

GeWare:

**A data warehouse for
gene expression analysis**

T. Kirsten, H.-H. Do, E. Rahm

WG 1, IZBI, University of Leipzig

www.izbi.de, dbs.uni-leipzig.de

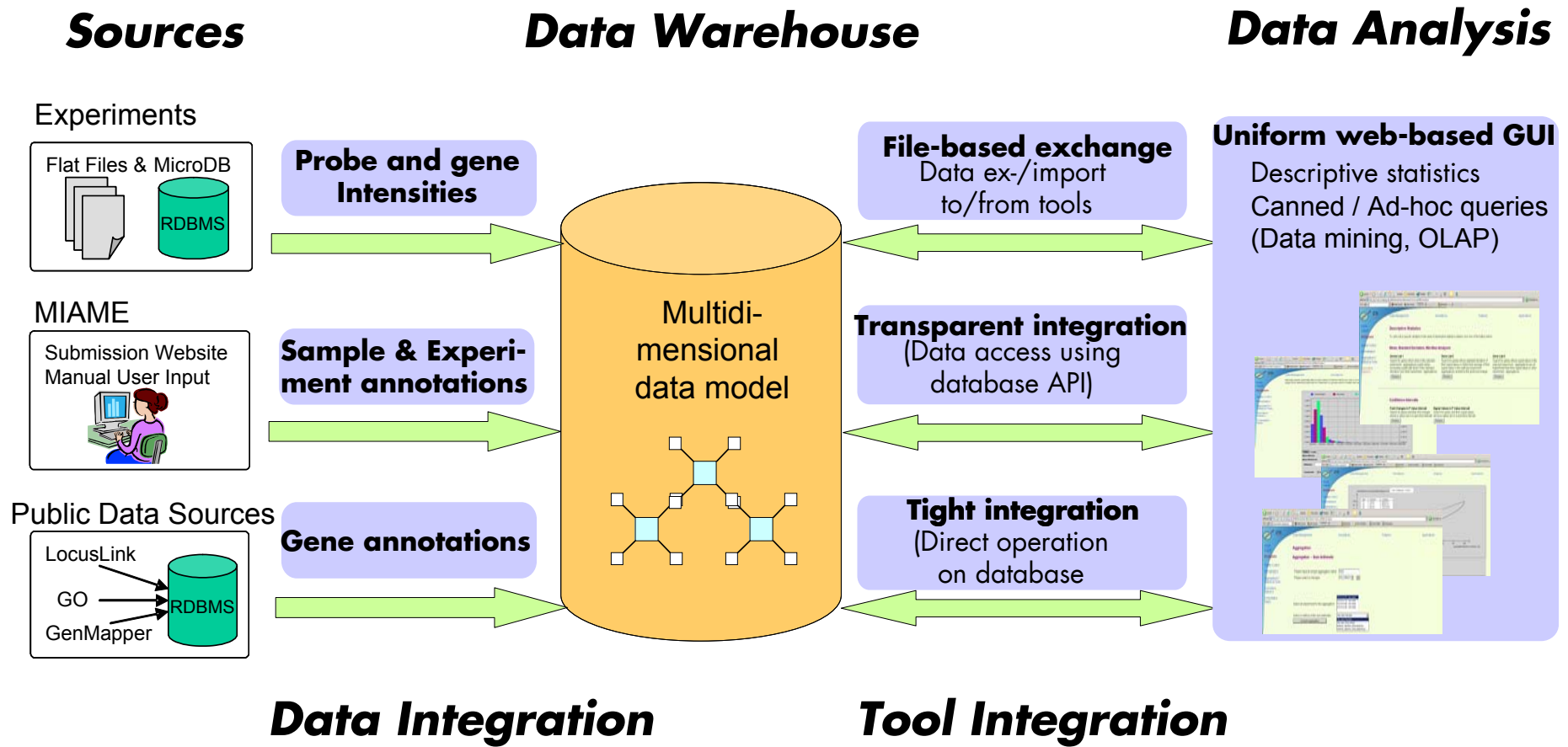
Outline

- Motivation
- GeWare Architecture
- Annotation Integration
- Expression Analysis
- Conclusion and Future Work

