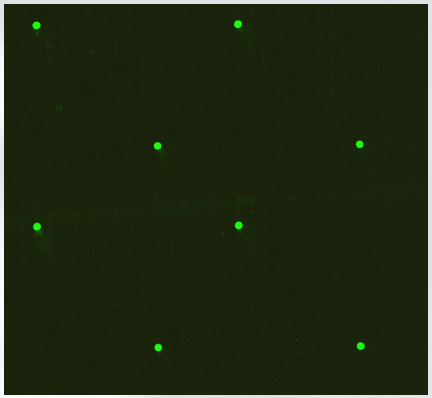
SIRS Lab

Post-production quality control @ SIRS-Lab

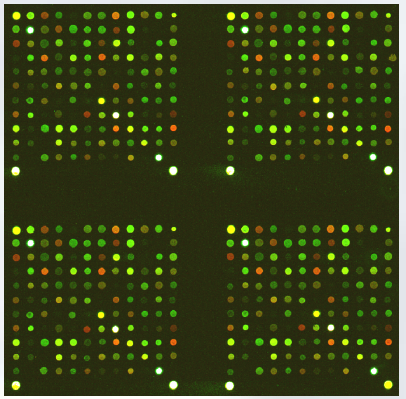
- spot morphology / homogeneity
- presence of each spot
- printing carryover
- sufficient DNA amount
- slide background
- hybridization performances



