

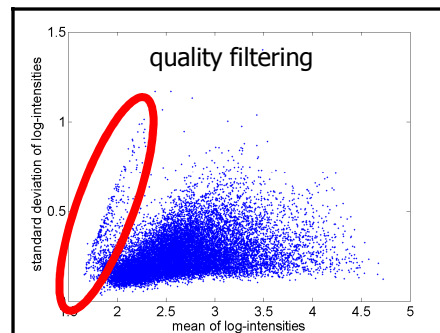
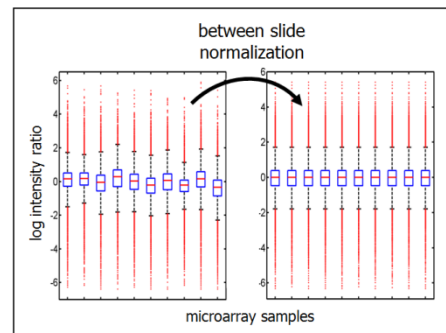
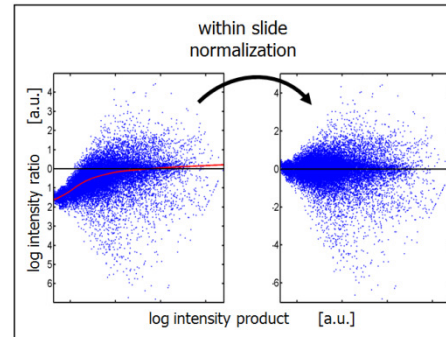
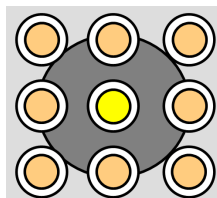
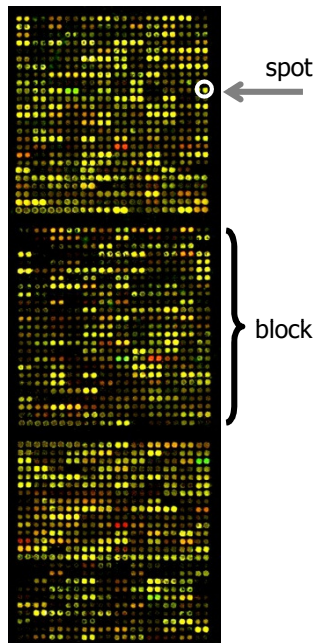
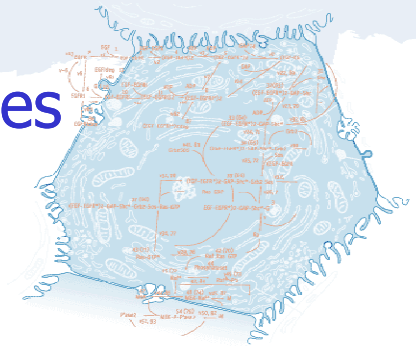
Processing of high-dimensional data for mathematical modeling

July, 14th 2009

Clemens Kreutz, Jens Timmer

ckreutz@fdm.uni-freiburg.de
jeti@fdm.uni-freiburg.de

Preprocessing: Many competing strategies



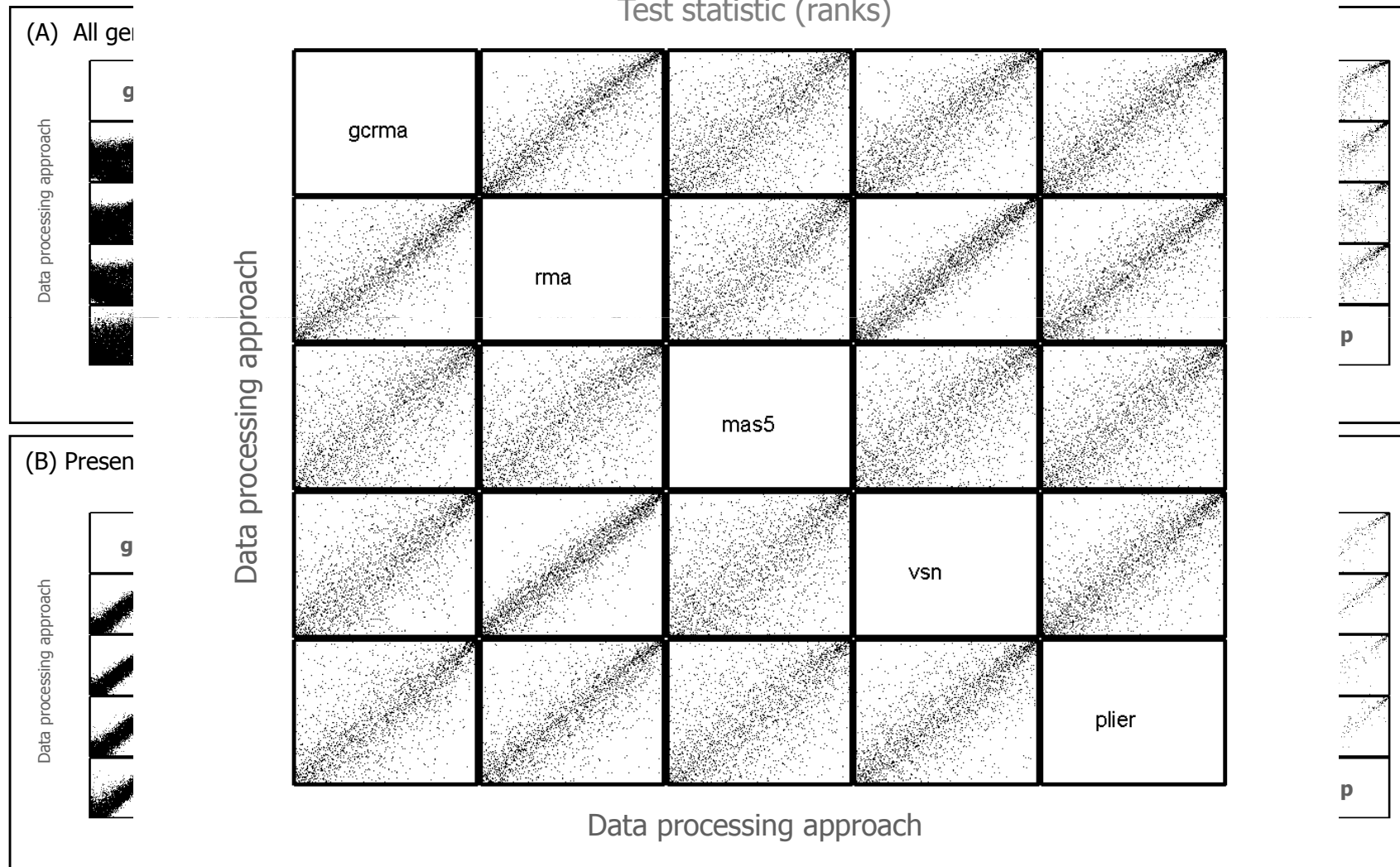
There are many competing reasonable approaches.

