

# *Erweiterungen in GeWare durch MMML*

Jörg Lange  
IZBI AG1

## Inhalt

---

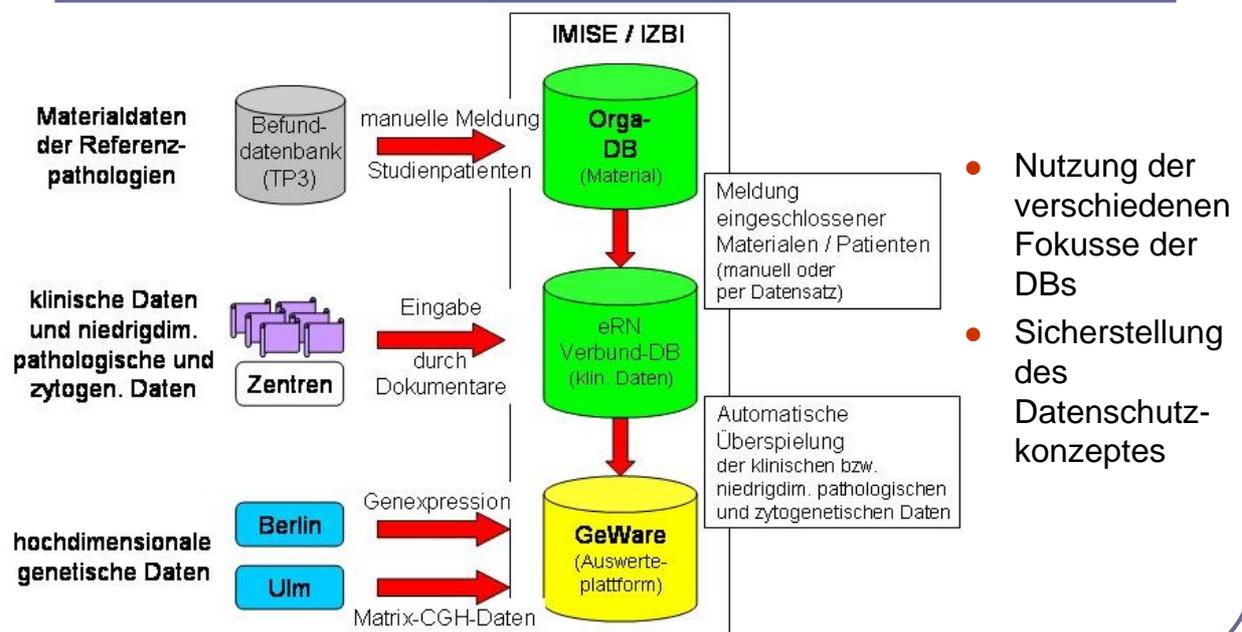
- Verbundprojekt MMML
- Matrix-CGH als neuer Gen-Datentyp
- Vorstellung neuer Module in GeWare
  - Import und Speicherung der CGH-Daten
  - Gruppenmanagement
  - Export von Gendaten
  - Visualisierungen
- Import der klinischen Daten von MMML aus eRN als Annotationen

# Molekulare Mechanismen der malignen Lymphome (MMML)

- **Verbundprojekt** mit über 10 angeschlossenen Zentren
- Ziel: Verbesserung von Prognose und Therapie von **Lymphomen** durch Nutzung und Erprobung neuer **molekulargenetischer** Strategien im Verbund mit klassischen **histopathologischen** und **klinischen** Ansätzen
- zentrale Speicherung der generierten und erhobenen Daten in **Datenbanken** am IMISE/IZBI
- Data Warehouse **GeWare** – Analyseplattform für hochdimensionale und niedrigdimensionale genetische Daten
  - Genexpressionsdaten – Funktionalität größtenteils vorhanden
  - **Matrix-CGH-Daten** – neuer Gendatentyp
- <http://www.lymphome.de/Projekte/MMML/index.jsp>

