

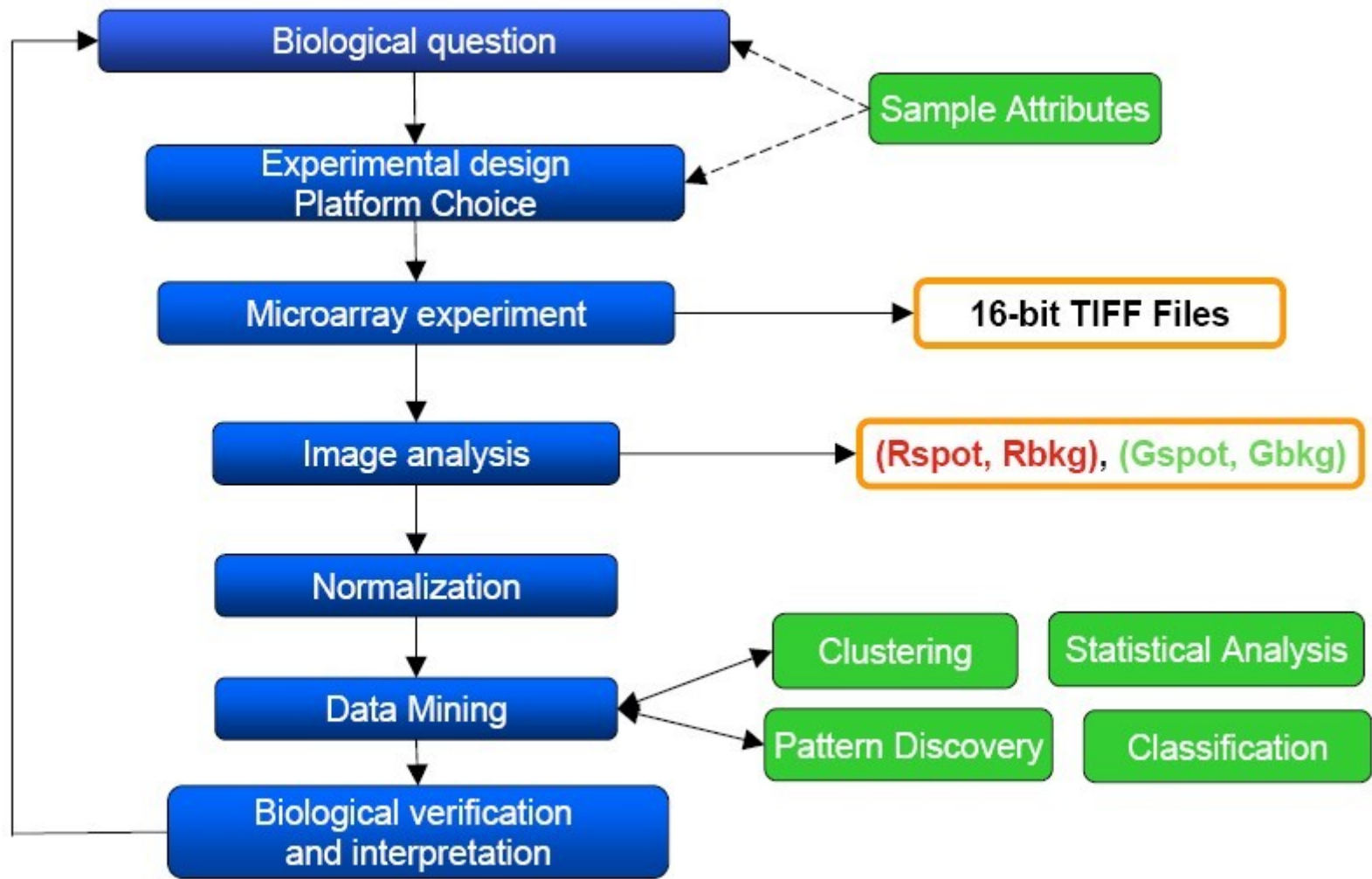
Microarray Data Analysis

Introduction to Bioinformatics
Dortmund, 16.-20.07.2007

Lectures:
Sven Rahmann

Exercises:
Udo Feldkamp, Michael Wurst

Stages of a Microarray Experiment



Gene Expression Omnibus (GEO)

- NCBI repository for gene expression data
- <http://www.ncbi.nlm.nih.gov/geo/>
- integrated into Entrez
- MIAME-compliant

