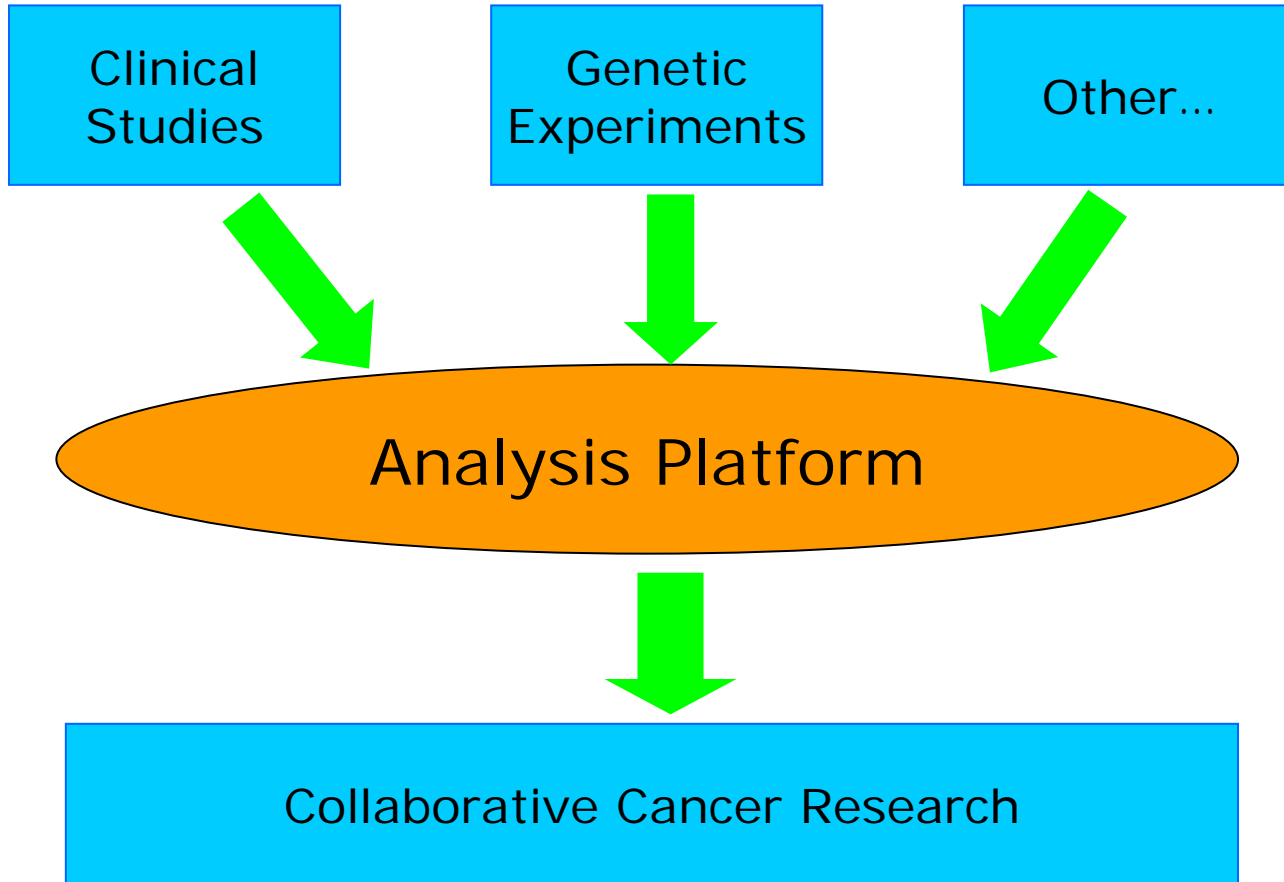


An integrated **analysis platform** for **experimental** and **clinical** data in modern cancer research studies