3

# Systematischer Aufbau der Datenbanken in MMML



4

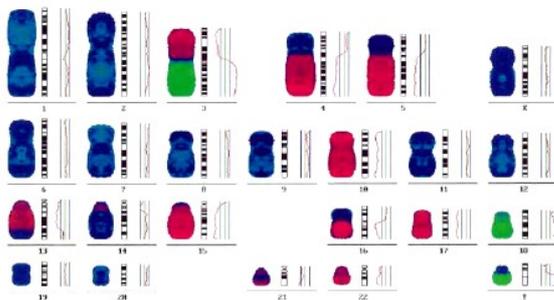


# Matrix-CGH

- „Matrix-based comparative genomic hybridization“
- Ziel: **Erkennung von tumorspezifischen DNA-Veränderungen**
- Methode: vergleichende Analyse von Tumor-DNA aller Fälle mit Kontroll-DNA nach genetischer Veränderungen
- Untersuchung des Erbgutes nach:
  - **Amplifikationen** d.h. erhöhte Kopienzahl in Tumor-DNA
  - **Deletionen** d.h. Allelverlust in Tumor-DNA

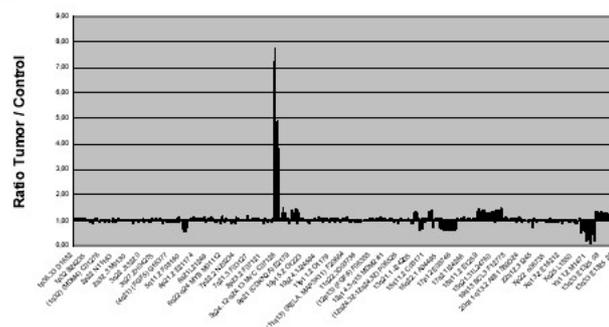
7

# Matrix-CGH (2) – Resultate



Summenkaryogramm  
mit Ratio-Kurve

## Identifikation der Gene



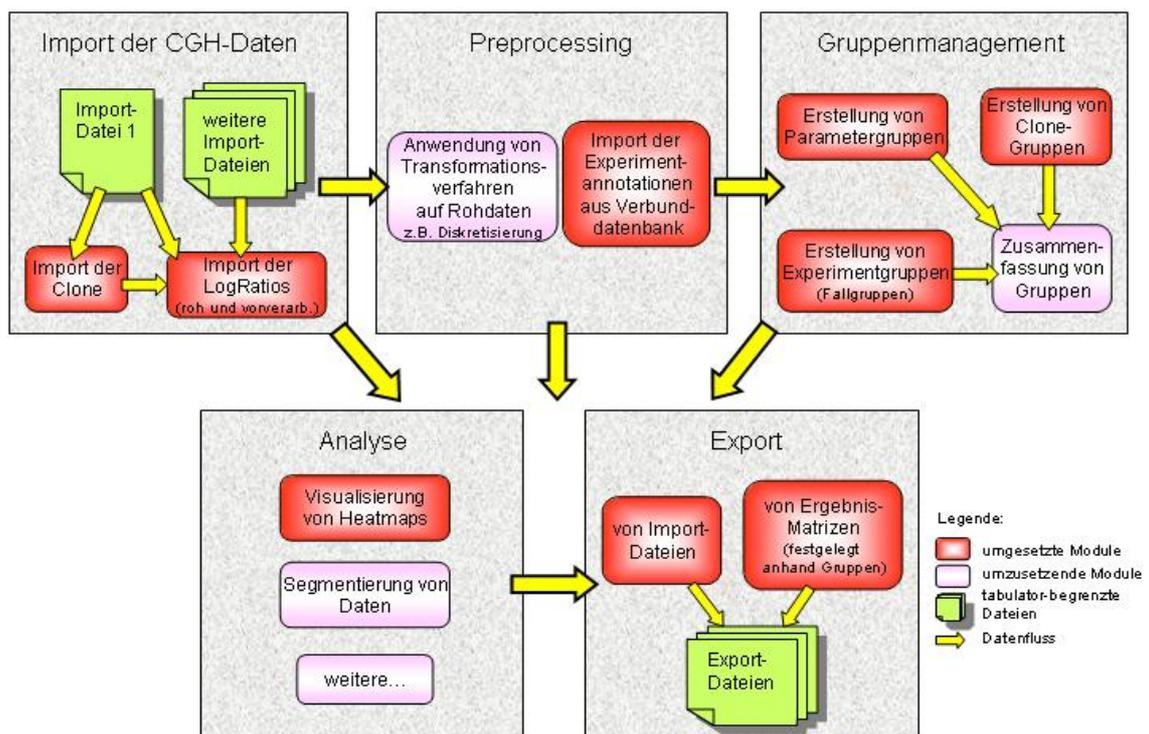
8

## Matrix-CGH (3)

- **Prinzip:**
  - Isolation der Metaphasenchromosomen aus Tumor und Kontrollgewebe
  - Duplikation der Gene durch Bakterien -> **BAC-Clone** (Bacterial Artificial Chromosomes)
  - Hybridisation der fluoreszenzmarkierten Tumor- und Kontroll-Clone
  - Intensitätsmessung der Fluoreszenz
  - Generierung von logarithmierten, numerischen Ratios (**Log-Ratios**) aus Rot/Grün-Ratios
- **Problem:** Chips werden (noch) von jedem Labor selbst hergestellt -> kein allg. Chiptyp

9

## CGH-Workflow in GeWare



10

# Import von CGH-Daten

- aus Tabulator-begrenzten Dateien
  - erste Spalte: Clone
  - erste Zeile: Experimente
  - Inhalt: Log-Ratios
- Aufspaltung des Importes
  - Import der Clone zur Bildung eines Chiptyps