printing



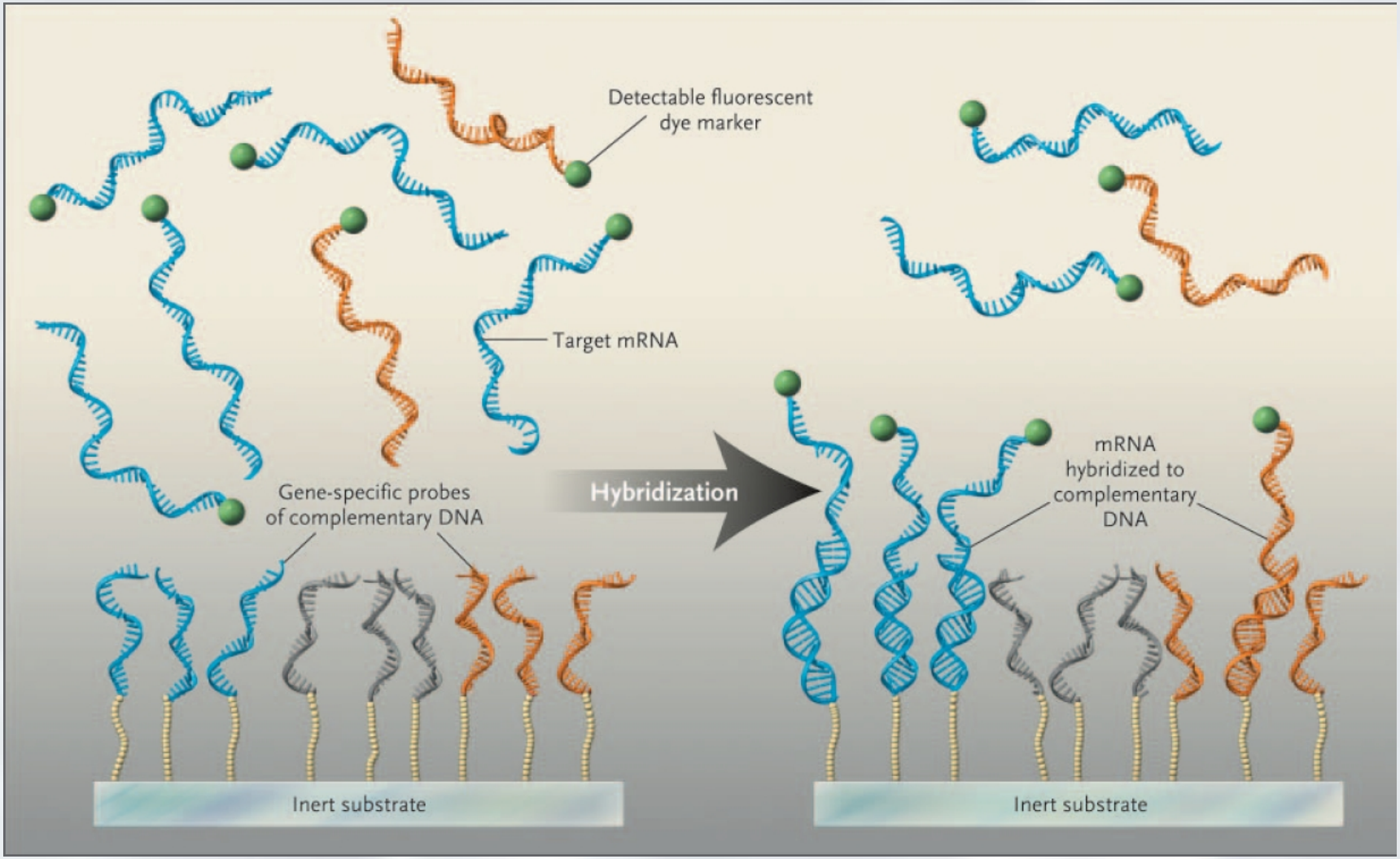
washing



hybridization

© SIRS-Lab GmbH, 2001 – 2006

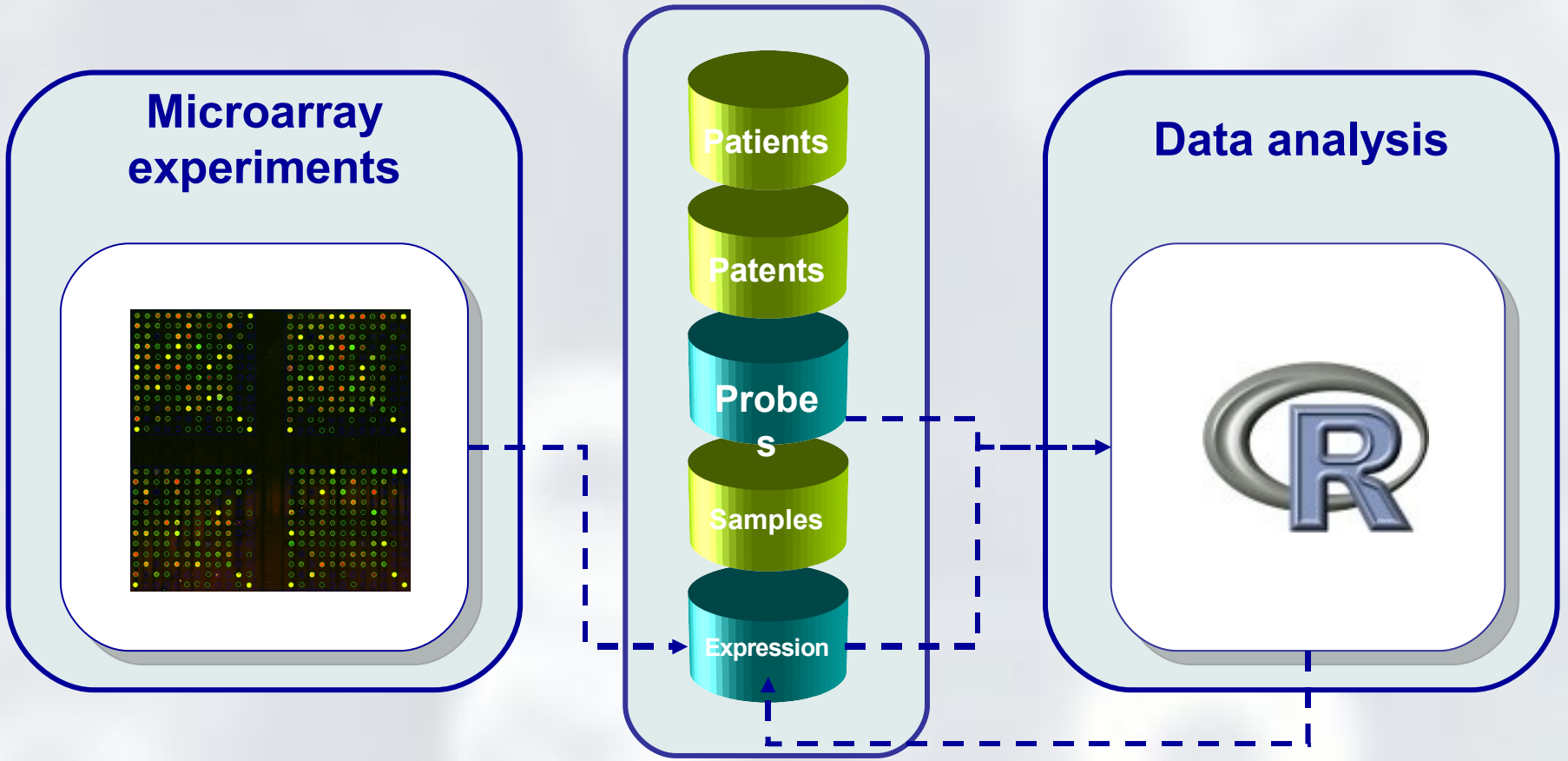
Hybridization with Gene Elements on a Microarray