A change of the data processing has noticeable effects on the outcomes.

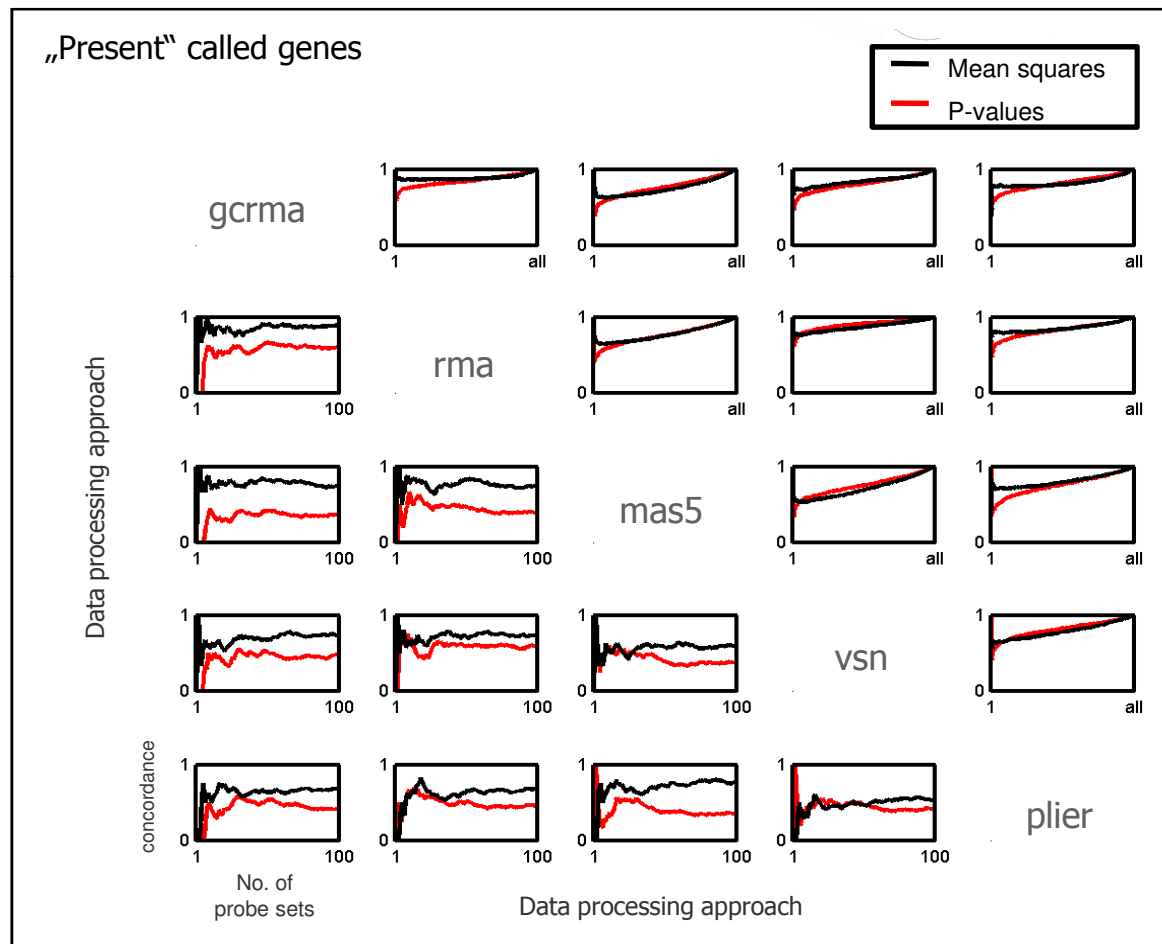
Label	Technology	Description
A	cDNA	Standard setting
B	cDNA	Ratio of medians (instead of median of ratios) over pixel intensities within a spot
C	cDNA	Mean of ratios (instead of median) over the pixel intensities within a spot
D	cDNA	<i>Bayesian background correction</i> (instead of local subtraction)
E	cDNA	<i>LOWESS</i> smoothing parameter of 0.2 (instead of 0.1)
F	cDNA	Block-wise <i>LOWESS</i> correction (instead of global)
G	cDNA	Elimination of data points close to the background and/or large within spot CV
H	cDNA	Elimination of data points with potential saturation effects
I	cDNA	Quantile normalization (instead of linear scaling)

g	Affymetrix	<i>Sequence Dependent Robust Multi-array Analysis (gcrma)</i>
r	Affymetrix	<i>Robust Multi-array Analysis (rma)</i>
m	Affymetrix	<i>Affymetrix Microarray Suite 5.0 (mas5) algorithm</i>
v	Affymetrix	<i>Variance Stabilization Normalization (vsn)</i>
p	Affymetrix	<i>Probe Logarithmic Error Intensity Estimate (PLIER)</i>