    - Parsing zur Generierung von Clone-Annotationen möglich (Chromosom, Band, Position, Clone-ID, Zusatz)
  - Import der Log-Ratios
    - Inklusive/Exklusive Anlegen der Experimente
    - Roh- und/oder vorverarbeitete Daten durch Angabe einer Preprocessing-Methode

11

# Import von CGH-Daten (2)

**Load Matrix-CGH values of various experiments**

**Matrix-CGH Import - Load Matrix-CGH Values**

This form inserts the Matrix-CGH log ratios of a tab delimited file into the warehouse. For this, you have to select the chip type which contains the Matrix-CGH clones, to choose preprocessing method of generated log ratios and to specify the experiment series which should include the chip experiments of your data file.  
*NOTE: Due to the large size of Matrix-CGH data file the import can take a few minutes. Please have patience.*

**1. Step** Do the experiments you want to import from the file already exist in GeWare?  
*NOTE: Choose yes if you already integrated the raw values and want to import the data of another preprocessing method.*  ja  nein

**2. Step** Please specify the Matrix-CGH tab-delimited data file including the log ratios.  

**3. Step** Please select the preprocessing method the log ratios were transformed with.

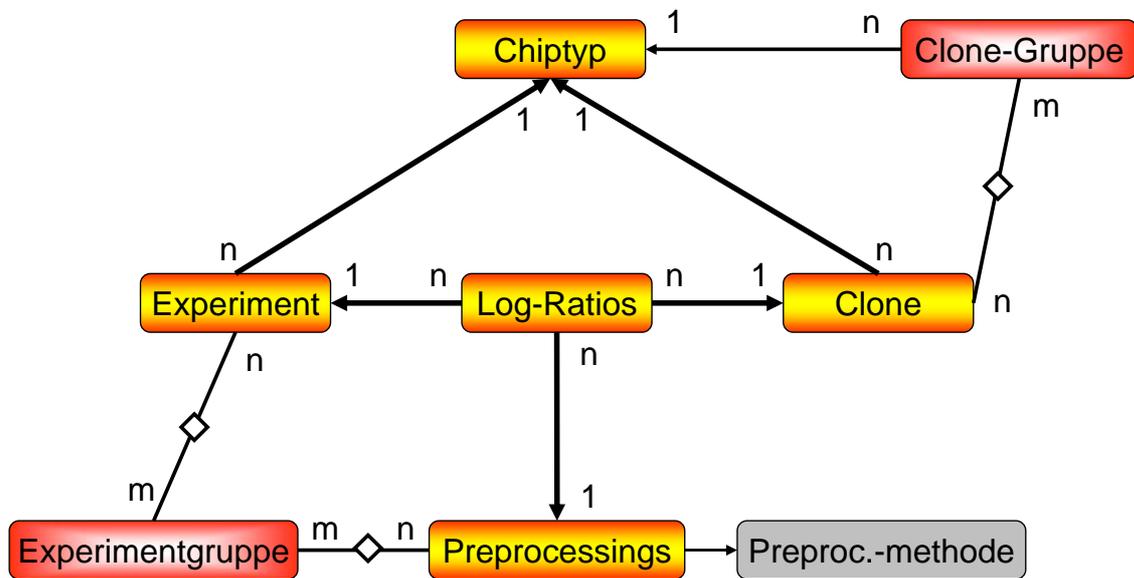
**4. Step** Please insert a name of the experiment series to characterize the set of chip experiments within the file.  
*NOTE: This name must be unique.*

