

An integrated platform for analyzing clinical and molecular-biological data in collaborative research projects

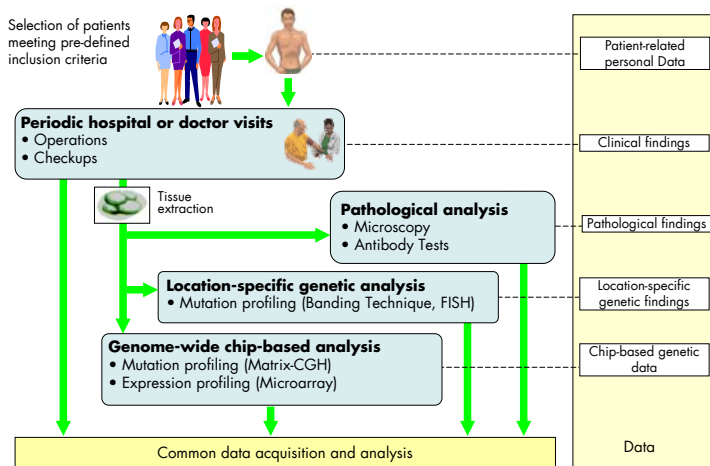
Jörg Lange^{1,3}, Toralf Kirsten¹, Erhard Rahm^{1,2}, Hilmar Berger³, Markus Löffler^{1,3}

¹ Interdisciplinary Center for Bioinformatics
http://www.izbi.de

² Database Group of the Department of Computer Science
http://dbs.uni-leipzig.de

³ Institute for Medical Informatics, Statistics and Epidemiology
http://www.imise.de

Project requirements for data management and analysis



Central data management and analysis platform for collaborative research projects funded by "Deutsche Krebshilfe" (German Cancer Aid)

- Molecular mechanisms in **malignant lymphoma** (MMML)¹
- German **Glioma Network**²

Project Requirements

Data Acquisition during complex workflows at multiple hospitals / labs

- Uniform metadata and data specifications
- Validate input data to minimize errors and missing values

Data Integration

- Clinical and pathological findings
- High-dimensional results of molecular-genetic examinations
- Molecular-genetic annotations from public databases

Data privacy and protection: anonymous patient identifier

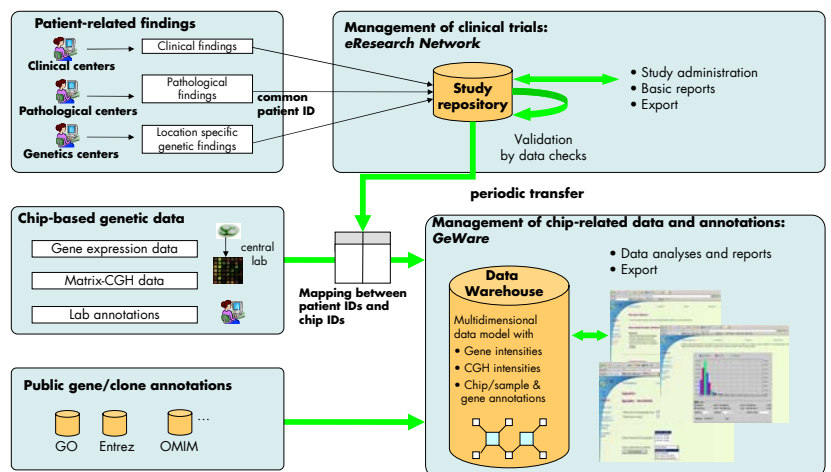
Comprehensive data analysis support

Usage of running software for study management and analysis

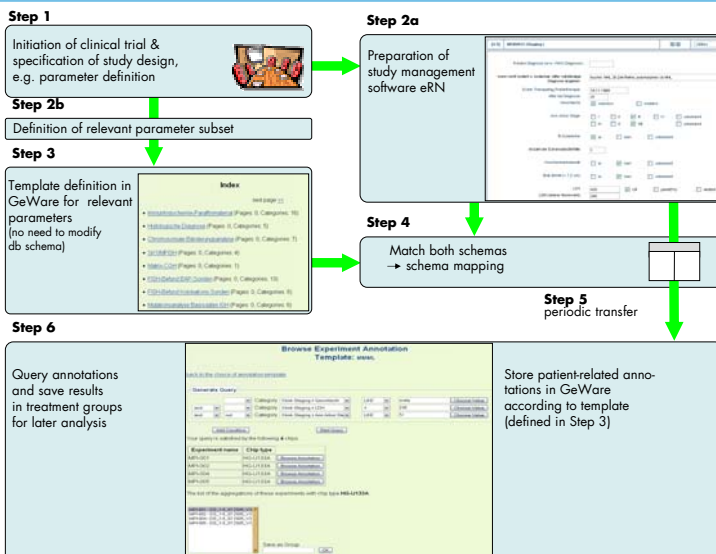
- 1 <http://www.lymphome.de/en/Projects/MMML/index.jsp>
- 2 <http://www.gliomnetzwerk.de>

Architecture of the analysis platform

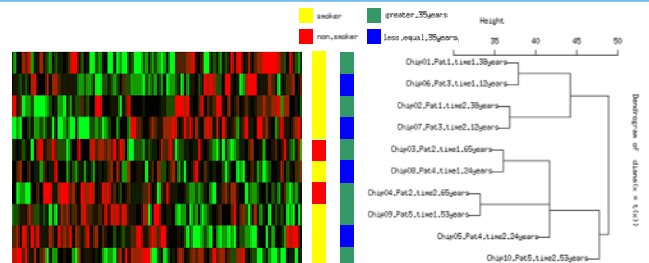
- All patient-related finding data are stored by study management system **eResearch Network**
 - Consistency checks to find imbalances or missing data
 - Basic reports for input management
- Chip-based genetic data of gene expression and matrix CGH analyses managed in **GeWare** data warehouse
 - Multidimensional data model with intensity values in fact tables
 - Laboratory annotations
 - Gene and clone annotations from publicly available databases like GO - integrated by a query mediator [KI05]
- Combined clinical-genetic analysis by transferring relevant patient-related findings from **eRN** to **GeWare** [KLR06]
 - Mapping between patientIDs and chip-IDs
 - Periodic data transfer for current data



Transfer of patient-related data into warehouse



Combined data analysis: example



- Patient-related data is used as classifier label in a heatmap visualization including hierarchical clustering

References

- [KLR06] Kirsten, T.; Lange J.; Rahm E.: An integrated platform for analyzing molecular-biological data within clinical studies, EDBT - Workshop on Information Integration in Healthcare, Munich, 2006 [submitted]
- [KI05] Kirsten, T.; Do, H.-H.; Rahm, E.; Körner, C.: Hybrid Integration of molecular biological Annotation Data. Proc. 2nd Int. Workshop on Data Integration in the Life Sciences, San Diego, Springer LNBI 3615, 2005