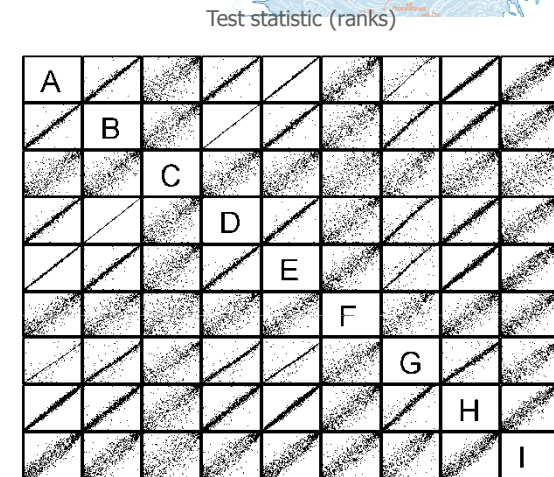
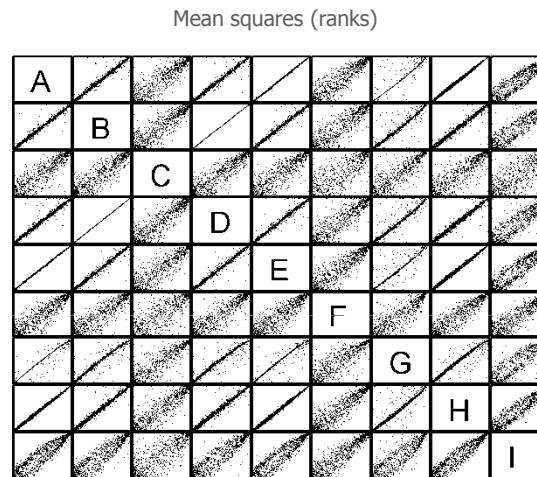
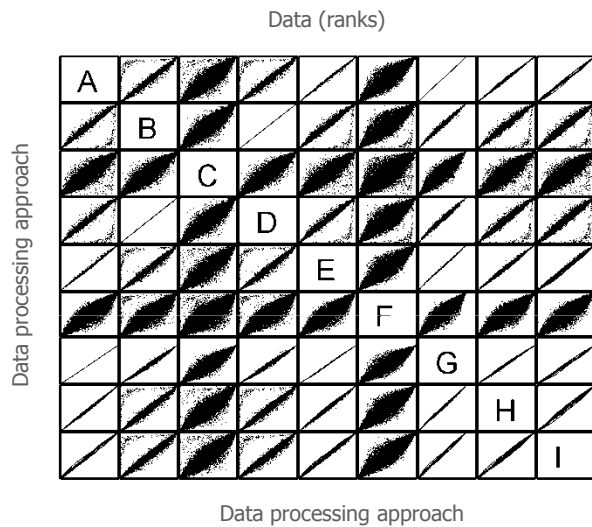
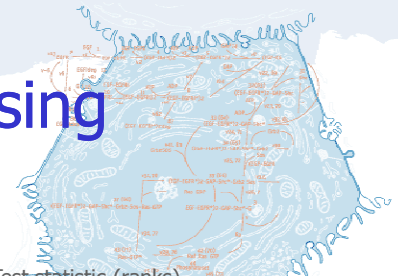
Affymetrix: Dependency on the preprocessing