**5. Step** Please select the chip type which contains the clones.  
*NOTE: The clone names in the file must be equal to clone codes of the chip type in the database.*

**6. Step** Please select the annotation template to which you want to annotate your experiments.

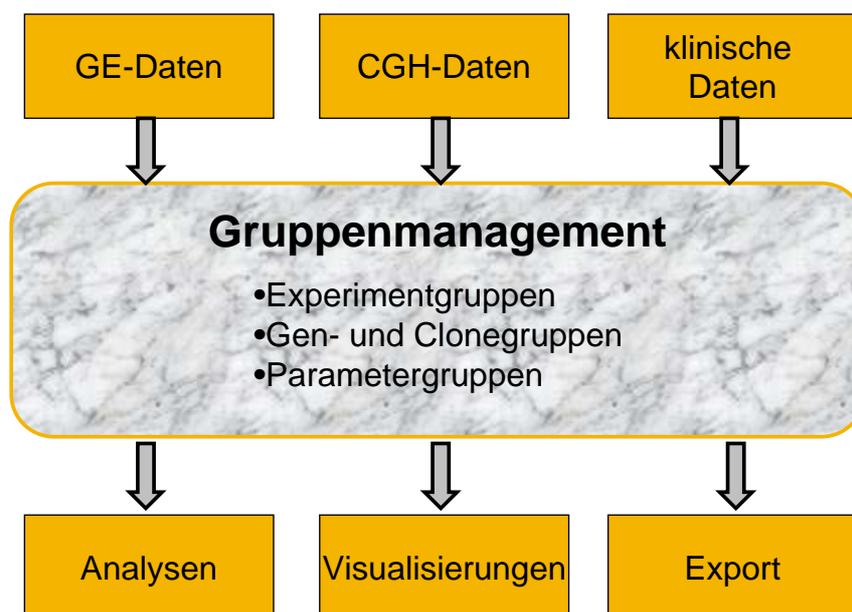
Optional Here you can specify the name of the preprocessing.  
*NOTE: If empty, GeWare uses the combination of the experiment series, the preprocessing method and the present date.*

## Speicherung der CGH-Daten



13

## Gruppenmanagement



14

## Gruppenmanagement (2)

- Experiment-Gruppen
  - Erstellung bei Genexpressions- und CGH-Daten
  - Angabe: Chiptyp, Experimentserien und infolge die Experimentaggregationen (Experiment-Preprocessing-Tupel)
- Clone-Gruppen bei CGH-Daten
  - Beschränkung der Clone eines Chiptyps
  - Auswahlhilfe: Filter über die Chromosomen und folgend über das chromosomale Band, Auswahl jeweils aller ist möglich
- Parametergruppen
  - Kollektion ausgewählter Experimentannotationen eines Annotationsschemas
- Gengruppen bei Genexpressionsdaten

15

## Experimentgruppen

### Experiment Group Definition / Modification

#### Experiment Groups - Definition/Modification/Deletion

Please use the the following steps to define a new group of experiments.

**Experiment group name:** Series2  
**Chip type:** TestChip  
**Experiment type:** Matrix-CGH

**1. Step:** Describe the group briefly.

only test

**2. Step:** Select the experiment series (use Shift/Ctrl to mark):

UlmSerie2  
UlmSerie3  
UlmSerie4

**3. Step:** Select the experiment - aggregations:  
*NOTE: In general they are preprocessings.*

11180 - UlmSerie2 - chip-wise mode normalization + automatic aberration analysis - 2005-04-05  
12702 - UlmSerie2 - chip-wise mode normalization + automatic aberration analysis - 2005-04-05  
12741 - UlmSerie2 - chip-wise mode normalization + automatic aberration analysis - 2005-04-05  
14006 - UlmSerie2 - chip-wise mode normalization + automatic aberration analysis - 2005-04-05  
14779 - UlmSerie2 - chip-wise mode normalization + automatic aberration analysis - 2005-04-05  
14804 - UlmSerie2 - chip-wise mode normalization + automatic aberration analysis - 2005-04-05  
15017 - UlmSerie2 - chip-wise mode normalization + automatic aberration analysis - 2005-04-05  
15027 - UlmSerie2 - chip-wise mode normalization + automatic aberration analysis - 2005-04-05  
15047 - UlmSerie2 - chip-wise mode normalization + automatic aberration analysis - 2005-04-05  
15902 - UlmSerie2 - chip-wise mode normalization + automatic aberration analysis - 2005-04-05