Jörg Lange, Toralf Kirsten, Erhard Rahm

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

Objective



Collaborative Cancer Research

○ **Clinical Studies**

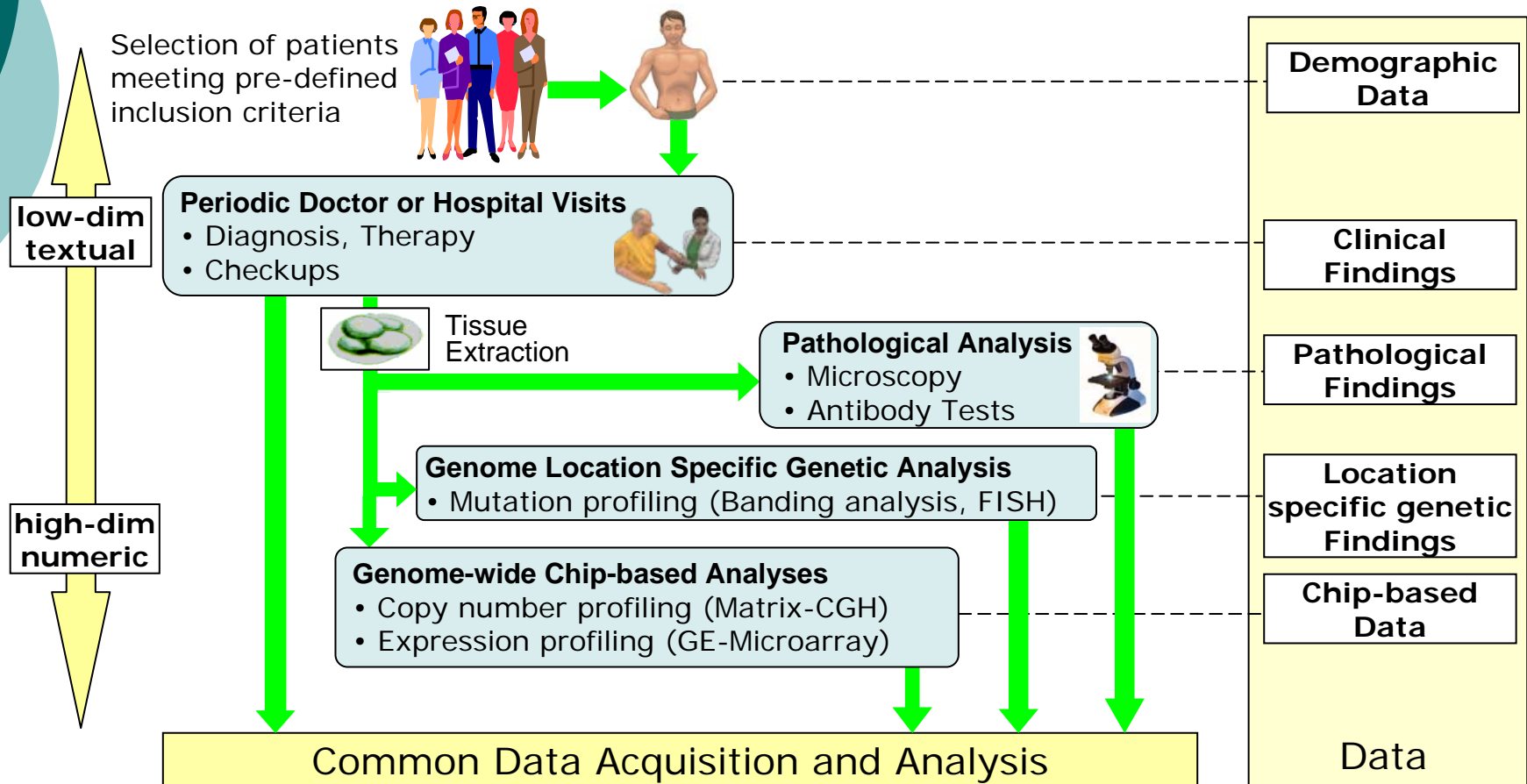
- Find best treatment
- Find factors that predict prognosis
- Improve diagnostic accuracy

○ **Recent development:** Combination of clinical and molecular-biological data including high throughput techniques, e.g. microarrays

○ Collaborative **Cancer Research Studies**

- Molecular Mechanism in Malignant **Lymphoma** with about 500 patients
 - German Glioma Network
- 10-20 network partners