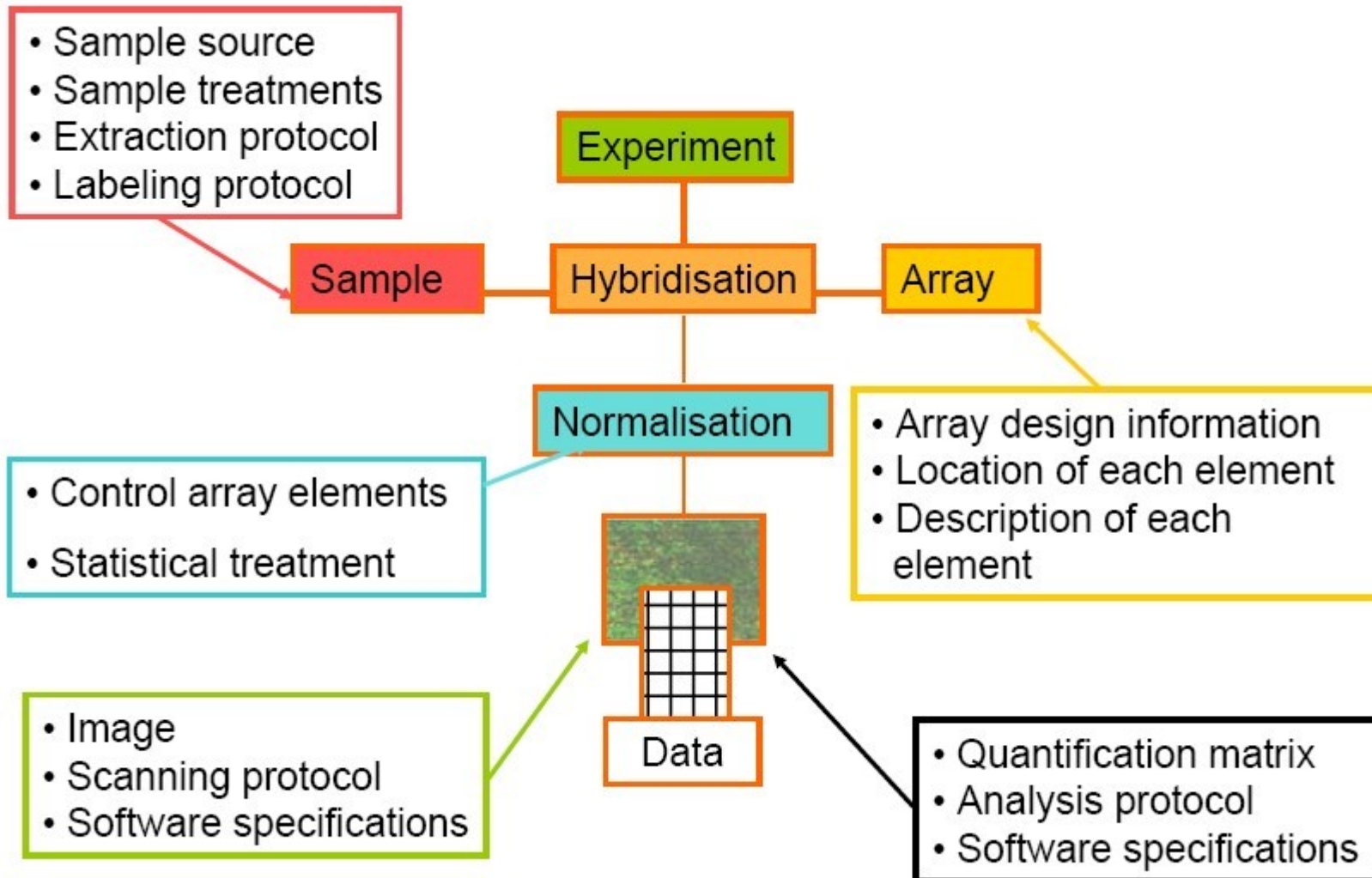
ArrayExpress

- EBI repository for gene expression data
- <http://www.ebi.ac.uk/arrayexpress/>
- MIAME-compliant

MIAME

- Minimum Information About a Microarray Experiment, <http://www.mged.org>
- Standardization effort by Alvis Brazma (EBI) and colleagues at 17 institutions
- 6 areas of information standardized
 - experimental design
 - microarray design
 - sample preparation
 - hybridization procedures
 - image analysis
 - controls for normalization

MIAME Information



Analysis Steps

- 1. Read data
- 2. Exploratory plots
- 3. Normalize (eliminate systematic biases)
- 4. Find differentially expressed genes
- 5. Output results

Software for Analysis

- Vendor-supplied software (e.g., Affymetrix)
 - Quality has increased considerably since early days
- Free community software packages
 - mostly not stand-alone products,
 - but packages for the statistical language R
 - in a framework called BioConductor
- Some statistical experience recommended to analyze microarray data
 - problem: many variables, few samples