Quackenbush J. N Engl J Med 2006; 354 (23): 2463

© SIRS-Lab GmbH, 2001 – 2006

Expression DB – data flow and analysis



© SIRS-Lab GmbH, 2001 – 2006

Imaging and data extraction

scanning

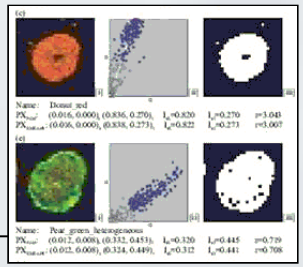
spot identification

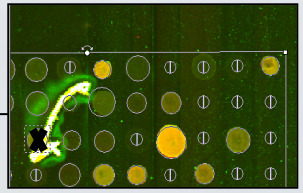
spot quality filtering

spot quantification

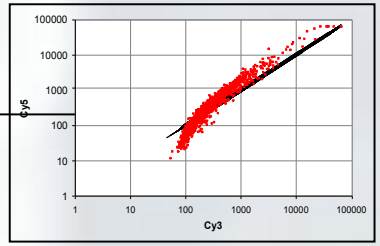
normalization

adjustment of batch effects





block	row	col	name	F532 median	F635 median
12	1	1	AA001392	35092	6095
12	1	2	AA001925	30729	28416
12	1	3	AA291398	23330	2303
12	1	4	AA400951	59703	56565
12	1	5	AA418009	51273	30860
12	1	6	AA420970	33707	5661

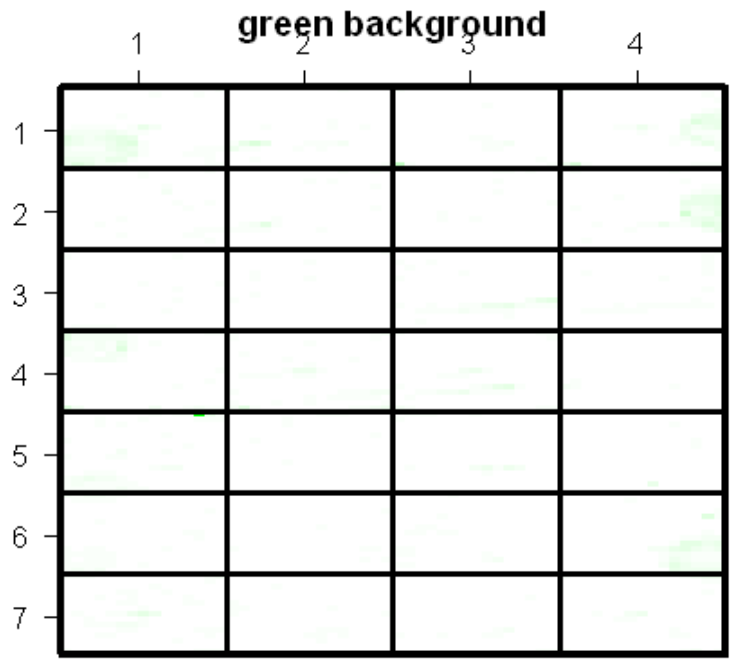


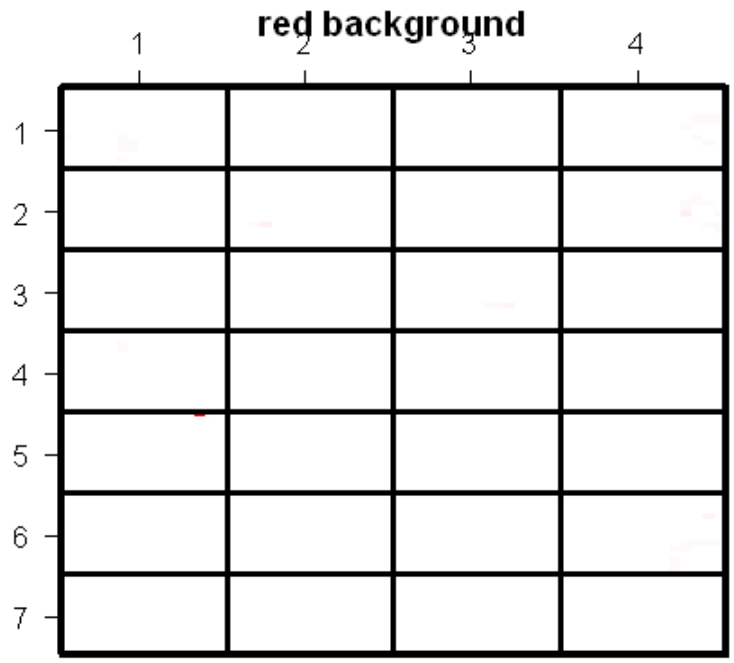
Quality evaluation using statistic based scanning reports