Affymetrix: Dependency on the preprocessing

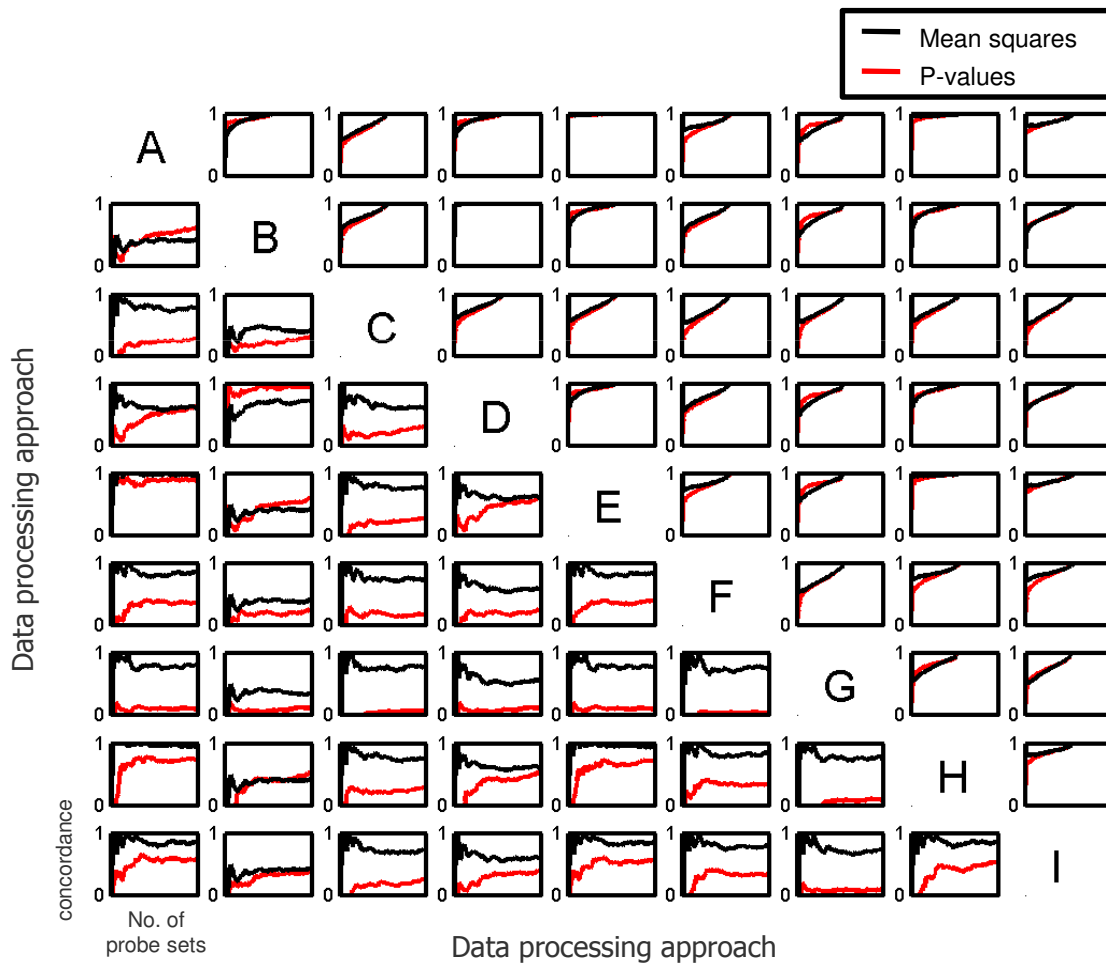


cDNA chips: Dependency on the preprocessing

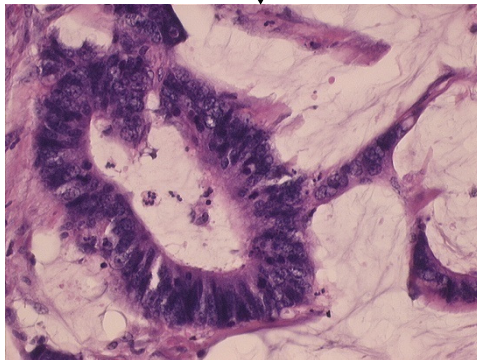
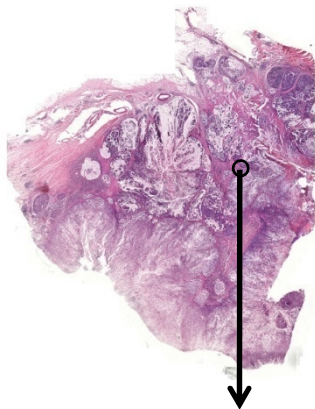
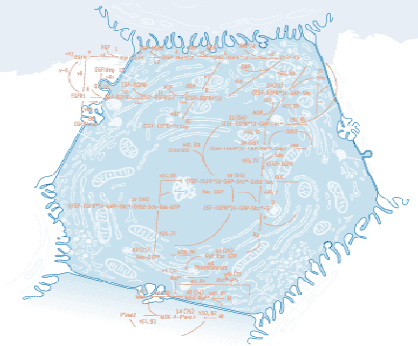


Label	Technology	Description
A	cDNA	Standard setting
B	cDNA	Ratio of medians (instead of median of ratios) over pixel intensities within a spot
C	cDNA	Mean of ratios (instead of median) over the pixel intensities within a spot
D	cDNA	<i>Bayesian background correction</i> (instead of local subtraction)
E	cDNA	<i>LOWESS</i> smoothing parameter of 0.2 (instead of 0.1)
F	cDNA	Block-wise <i>LOWESS</i> correction (instead of global)
G	cDNA	Elimination of data points close to the background and/or large within spot CV
H	cDNA	Elimination of data points with potential saturation effects
I	cDNA	Quantile normalization (instead of linear scaling)

cDNA chips: Dependency on the preprocessing



Microarrays for FFPE tissues



Formalin fixation and paraffin (FFPE) embedding is a standard sample preparation in histopathology.