Motivation

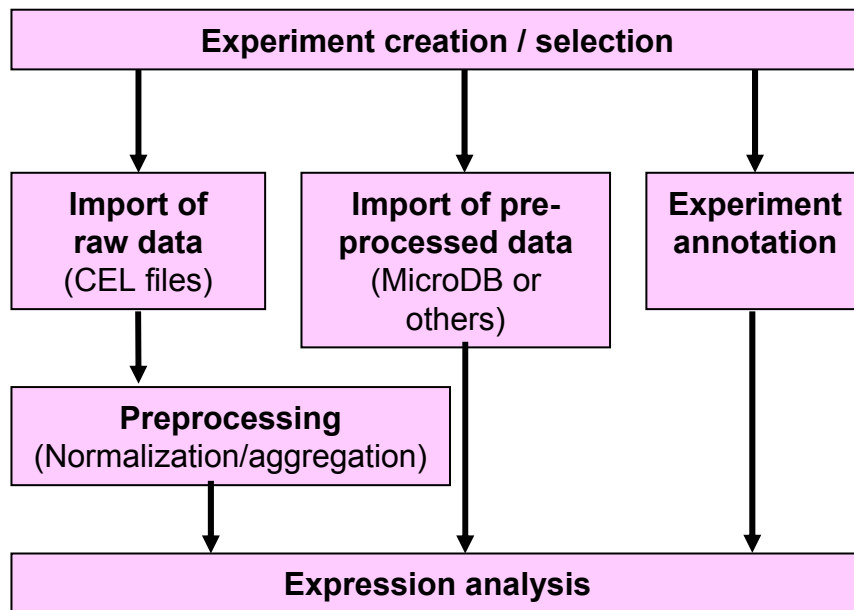
- Microarrays to measure expression of thousands of genes at the same time
- Various kinds of data with different characteristics and requirements
- Central data management and analysis platform
- Data Warehouse approach
 - Expression data import, e.g. from Affymetrix system
 - Fact tables to store both raw and derived data
 - Uniform specification of experiment annotations
 - Integration of gene annotations from public sources
 - Integration of analysis and data mining algorithms/tools

System Architecture

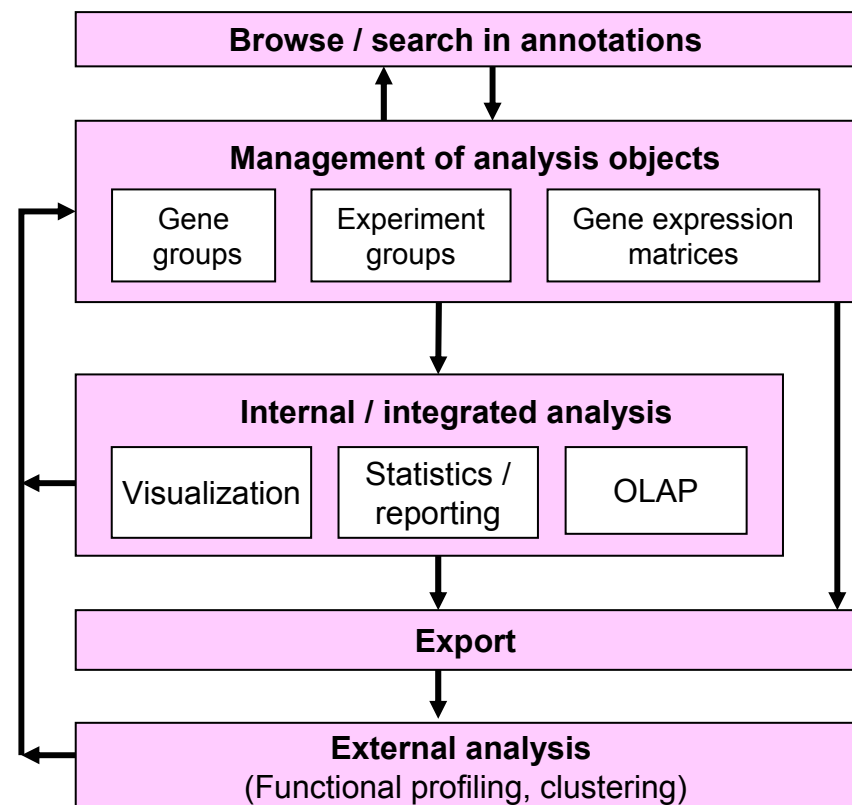


System Workflows

Import workflow



Expression analysis workflow



Experiment Annotation

- Uniform and consistent annotation
- Controlled annotation vocabularies
- Annotation templates as collections of annotation categories for which the annotation values has to be captured
 - Hierarchical arrangement of categories
 - Definition of MIAME compliant templates by biologists
- MAGE-ML export (data exchange)

Experiment Annotation Specification
Template: *Cell biological microarray experiments 1*
Submission: *Experiment1*

culture conditions

[<< previous page](#) [next page >>](#)

[back to the start page of the submission](#)

medium

medium type*

serum* fetal calf serum
 horse serum

antibiotics penicillin
 streptomycin

atmosphere

oxygen

carbondioxide

temperature

Experiment and Gene Groups

- Collections of objects with common patterns
- Input for reporting and further analysis
- Definition by
 - User selection
 - Search in annotation
 - Result of reporting and advanced analysis

- Example: p53 tumor

Gene Annotation Browsing

GeWare Organism

AND GeWare Probe Set Title

Add new Condition

Retrieve Data using the database path Ne

Save as gene group Save

Select?

