Three ways to analyse GeneChip data

- **Absolute analysis (1 chip):**
  Signals, detection p-values, detection calls

- **Comparison analysis (1 chip vs 1 chip):**
  SLRs, change p-values, change calls

- **Group analysis (m chips vs n chips):**
  Statistics on absolute and comparison analysis results
  e.g.: t-test on signals, t-test on SLRs, percent of increase or decrease calls

Toy data: 6 hybridizations from the Affymetrix Latin Square Experiment (HG-U133A)

- Comparison of EXP1 (3 replicates) versus EXP2 (3 replicates)
- 30 spikes were mixed into the background RNA at 14 concentrations (0, 0.125, 0.25, ..., 512 pmol)
- Concentrations of spikes differ between EXP1 and EXP2 (fold change = 2)
- SLRs and change calls were calculated between each pair of chips from EXP1 and EXP2 (9 comparisons)

Toy data: detection of the spiked transcripts

- Selection of candidates by thresholds on
  - fold change
  - t-statistics signals
  - t-statistics SLRs
  - percent of change calls indicating in- or decrease
- Count of the number of true and false positives
**Data processing**

- **Data sources:**
  - Expressions profiles: CEL files
  - Chip annotations: Excel sheet

- **Methods für absolute analysis:**
  - Available in `library(affy)` (Bioconductor project)
  - Functions `mas5()` and `mas5calls()` yield signals and detection p-values, respectively

- **Methods for comparison analysis:**
  - To my knowledge: not available from Bioconductor or other open source projects
  - cf. Affymetrix: *Statistical algorithms description document*
  - Own implementation (R code with integrated C functions)

**Data annotation**

- **Sample annotation:**
  - Entity-attribute-value (EAV) system
  - Sample names (CEL file names) → row names
  - Attributes names → column names
  - Matrix entries assign values of attributes to samples

- **Probe annotation:**
  - Mapping of probes to genes
  - Annotation of genes
    - e.g. cytoband, function (gene ontology), pathway (KEGG), references (PubMed)

**Data upload und retrieval**

- **ODBC is a generic interface to relational databases**
- **ODBC is supported by MS Access, PostgreSQL, MySQL, Oracle, …**
- **The library `RODBC` implements the ODBC database connectivity under R**

- RODBC allows an easy and generic database management including definition of tables, data upload and data retrieval
**oligoExpress - conclusion**

- Concise management of all information from Affymetrix absolute and comparison analysis
- Flexible sample annotation by an EAV system, analysis of the corresponding biological groups
- Compatibility to all common database systems by usage of the RODBC interface

**Toy data: detection of the spiked transcripts spikes with concentrations ≥ 1 pmol**

- Selection of candidates by thresholds on
  - fold change
  - t-statistics signals
  - t-statistics SLRs
  - percent of change calls indicating in- or decrease
- Count of the number of true and false positives