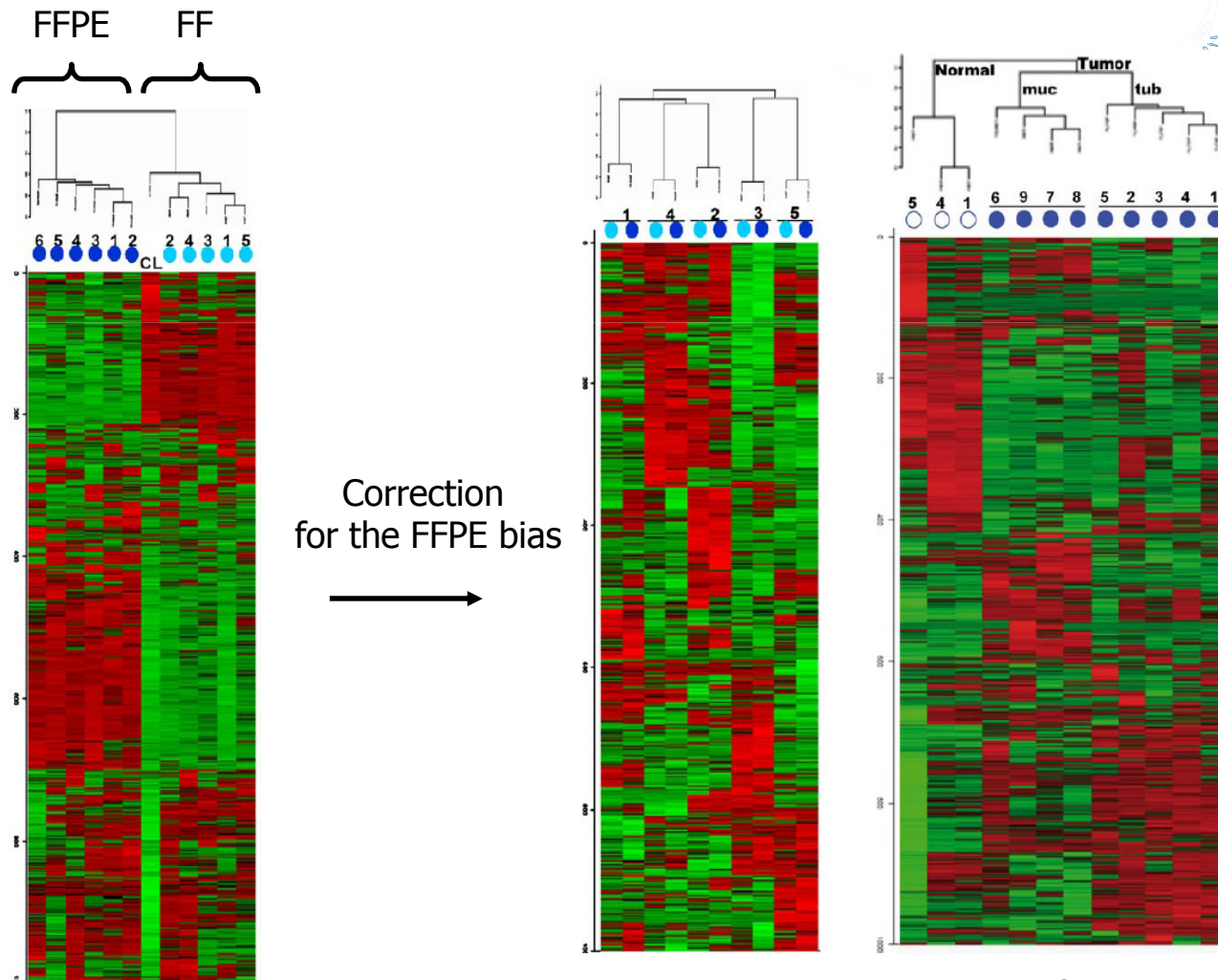
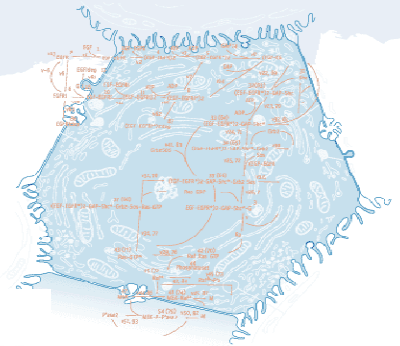
So far, microarrays were not applicable to FFPE samples.

We could show for FFPE colorectal tumor tissues that

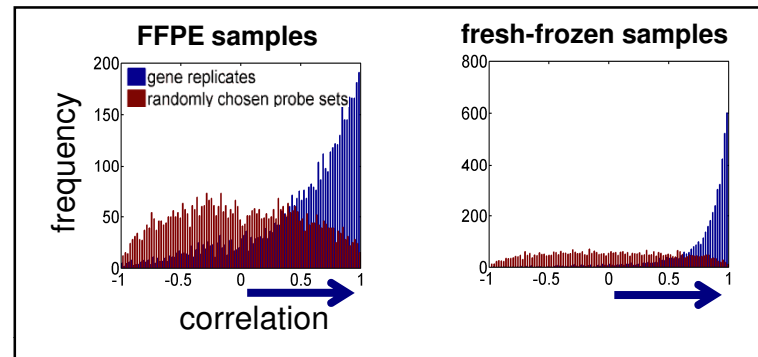
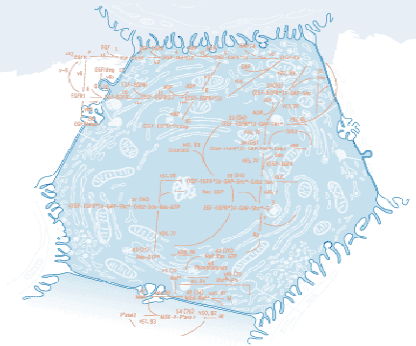
1. Adapted sample preparation
2. Latest chip technology
3. Adapted normalization

yield informative data.

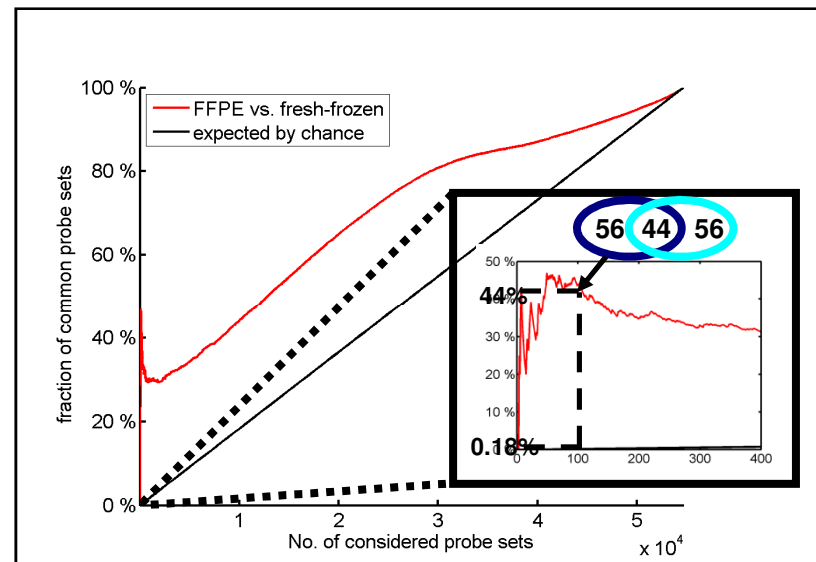
FFPE tissue yield informative data



FFPE tissue yield informative data

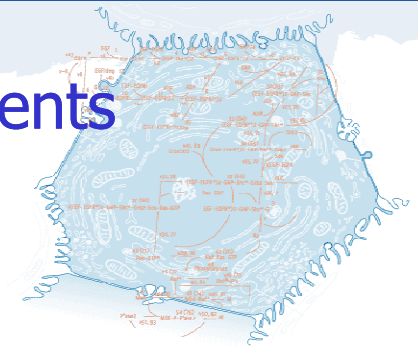


Probe-sets from the same gene are correlated.