Study Environment and Resulting Data



Requirements

- **Data acquisition** during complex workflows running in different hospitals and labs
 - Uniform data specification (metadata, values)
 - Autonomous data input (online data input)
- **Data integration**
 - Patient-related findings + chip-based data
 - Publicly available molecular-biological annotations
- **Privacy** aspects: Protect patients' privacy
- **Comprehensive** data analysis support
- Build on **existing** software for data management and analysis

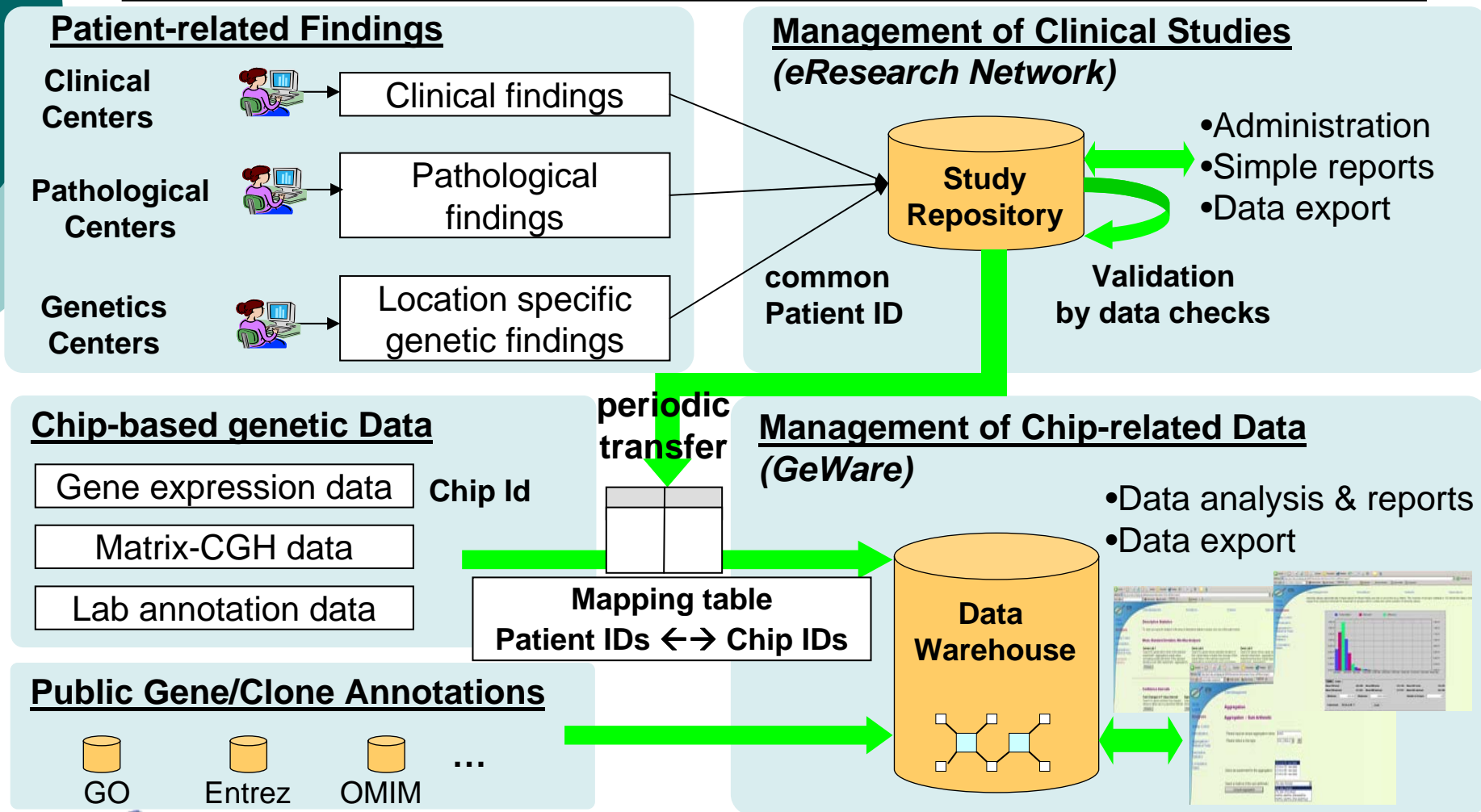
Selected Available Solutions

- Special solutions for clinical study management
 - Examples: eResearch Network, Oracle Clinical, ...
 - No support for array-based experimental data
- Specific systems maintaining data of molecular-biological experiments
 - Examples: ArrayExpress, GEO, SMD, GeWare, ...
 - Focused on microarray data, but no support for clinical annotations
- caBIG: Grid infrastructure developing by NCI
 - Comprehensive software collection e.g. caGrid-Browser, caArray
 - In development process (at time of decision)

Genetic Data Warehouse

- Handling, analysis and storage of large chip-based genetic data
 - Microarray-based **gene expression** data (Affymetrix)
 - **Matrix-CGH** data (Laboratory)
- Contains about 1700 preprocessed chips, thereof about 200 in MML
- Integrate **clinical data** from eResearch Network
- Integrate **public annotation** data using **SRS**
- Coupling with **statistic** tool **R** for powerful and flexible analysis
- Web-Client Application using JavaServlets