The R Project

- URL: <http://www.r-project.org>



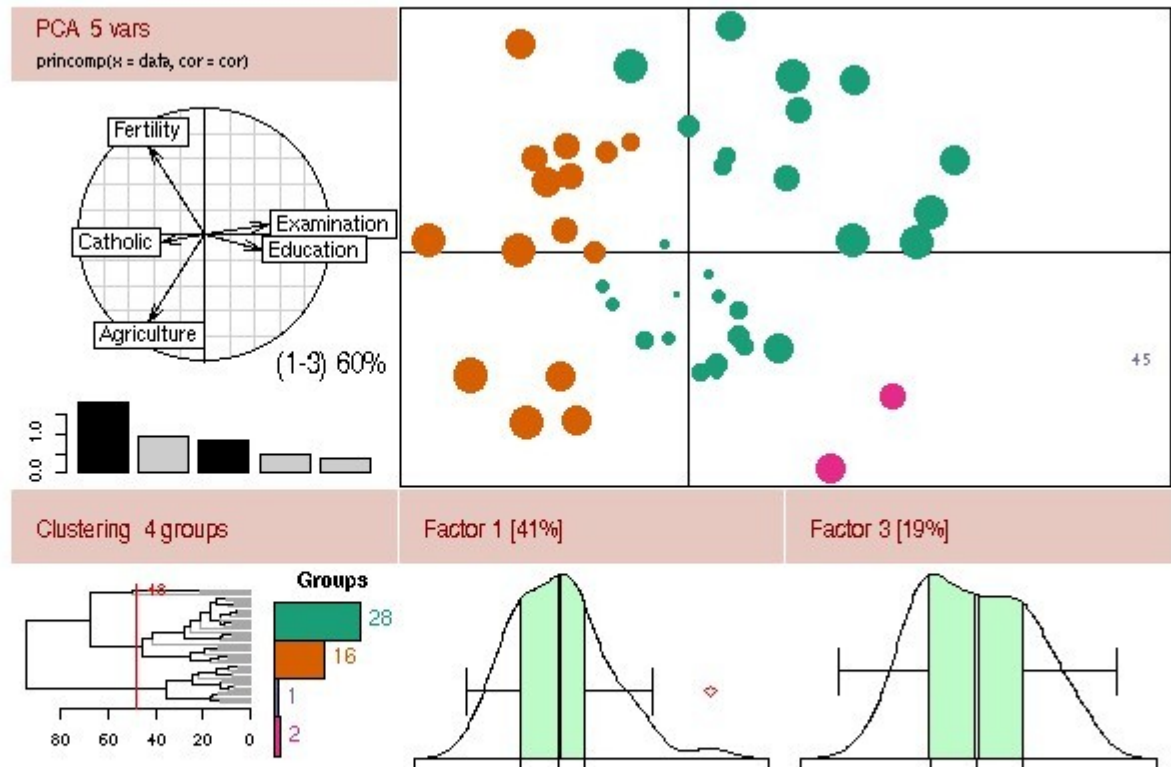
The R Project for Statistical Computing

About R
[What is R?](#)
[Contributors](#)
[Screenshots](#)
[What's new?](#)

Download
[CRAN](#)

R Project
Foundation
[Members & Donors](#)
[Mailing Lists](#)
[Bug Tracking](#)
[Developer Page](#)
[Conferences](#)
[Search](#)

Documentation
[Manuals](#)
[FAQs](#)
[Newsletter](#)
[Wiki](#)
[Books](#)



Getting started:

- R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To download R, please choose your preferred [CRAN mirror](#).
- If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

<http://www.bioconductor.org/>

- R packages for biological data analysis

BIOCONDUCTOR
open source software for bioinformatics

Bioconductor is an open source and open development software project for the analysis and comprehension of genomic data.

startseite | what is it? | herunterladen | dokumentation | publications | workshops | cabig

project news

- ▶ [2007-06-08](#)
Changes in BioC Devel, May 2007
- ▶ [2007-05-02](#)
Changes in BioC Devel, April 2007

[Mehr ...](#)

▶ What is it?