© SIRS-Lab GmbH, 2001 – 2006

SIRS Lab

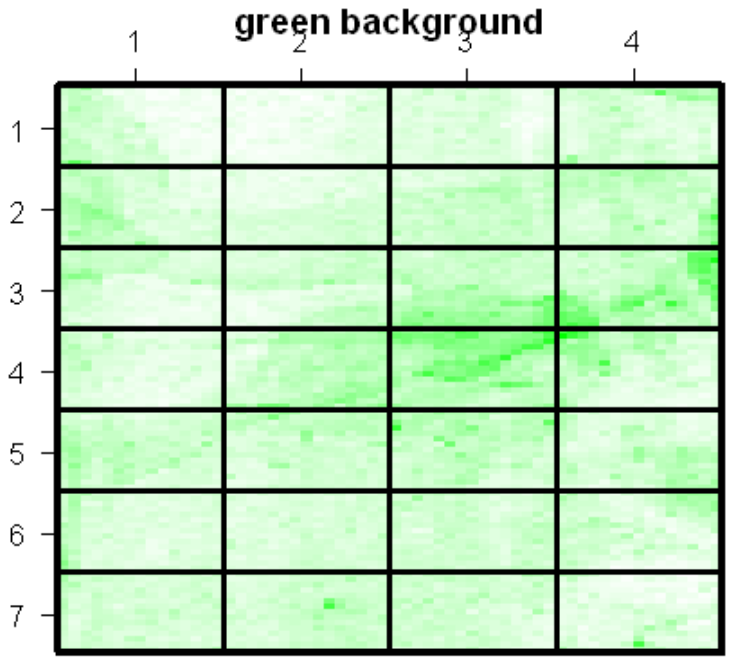
Quality control - background

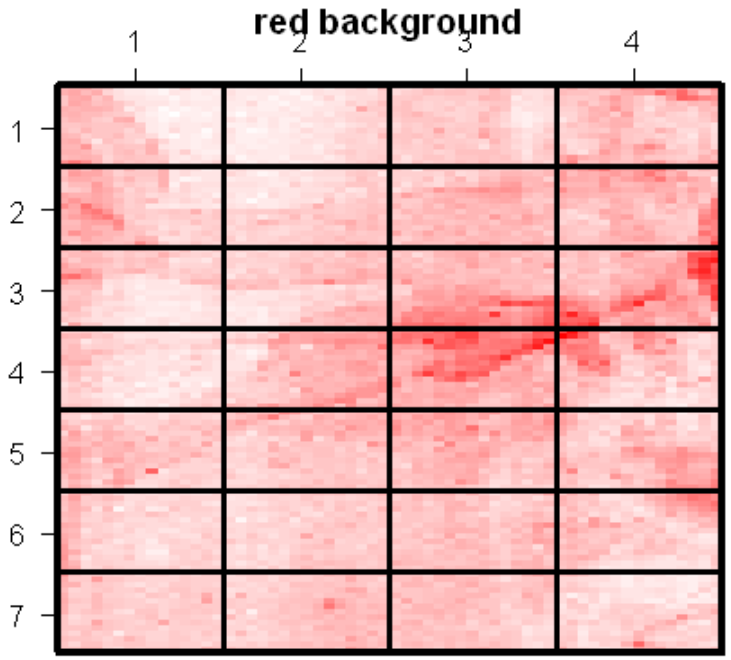




© SIRS-Lab GmbH, 2001 – 2006

Quality control - background





© SIRS-Lab GmbH, 2001 – 2006

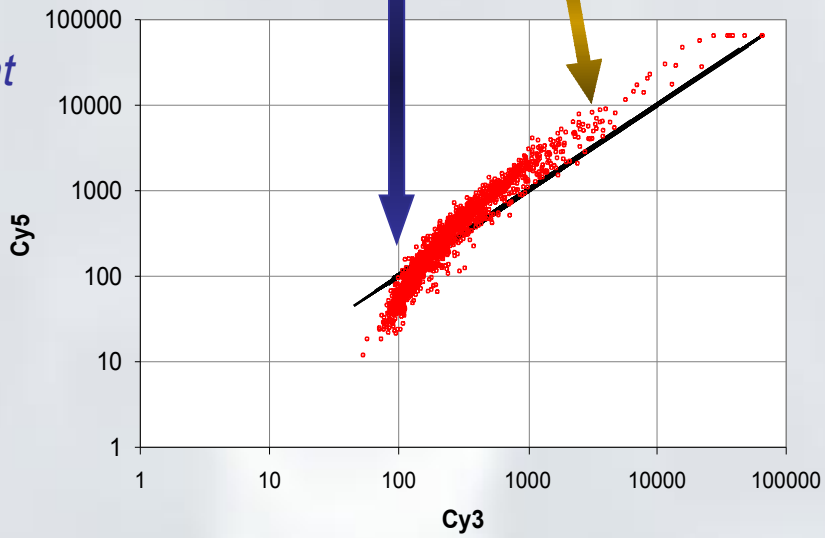
Systematic Bias

- **two-component error model** (*Rocke and Durbin[1], 2001):*

signal intensity = offset + (gain * abundance)