44 % overlap within the 100 mostly regulated genes between the samples.

Microarray based classification of CLL patients



Chronic lymphatic leukemia, N=51, Affymetrix

Prediction of the mutational status
of the immunoglobulin heavy-chain gene:

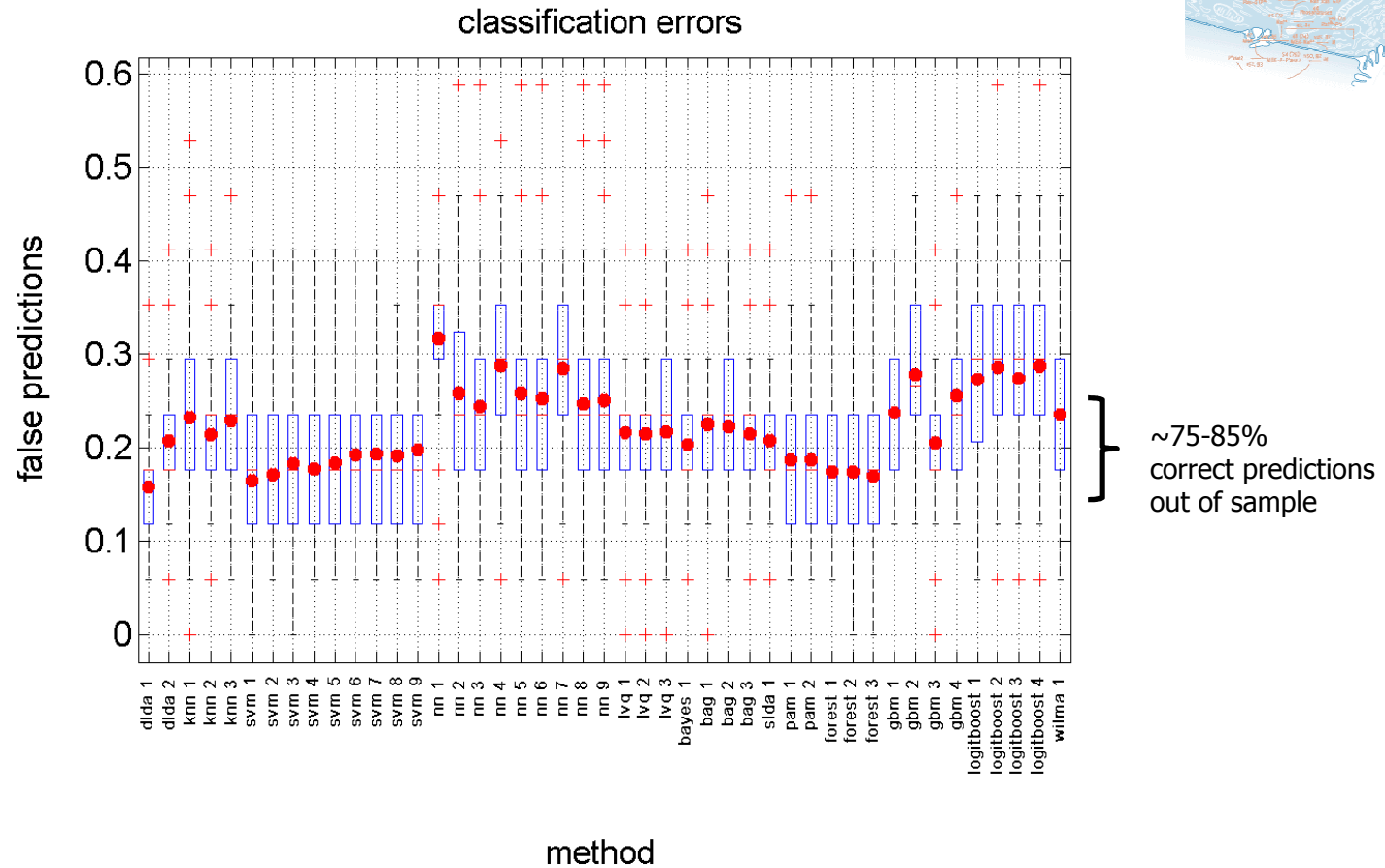
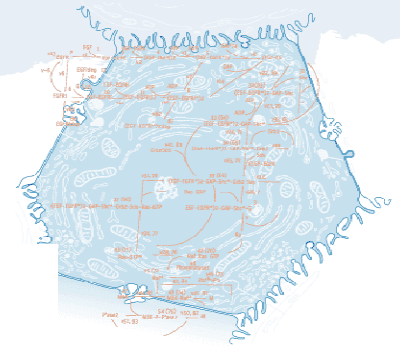
IgV (H) mutated *vs.* IgV (H) not mutated (34 *vs.* 17)