www.izbi.de/GEWARE

Integration Architecture



Linking Patient Data

- **Goal:** Utilization of uniform and comprehensive annotation for later analysis
- Focus-dependent annotation data in different clinical studies, e.g. Lymphoma
- **Annotation templates**
 - Collections of parameters for which the annotation values has to be captured
 - Definition of **MIAME-compliant** templates
 - Generic management of metadata and values
- Controlled **vocabularies** (predefined terms)

Combined Analysis (1)

Browse Experiment Annotation
Template: *MMML*

[back to the choice of annotation template](#)

Generate Query

Category Klinik Staging > Geschlecht LIKE male

and Category Klinik Staging > LDH > 240

and not Category Klinik Staging > Ann Arbor Sta LIKE IV

Your query is satisfied by the following 4 chips.

Experiment name	Chip type	
MPI-001	HG-U133A	<input type="button" value="Browse Annotation"/>
MPI-002	HG-U133A	<input type="button" value="Browse Annotation"/>
MPI-004	HG-U133A	<input type="button" value="Browse Annotation"/>
MPI-005	HG-U133A	<input type="button" value="Browse Annotation"/>

The list of the aggregations of these experiments with chip type HG-U133A:

- MPI-001 - DS_1-5_07.2005_V3
- MPI-002 - DS_1-5_07.2005_V3
- MPI-004 - DS_1-5_07.2005_V3
- MPI-005 - DS_1-5_07.2005_V3

unbekannt

andere:

Search for relevant mol.-biol. data in GeWare using clinical data

Group data for later reuse in other analysis

Combined Analysis (2)

Additional annotation class label:
Stage (location infected Lymph nodes)