dominant for low intensity signals
source:
scanner adjustment

dominant for high intensity signals,
source: labelling efficiency bias

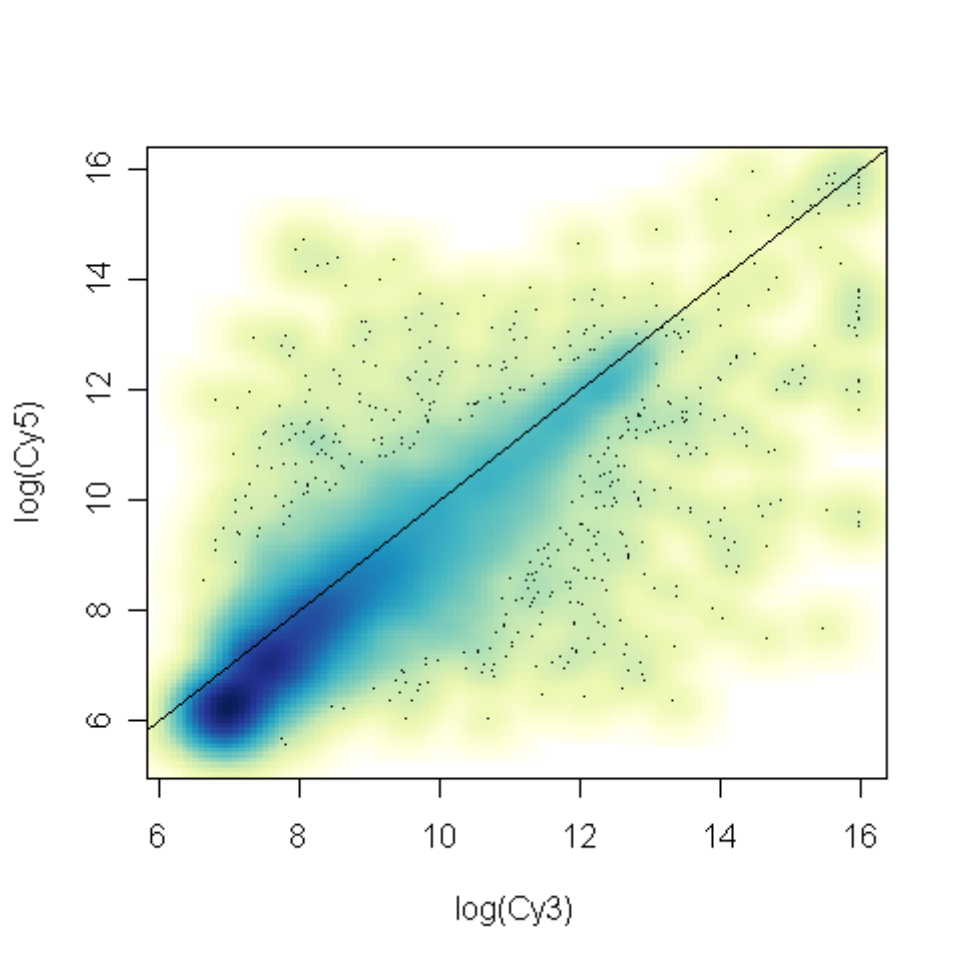


- **results in non-linear systematic bias** observable in self-to-self experiments, where no different expression levels are expected for a gene