<input checked="" type="checkbox"/>	1711_at	Homo sapiens	tumor protein p53 binding protein, 1
<input checked="" type="checkbox"/>	1859_s_at	Homo sapiens	Mdm2, transformed 3T3 cell double minute 2, p53 binding protein (mouse)
<input checked="" type="checkbox"/>	1860_at	Homo sapiens	tumor protein p53 binding protein, 2
<input checked="" type="checkbox"/>	1880_at	Homo sapiens	Mdm2, transformed 3T3 cell double minute 2, p53 binding protein (mouse)
<input checked="" type="checkbox"/>	1935_at	Homo sapiens	Mdm4, transformed 3T3 cell double minute 4, p53 binding protein (mouse)
<input checked="" type="checkbox"/>	1974_s_at	Homo sapiens	tumor protein p53 (Li-Fraumeni syndrome)
<input checked="" type="checkbox"/>	2008_s_at	Homo sapiens	Mdm2, transformed 3T3 cell double minute 2, p53 binding protein (mouse)
<input checked="" type="checkbox"/>	201746_at	Homo sapiens	tumor protein p53 (Li-Fraumeni syndrome)

Preprocessing

■ Preprocessing

- Generation of expression values on gene level
- Several predefined algorithms (Mas5, Rma, Li/Wong)
- P. by user defined methods for normalization, etc.

Probe Intensity Preprocessing

External Transformation - ArrayAnalyzer / R BioConductor

[List of available aggregations](#)

This function provides the possibility to summarize raw probe intensity values into gene signal values by using different methods which also includes background subtraction and normalization.

Following methods can be used:

- MAS5, developed by Affymetrix
- RMA, introduced by Terry Speed's group at UCSC
- LWF, Li/Wong's described model

1. **Step** Please input an unique name for the calculation:

2. **Step** Please select an aggregation method: Advanced

- Speeds RMA Transformation (R)
- Affy MAS5 Transformation (R)
- Speeds RMA Transformation (R)
- Full Li/Wong Transformation (R)
- Reduced Li/Wong Transformation (R)

A detailed comparison between the MAS 5.0 values that are computed by affy and by Affymetrix's software can be found at <http://stat-www.berkeley.edu/~bolstad/MAS5diff/Mas5difference.html>

Expression Analysis

- Several statistical reports used for analysis entry and outlier detection
 - Descriptive statistics (mean, sd, se ...)

Statistical Reports

Standard Reports -

Gene Group Name:

Please specify the filter criteria To download the results please use this [link](#)

In addition, annotation can be viewed for the selected probe sets.