Proceed

16

# Clone-Gruppen

## Clone Group Definition

### Clone Groups - Definition/Modification/Deletion

**1.Step:** According the large number of clones in a chip type please specify the chromosome and the chromosomal band to preselect the potential entries of your clone group.

*NOTE: There are possible limitations if the clone annotations exist in your selected chip type.*

Chromosome  Chromosomal band

**2.Step:** Please select the clones you want to subsume to the group 'JL-11q'

<input checked="" type="checkbox"/> All	Clone name	Chromosome	Chromosomal band	Clone-ID	Position of first base	Note
<input checked="" type="checkbox"/>	11q_57813613_RP11-10E21_939	11	q	RP11-10E21	57813613	939
<input checked="" type="checkbox"/>	11q12_59405724_RP11-872D17_SSRP1; PRG3;	11	q12	RP11-872D17	59405724	SSRP1; PRG3;
<input checked="" type="checkbox"/>	11q_61539292_RP11-384C21_940	11	q	RP11-384C21	61539292	940
<input checked="" type="checkbox"/>	11q12-q13.1_62568353_RP11-729B4_MS4A2; M	11	q12-q13.1	RP11-729B4	62568353	MS4A2; M
<input checked="" type="checkbox"/>	11q_64108196_RP3-466A11_941	11	q	RP3-466A11	64108196	941
<input checked="" type="checkbox"/>	11q_65027264_RP11-163K24_942	11	q	RP11-163K24	65027264	942
<input checked="" type="checkbox"/>	11q_65676845_Rp11-783K16_Bad;ESRRA;PRDX5	11	q	Rp11-783K16	65676845	Bad;ESRRA;PRDX5
<input checked="" type="checkbox"/>	11q12.3_65676845_RP11-783K16_BAD;RPS6KA4	11	q12.3	RP11-783K16	65676845	BAD;RPS6KA4

17

# Parametergruppen

## Parameter Group Definition

### Common Group Management - Parameter Groups

**1. Step:** Here you can limit the number of categories by using the page filter.

Page:

**2. Step:** Please select the annotation categories you want to subsume to the parameter group 'MMML-PG4'.

<input type="checkbox"/> All	Annotation Template	Page	Annotation Category
<input type="checkbox"/>	MMML	Immunhistochemie-Paraffinmaterial	CD5 - Status
<input type="checkbox"/>	MMML	Immunhistochemie-Paraffinmaterial	CD5 - Anteil markierter Zellen
<input type="checkbox"/>	MMML	Immunhistochemie-Paraffinmaterial	CD10 - Status
<input type="checkbox"/>	MMML	Immunhistochemie-Paraffinmaterial	CD10 - Anteil markierter Zellen
<input type="checkbox"/>	MMML	Immunhistochemie-Paraffinmaterial	CD20 - Status
<input type="checkbox"/>	MMML	Immunhistochemie-Paraffinmaterial	CD20 - Anteil markierter Zellen
<input type="checkbox"/>	MMML	Immunhistochemie-Paraffinmaterial	BCL6 - Status
<input type="checkbox"/>	MMML	Immunhistochemie-Paraffinmaterial	BCL6 - Anteil markierter Zellen
<input type="checkbox"/>	MMML	Immunhistochemie-Paraffinmaterial	BCL2 - Status
<input type="checkbox"/>	MMML	Immunhistochemie-Paraffinmaterial	BCL2 - Anteil markierter Zellen

18

## Export von Gendaten

- Export der importierten Dateien der eigenen Nutzergruppe
- Export einer Matrix von Signalwerten / LogRatios, begrenzt von einer ausgewählten Experiment- und einer Gen- bzw. Clone-Gruppe desselben Chiptyps
  - Zieldatei tabulatorbegrenzt und komprimiert
  - Optional kann zudem eine Parametergruppe angegeben werden