[1] *Journal of Computational Biology*, 8 (2001): 557-569

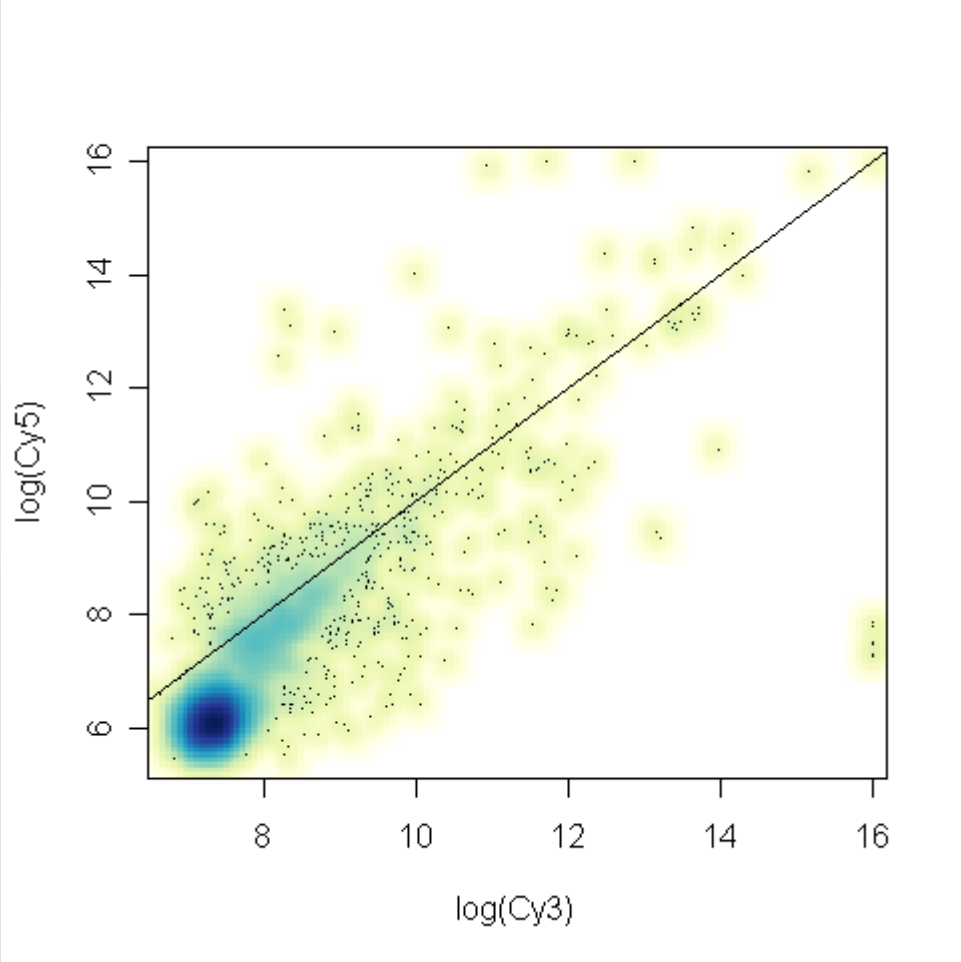
© SIRS-Lab GmbH, 2001 – 2006

Quality control – raw signals



© SIRS-Lab GmbH, 2001 – 2006

Quality control – raw signals



© SIRS-Lab GmbH, 2001 – 2006

Normalization: removal of systematic bias

- **parametric normalization:**

  location via median, scale via MAD, variance-stabilizing transformation via LTS-regression and arcsinh, etc.

- **nonparametric normalization:**

  spatial or intensity dependent local regression (loess), quantile normalization, etc.

© SIRS-Lab GmbH, 2001 – 2006

Variance-stabilizing normalization

- **two-component error model** (*Rocke and Durbin[1], 2001):*

signal intensity = offset + (gain * abundance)

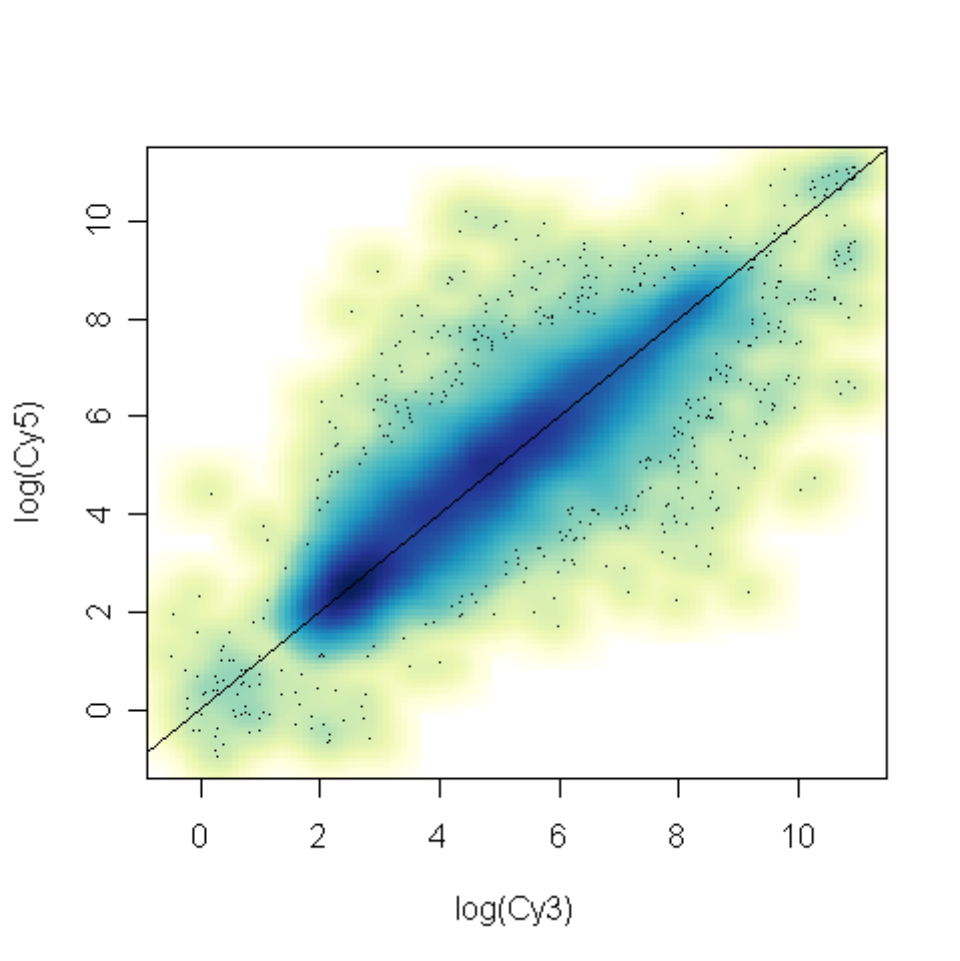
Huber et al.[2] *(2002):*

- **estimation of *offset* and *gain* via least trimmed sum of squares (LTS) regression**
- **application of *arcsinh* as variance-stabilizing transformation**

[1] *Journal of Computational Biology*, 8 (2001): 557-569
[2] *Bioinformatics*, Vol. 18 Suppl. 1 (2002): 96-104

© SIRS-Lab GmbH, 2001 – 2006

Variance-stabilized, normalized signals



© SIRS-Lab GmbH, 2001 – 2006

Statistical Analysis

- **univariate analysis per gene:**

  t-Test, Wilcoxon-Test, ANOVA, regression models, Bayesian approaches, etc.