The mutational status is known to be related to the prognosis

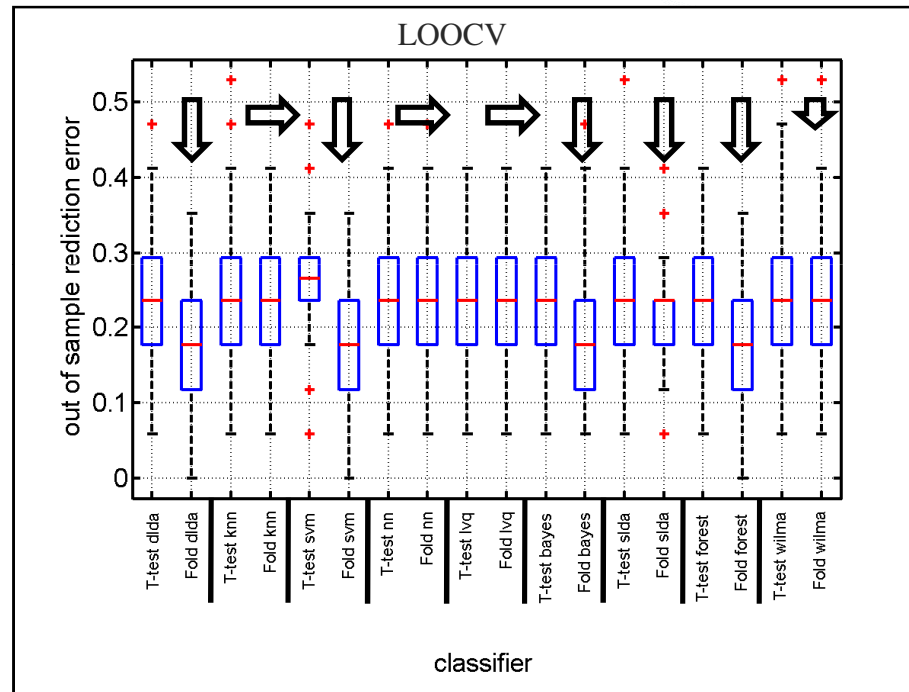
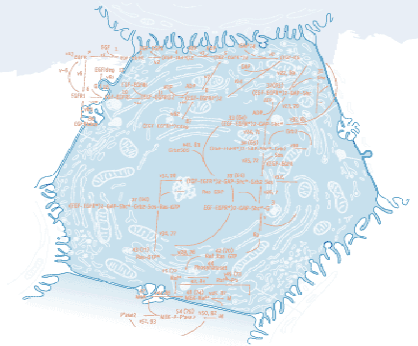
Proof of concept study

Classification performance evaluated for
many classifier-, dimension reduction-,
and cross validation approaches

Prediction of the mutational status



Effect size based dimension reduction and classification

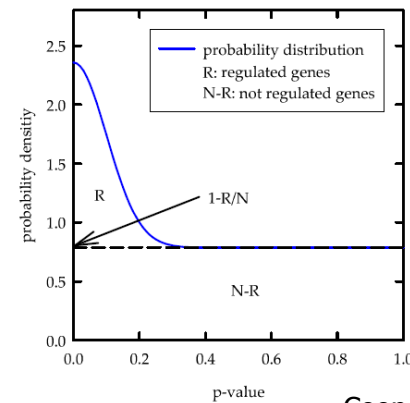
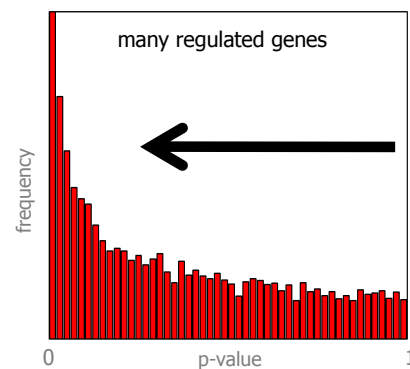
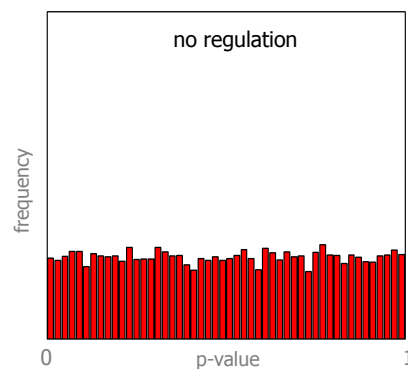
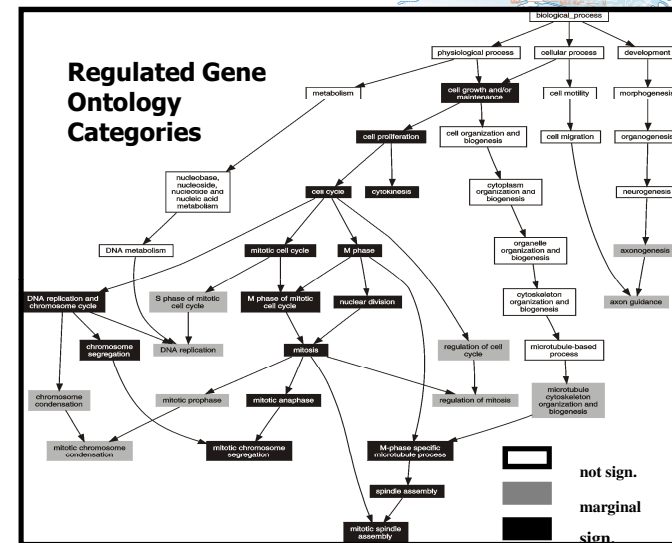