1. Step: Select gene annotation

2. Step: Choose a defined gene

3. Step: Choose a defined gene

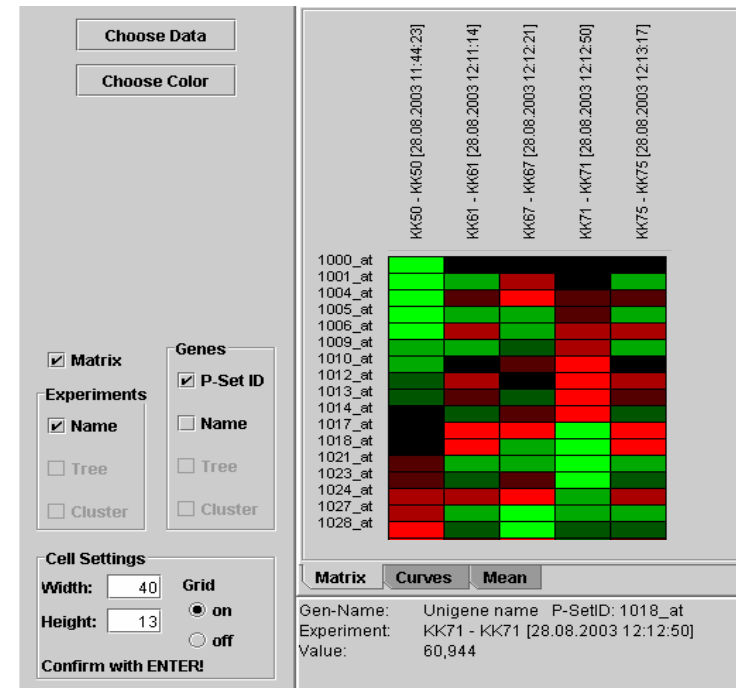
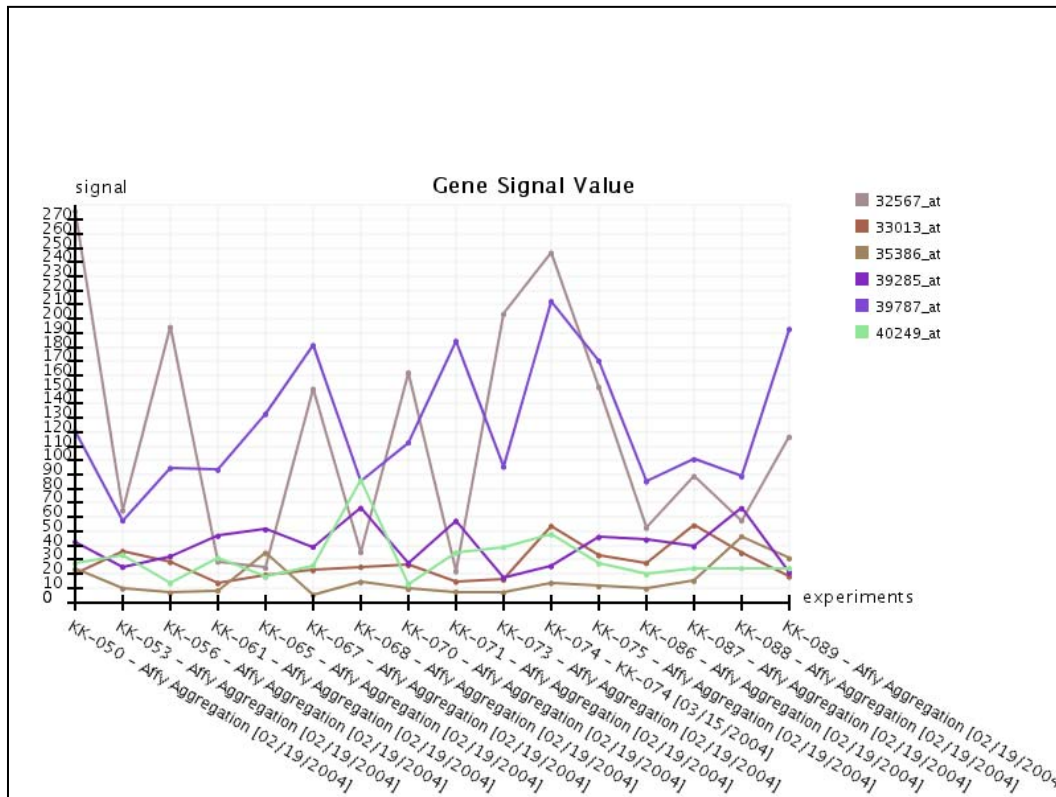
4. Step: Mark the deviation

5. Step: Specify z, if Range

Select?	Probe Set Name	Gene Symbol	Map Location	LocusLink	maximal Signal	Avg (Signal)	StdDev (Signal)
<input checked="" type="checkbox"/>	35083_at	FTL	19q13.3-q13.4	2512	22246.889	7126.647	1634.848
<input checked="" type="checkbox"/>	36130_f_at	MT2A	16q13	4502	20413.998	10935.716	2837.217
<input checked="" type="checkbox"/>	35928_at	TPO	2p25	7173	10014.709	5040.232	1104.781
<input checked="" type="checkbox"/>	609_f_at	---	---	---	8410.894	4051.537	1203.097
<input checked="" type="checkbox"/>	36455_at	COL9A3	20q13.3	1299	7092.540	220.447	197.615
<input checked="" type="checkbox"/>	870_f_at	---	---	---	7089.801	2501.577	637.314
<input checked="" type="checkbox"/>	37039_at	HLA-DRA	6p21.3	3122	6823.413	1876.503	535.567
<input checked="" type="checkbox"/>	41206_r_at	COX6A1	12q24.2	1337	6819.760	4022.020	576.162
<input checked="" type="checkbox"/>	1120_e_at	FIE1A2	3q28	1071	6708.081	3711.721	635.355

Visualization

- Different kinds of visualization of gene expression values (line, bar charts, heat maps)



Conclusions / Future Work

■ GeWare

- Management of a high volume of expression data
- Flexible and uniform experiment annotation
- Storing experiment and gene groups, expression matrices
- Different kinds of analysis, export
- Web-Interface:
<http://www.izbi.de/izbi/AG1/GeWare.html>

■ Future work

- Coupling with advanced analysis/ data mining routines
- Visualization extension