▶ Install - How To

▶ Browse Packages

▶ FAQ

▶ For Developers

▶ Upload a package

▶ Join mailing list

BioC Release 2.0

Bioconductor 2.0 was released 26 April, 2007. This release is designed for R 2.5.0. View the packages [here](#)

BioC2007: Where Software and Biology Connect

Mark your calendars! The conference will take place on the 6th and 7th; the 8th will be a developer meeting.
06-08 August 2007, Seattle, USA
[🔒 Details and registration](#)

Bioconductor Advanced Course

This 3-day course on Northwestern University Medical School's downtown Chicago campus will cover advanced topics in genomics and proteomics data analysis. Morning lectures and

R Example: t-test

- `c()` -- build a vector by concatenation
- `r***()`, e.g., `rnorm()` -- generate random numbers
- `boxplot()` -- plot distribution of vector
- `t.test()` -- compare mean of two vectors

```
>t.test(rnorm(100),rnorm(10))
```

```
t = -0.5098, df = 17.577, p-value = 0.6165  
alternative hypothesis: true difference in means is not equal to 0  
95 percent confidence interval:  
 -1.0844964  0.6615357  
sample estimates:  
 mean of x  mean of y  
-0.3436742 -0.1321938
```

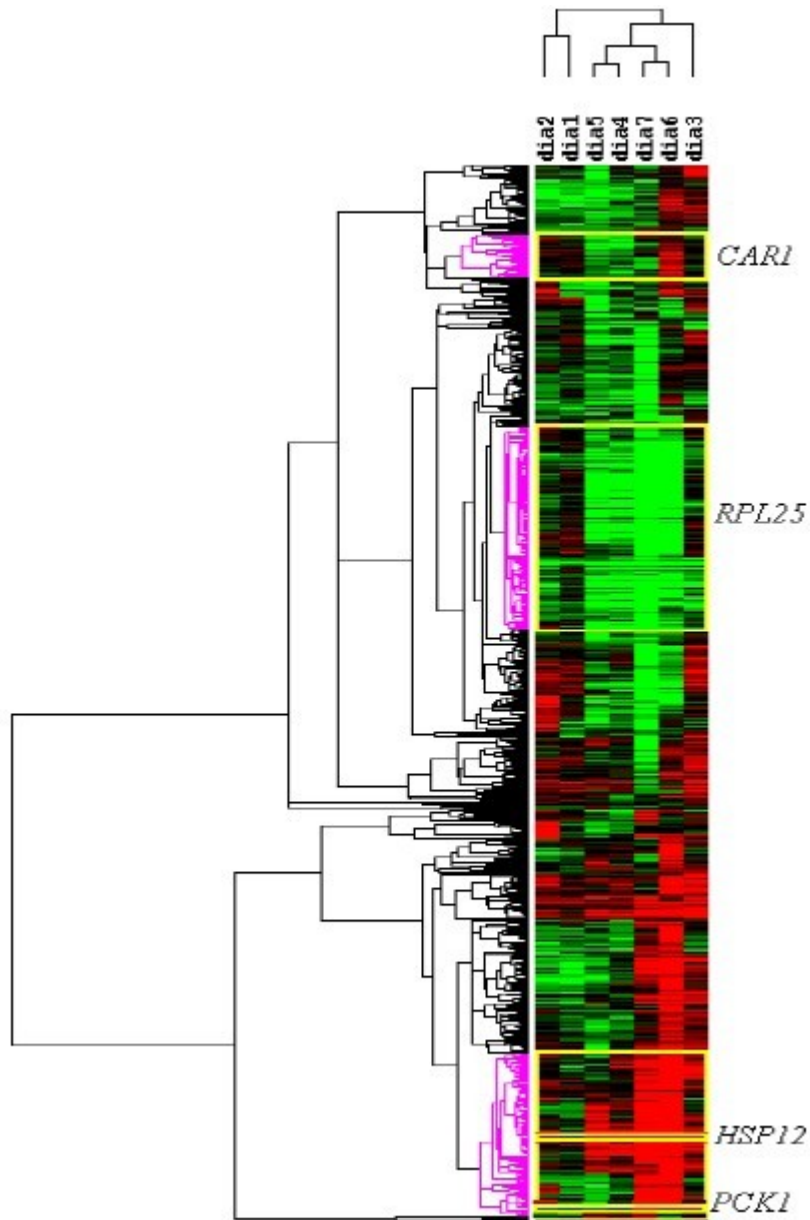
Installing BioConductor in R

- `source("http://www.bioconductor.org/biocLite.R")`
- installs most frequently needed BioConductor packages from the WWW
- R Windows version has convenient package management

Results of a Microarray Experiment

- Starting with a hypothesis
 - support for hypothesis
 - falsification of hypothesis
- Starting with no idea
 - exploratory analysis, hypothesis generation
- Interpreting visualization of outcome

Clustering Genes and Experiments



Some Clustering Methods

- Agglomerative / divisive hierarchical clustering
- K-means
- Self-Organizing Maps (SOMs)
- Clustering via Principal Component Analysis (PCA)