- **multivariate analysis:**

  PCA, nonnegative matrix decomposition, clustering, multidimensional scaling, support vector machines, neural networks, Bayesian Decomposition, etc.

© SIRS-Lab GmbH, 2001 – 2006

Multiple testing problems

Analysis per gene: many (>> 1000) simultaneous tests

Problem: false positives

Controlled via: Bonferroni, family-wise error rate, false discovery rate (FDR), tail-probability of the proportion of false-positives

© SIRS-Lab GmbH, 2001 – 2006

False discovery rate - q value

False positive rate (FPR):

$$FPR \approx \frac{\#\ \text{false positives}}{\#\ \text{truly null}}$$

False discovery rate (FDR):

$$FDR \approx \frac{\#\ \text{false positives}}{\#\ \text{significant tests}}$$

p value:

"Minimum FPR at which the test statistic can be called significant"

q value (Storey, 2002[1]):

"Minimum FDR at which the test statistic can be called significant"

[1] *Journal of the Royal Statistical Society , Series B,* 64 (2002): 557-569

© SIRS-Lab GmbH, 2001 – 2006

Peritonitis vs. SIRS + OD vs. ICU-Controls

	ICU-Controls	SIRS + OD	Peritonitis
No. of patients	17	18	15
Mortality	0 (0 %)	1 (5.6 %)	9 (60 %)
Gender [m/f]	14/3	9/9	11/4
Age [y]	64 (11)	75 (10.75)	66 (9)
SOFA score	3* (2) [16]	8* (2)	10 (2)
No. of OD	1* (1)	3 (1)	3 (0.5)
PCT [ng/ml]	0.37* (0.27) [10]	3.97 (6.27) [12]	8.85 (25.02) [13]
CRP [mg/l]	66.5* (40.73) [16]	80.1* (29.6) [17]	146 (97)
WBC [no/l]	9400 (4500)	13600 (8525)	11200 (10000)

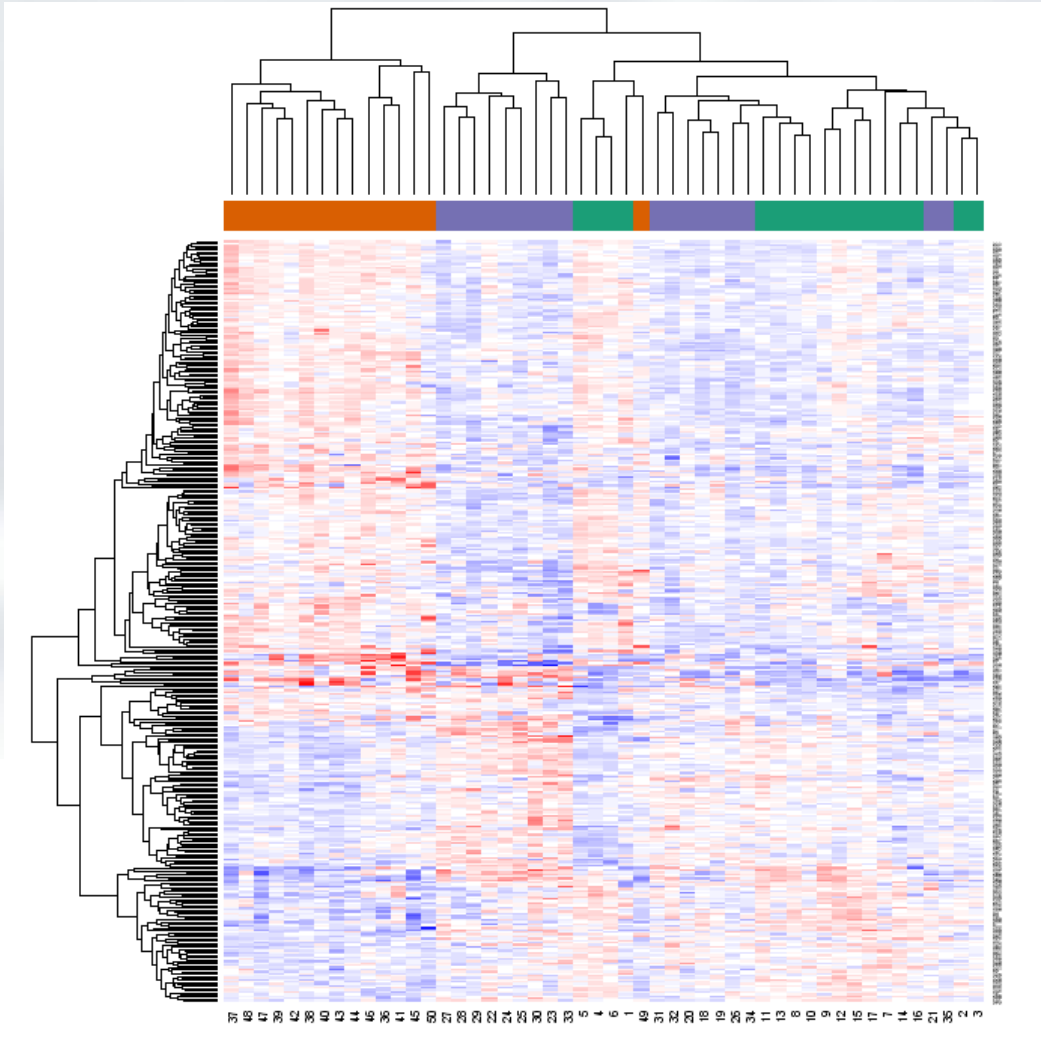
median (IQR) [No. of patients]
***p < 0.05 (Wilcoxon rank sum test)**