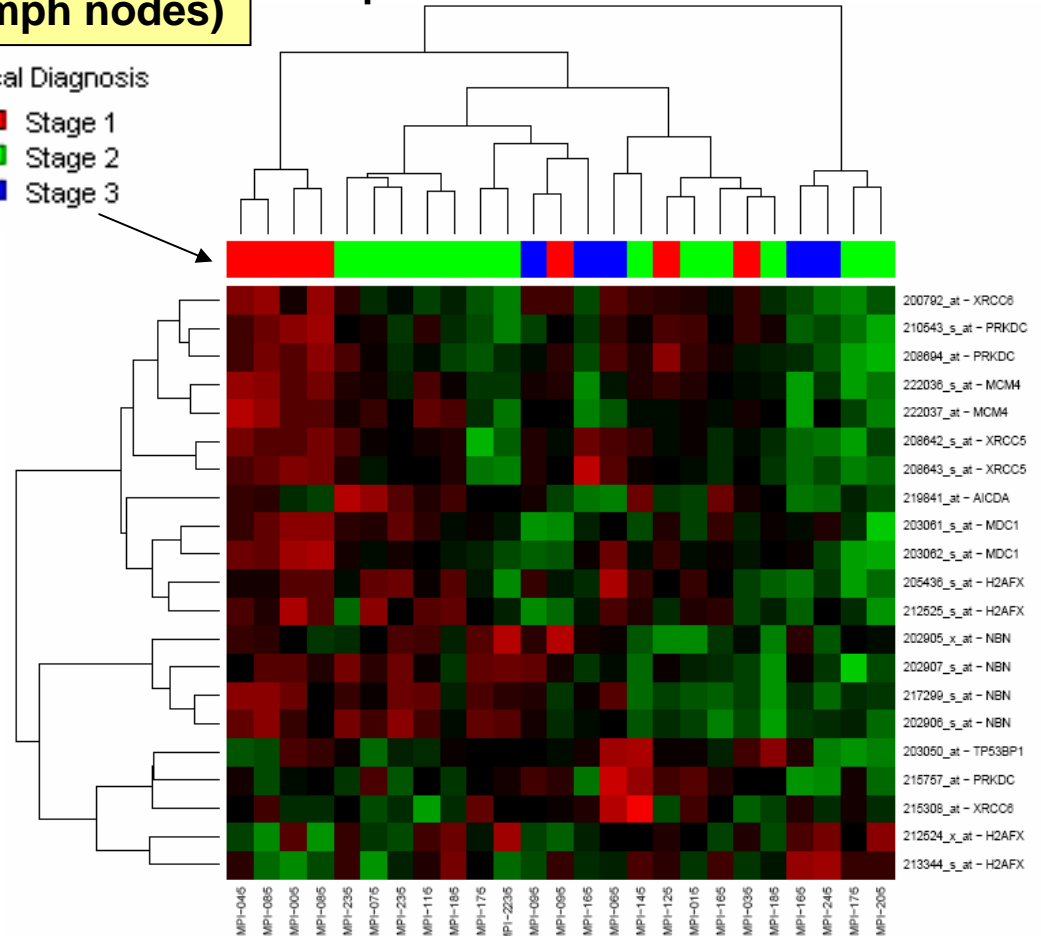
Chips / Patients

Clinical Diagnosis

- Stage 1
- Stage 2
- Stage 3

Heatmap including
hierarchical
cluster analysis

Genes



Combined Analysis (3)

- Report example

Statistical Reports

Chip and Gene Correlation - Gene Correlation

This report has following filter conditions:

Chip group: 10Chips
Gene group: housekeeping_genes
Correlation coefficient limitation: >=abs(0.50)
Single gene: 202431_s_at

Gene Group Name: Save as gene

To download the results please use this [link](#)

Biological Annotation

Select?	Probe Set Name	Gene Symbol	Chip Type	Organism	Gene Description	Correlation Coefficient
<input checked="" type="checkbox"/>	202431_s_at	MYC	HG-U133A	Homo sapiens	v-myc myelocytomatosis viral oncogene homolog (avian)	1.0
<input checked="" type="checkbox"/>	201217_x_at	RPL3	HG-U133A	Homo sapiens	ribosomal protein L3	0.9108847544823635
<input checked="" type="checkbox"/>	211073_x_at				teir	0.908144670346516
<input checked="" type="checkbox"/>	212039_x_at					0.9067696871181317
<input checked="" type="checkbox"/>	211666_x_at			sapiens	L3	0.8802670324105074
<input checked="" type="checkbox"/>	208646_at	---	HG-U133A	Homo sapiens	Homo sapiens PRO2640 mRNA, complete cds	0.8683325176758565

Probe sets of the group sorted by correlation

Conclusions

- Analysis platform integrating clinical and molecular-biological data
- Build on two existing systems
 - Uniform and autonomous input of findings in eRN
 - Considering privacy aspects: pseudonymous patient ID
 - Management of molecular-biological data in GeWare
 - Periodic data transfer by using a mapping table
 - Combined data analysis in GeWare
- Application in large collaborative cancer research studies

Acknowledgement

- IMISE, Univ. Leipzig
 - Markus Löffler
 - Hilmar Berger
 - Maciej Rosolowski
- IZBI, Univ. Leipzig
 - Hans Binder
- Members of the collaborative cancer research studies



Thank you for your attention!