19

## Export von Gendaten (2)

**Gene Intensität**

Please use the the local computer.  
First the file name (step 2). The following expressions data of the genes. After you If you want to download

1. Step: Please, sp
2. Step: Please, sp
3. Step: Please, cf
4. Step: Please, cf
5. Step: You can o

Probeset	Affy-Expt01-R1 - Affy Replicate 1 [2003-12-22]				
experimenter > Name	Lange	Lange			
experimenter > First name	Jörg	Jörg			
experimenter > Institute	I2BI	I2BI			
cell characteristics > species	Homo Sapiens	Homo Sapiens			
cell characteristics > organ	brain	brain			
cell characteristics > disease	cancer	cancer			
200665_s_at	300.478	326.797	323.632	511.828	
203471_s_at	797.357	1226.160		1262.462	1483.111
203508_at	10.622	75.604	77.460	18.339	
204205_at	16.549	88.808	98.259	31.723	
204417_at	119.624	153.520	149.705	172.293	
204430_s_at	411.505	459.675	460.454	643.301	
204513_s_at	6.064	152.918	147.006	22.273	
204563_at	11.262	69.725	69.989	27.306	
204836_at	29.705	105.352	99.639	51.976	
204912_at	36.678	120.781	120.005	53.557	
204951_at	649.607	794.406	792.424	1261.147	
204959_at	1.033	106.329	110.412	1.395	
205267_at	150.450	221.499	222.608	285.440	
205291_at	135.148	214.998	245.403	142.913	
205398_s_at	213.000	412.182	454.594	321.988	
205569_at	28.811	103.972	100.654	54.146	
205692_s_at	124.542	234.568	221.573	130.569	
205790_at	202.912	309.722	340.425	311.108	
206060_s_at	238.031	311.954	300.284	281.421	

the system to your

in the output file

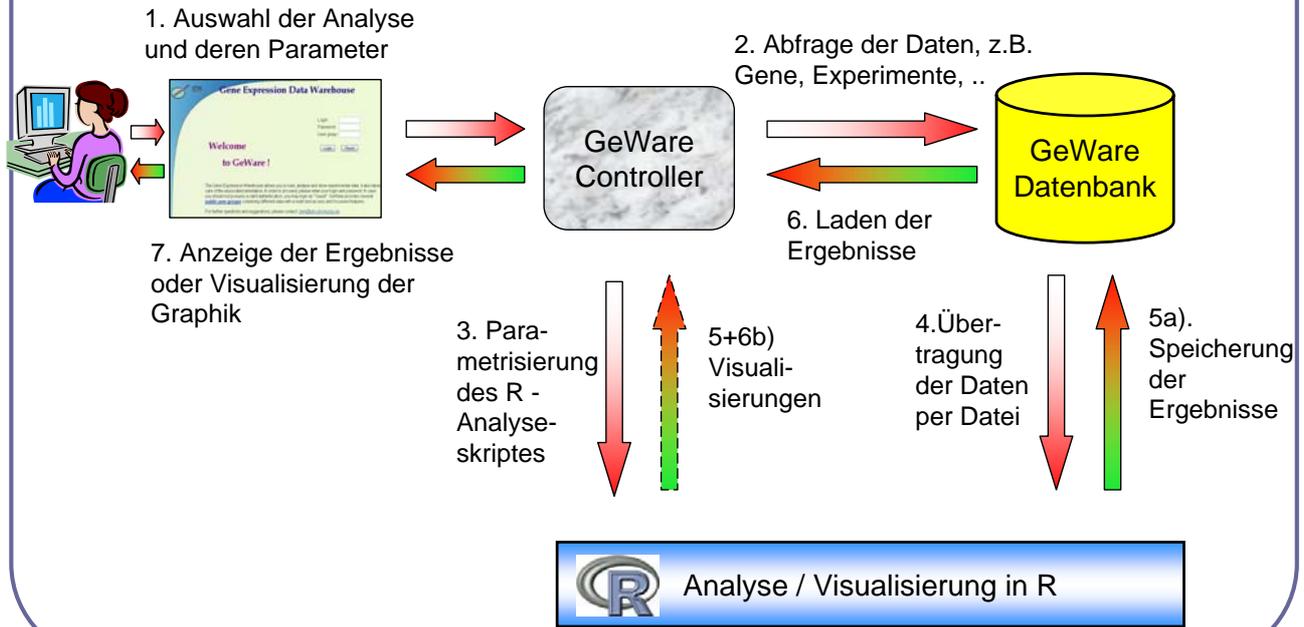
ou to focus on

exported in front of

export data.