© SIRS-Lab GmbH, 2001 – 2006

Peritonitis vs. SIRS + OD vs. ICU-Controls

Hierarchical cluster analysis
- complete linkage method

1-way ANOVA per gene:
p-value <= 0.01
368 significant genes
6 expected false positives



© SIRS-Lab GmbH, 2001 – 2006

SIRS Lab

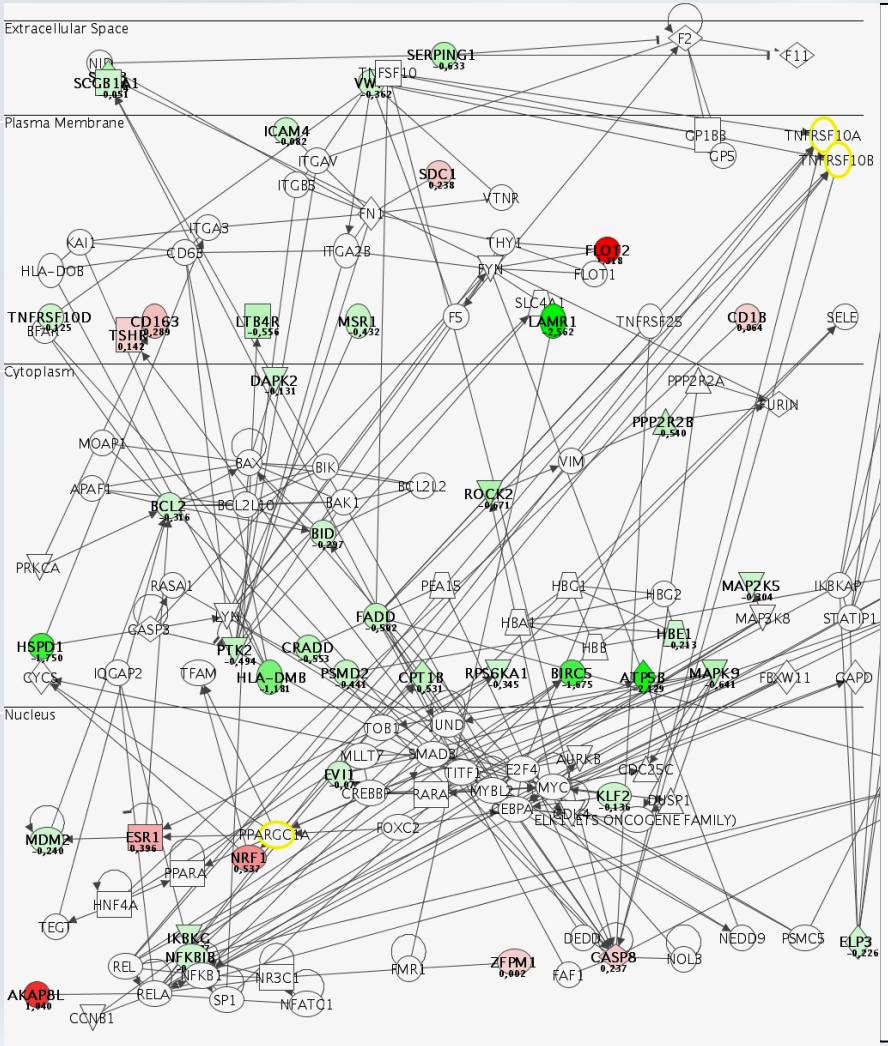
Signal transduction pathways differentially expressed in patients with infectious vs. non~ causes of organ dysfunction using Ingenuity® pathway analysis)

Extracell. space

Plasma membrane

Cytoplasma

Nucleus



- PI3KInase
- NFkB
- Interleukines Signaling
- Chemokines Signaling
- T Cell Receptor
- B Cell Receptor
- p38 MAPK
- JAK-STAT
- Apoptosis Signaling
- SAPK/JNK Signaling
- Interferon Signaling
- Estrogen receptor
- GPCR Signaling
- IGF-1
- TGF-B
- PPAR Receptor
- PDGF Signaling.....

© SIRS-Lab GmbH, 2001 – 2006

SIRS Lab