New approach for gene set analyses

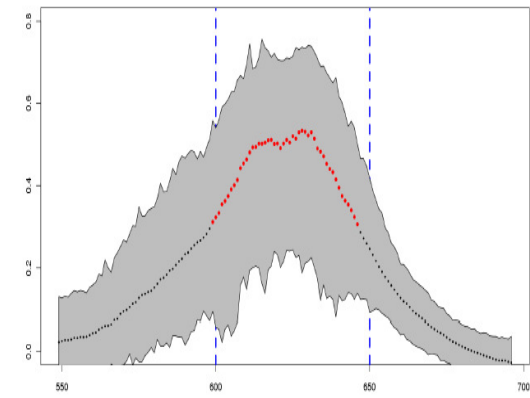
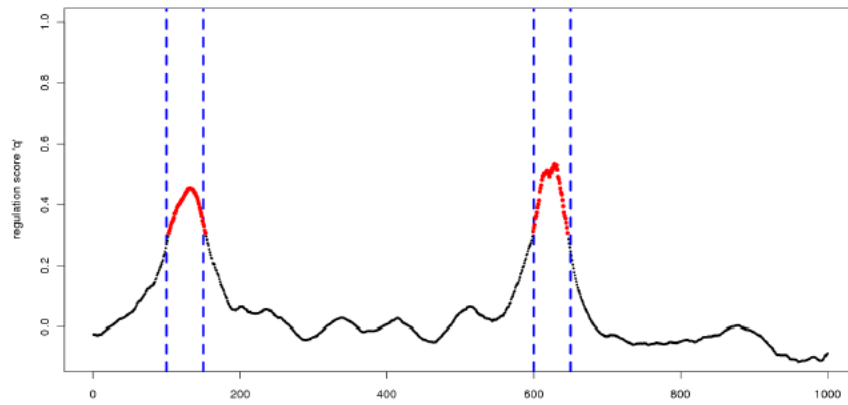
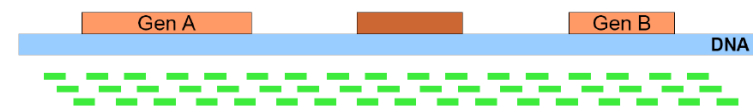
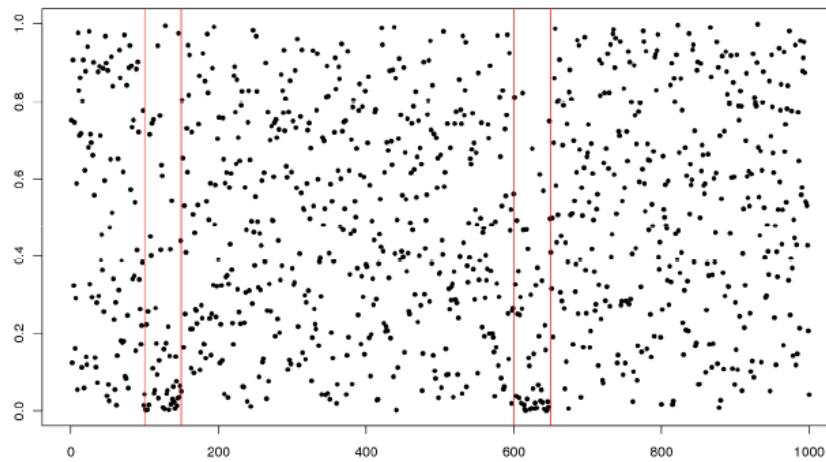
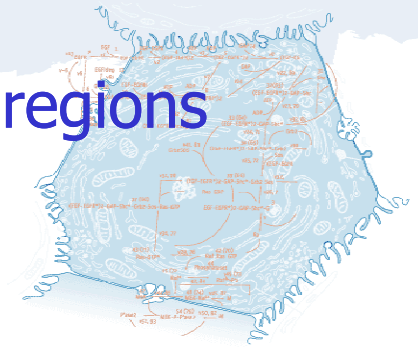


Investigation of the behavior of functionally related genes

We estimate the fraction of regulated genes in in a gene set.

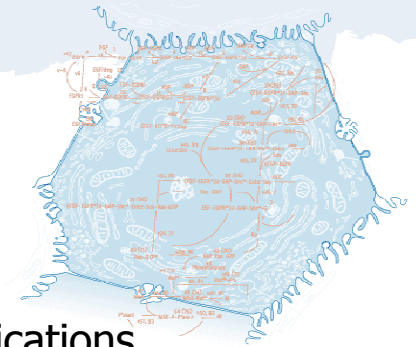


Tiling array data and differentially expressed regions



Position on the DNA

AG Timmer vs. AG Schumacher



Identification of the processes
at the molecular level



Clinical applications
of high throughput data

Exp. condition as predictor,
gene expression as response



Gene expression as predictor,
clinical response

Homogeneous biological
samples (e.g. cell lines)



Heterogeneous clinical samples
(e.g. blood, clinical covariates)

Different data processing,
platform specific calibration



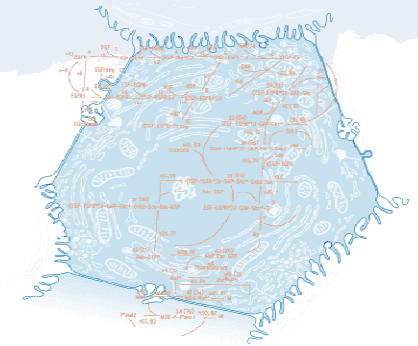
Established data processing

Mechanistic explanation and
modeling of the dynamics
(e.g. regulatory networks)



Phenomenological prediction
(e.g. for diagnosis, prognosis)

Core Facility for Data Analysis (ZBSA)



Adequate statistical analyses

Experimental design

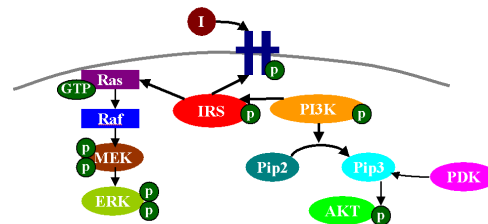
Development of new methodological approaches

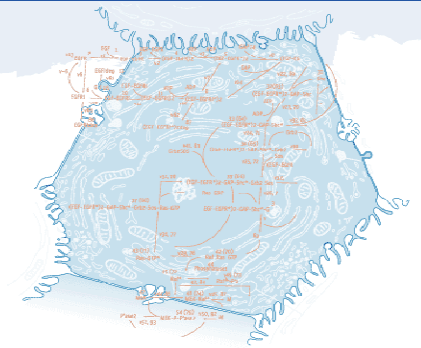
Platform specific data processing

Cooperation with the experimental core facilities
(e.g. standard processing workflows)

Cooperation with the IMBI

Dynamic modeling of molecular interaction
networks





Thanks for your attention.

ckreutz@fdm.uni-freiburg.de
jeti@fdm.uni-freiburg.de