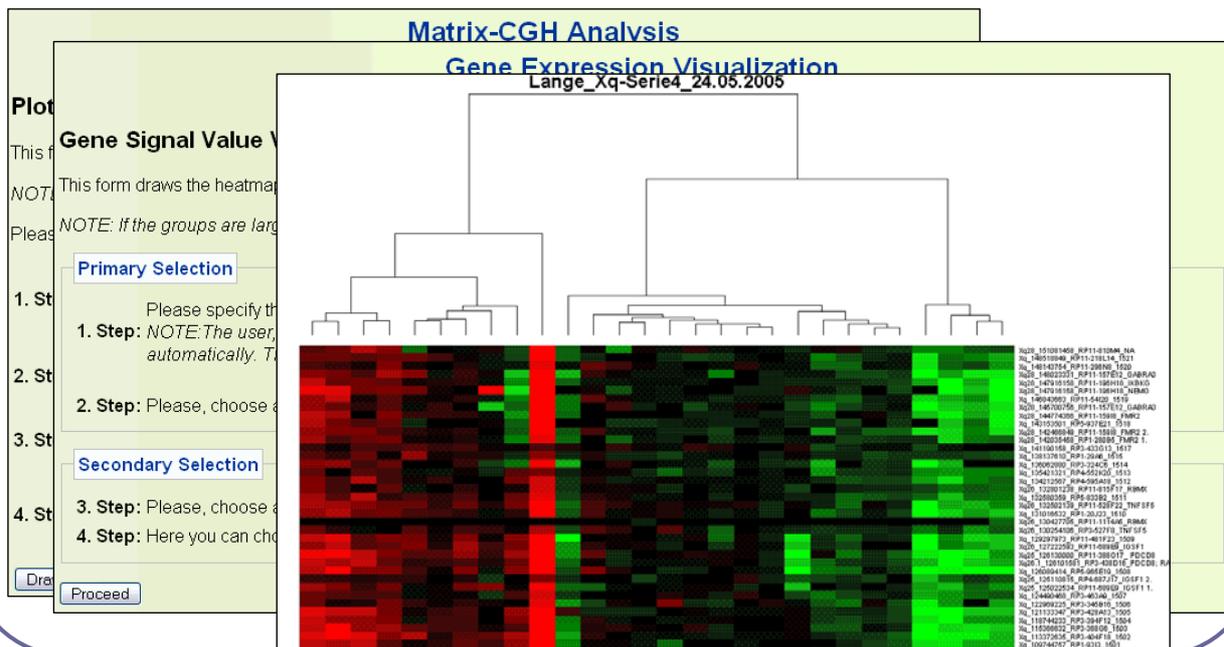
20

# Datenfluss während einer Analyse



21

# Visualisierung in Heatmaps



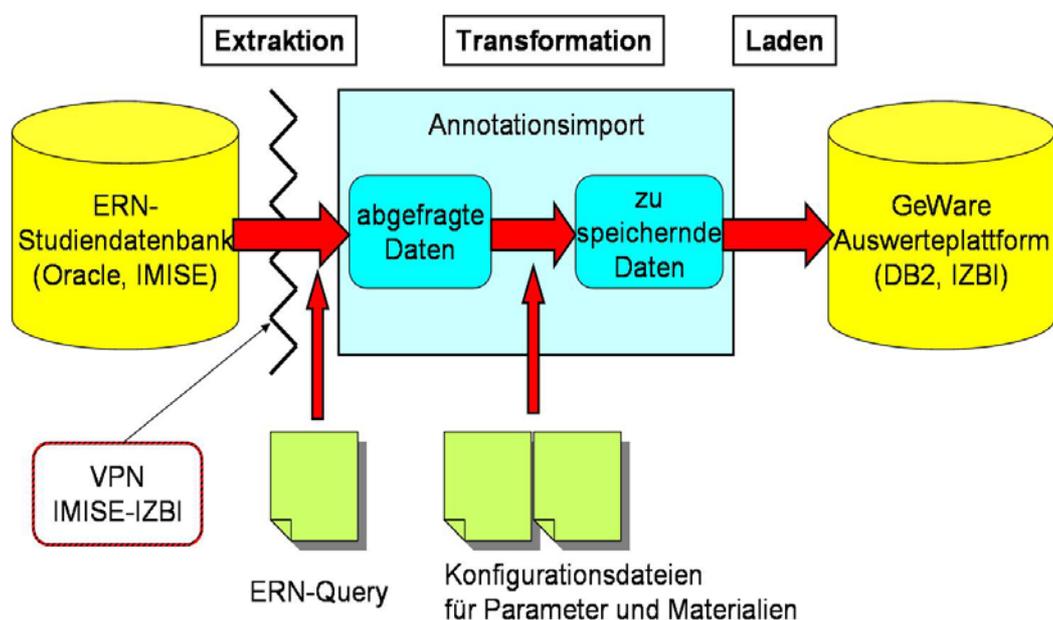
22

## Import von klinischen Daten von MMML als Annotationen

- In den Verbundprojekten MMML und Gliomnetzwerk werden die klinischen Daten in einer **eResearchNetwork (eRN)** Studiendatenbank am IMISE gespeichert
  - die für die Auswertung in **GeWare** relevanten Daten werden **tagesaktuell** in das Data Warehouse übernommen
  - klinische, histopathologische und zytogenetische Daten werden automatisch einem Experiment zugeordnet
- => gilt sowohl für Matrix-CGH als auch für Genexpressionsexperimente

23

## Schnittstelle des Imports der klinischen Daten von MMML (Schema)



24

# Aufbau der Konfigurationsdateien

- Parametertransformation (Items)

ERN-Parameter	GeWare-Schema	GeWare-Seite	GeWare-Parameter
Item1	Schema	Page1	Kategorie1
...			

- Materialtransformation (Fälle)

ERN-Material (Fall)	GeWare-Experimentserie	GeWare-Experimentname
B-12345	Berlin 1	MPI-75328
...		

Bemerkung: Sind für einen Fall in der Studiendatenbank sowohl Genexpressions- als auch Matrix-CGH-Daten in GeWare vorhanden, so werden die klinischen Annotationen in beide Experimenten eingefügt.

# Beispielimport klinischer Daten

[1/1] MOD0032 (Staging)

**Experiment Annotation**  
 Template: *MMML*  
 Experiment: *MPI-013*

**Klinik Staging**

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

Primäre Diagnose (erw. WHO-Diagnose):

wenn nicht kodiert o. kodierbar, bitte vollständige Diagnose angeben:

Erster Therapietag Primärtherapie:

Alter bei Diagnose:

Geschlecht:

Ann Arbor Stage:

B-Symptome:

Anzahl der Extranodal-Befälle:

Knochenmarksbefall:

Bulk-Befall (> 7,5 cm):

LDH:

LDH (oberer Normwert):

Primäre Diagnose	<input type="text"/>
Diagnose Freitext	hochm. NHL, B-Zell-Reihe, polymorphes cb NHL
Erster Therapietag Primärtherapie	1999-11-18 00:00:00.0
Alter bei Diagnose	37
Geschlecht	male
Ann Arbor Stage	III
Ann Arbor Extranodal Status	NE
B-Symptomatik	yes
Anzahl der Extranodalbefälle	1
Knochenmarksbefall	no
Bulk	no
LDH	430
LDH Einheit	U/l
LDH oberer Normwert	240

## Zusammenfassung und Ausblick

---

- GeWare als **Auswerteplattform** für verschiedene hoch- und niedrigdimensionale experimentelle Daten
- Matrix-CGH als **neuer Gen-Datentyp** in GeWare
- Entwicklung anhand verbundspezifischer Anforderungen und Workflows, jedoch weitestgehend allgemeine Konzeption der Module, so dass eine Nutzung für andere Projekte und Verbände möglich ist
- **ToDo:**
  - Integration verschiedener Preprocessing-Methoden zur Diskretisierung der Daten
  - Integration häufig benutzter Analysen und Visualisierungen