FMI

Friedrich Miescher Institute
for Biomedical Research

Tim Roloff Handschin

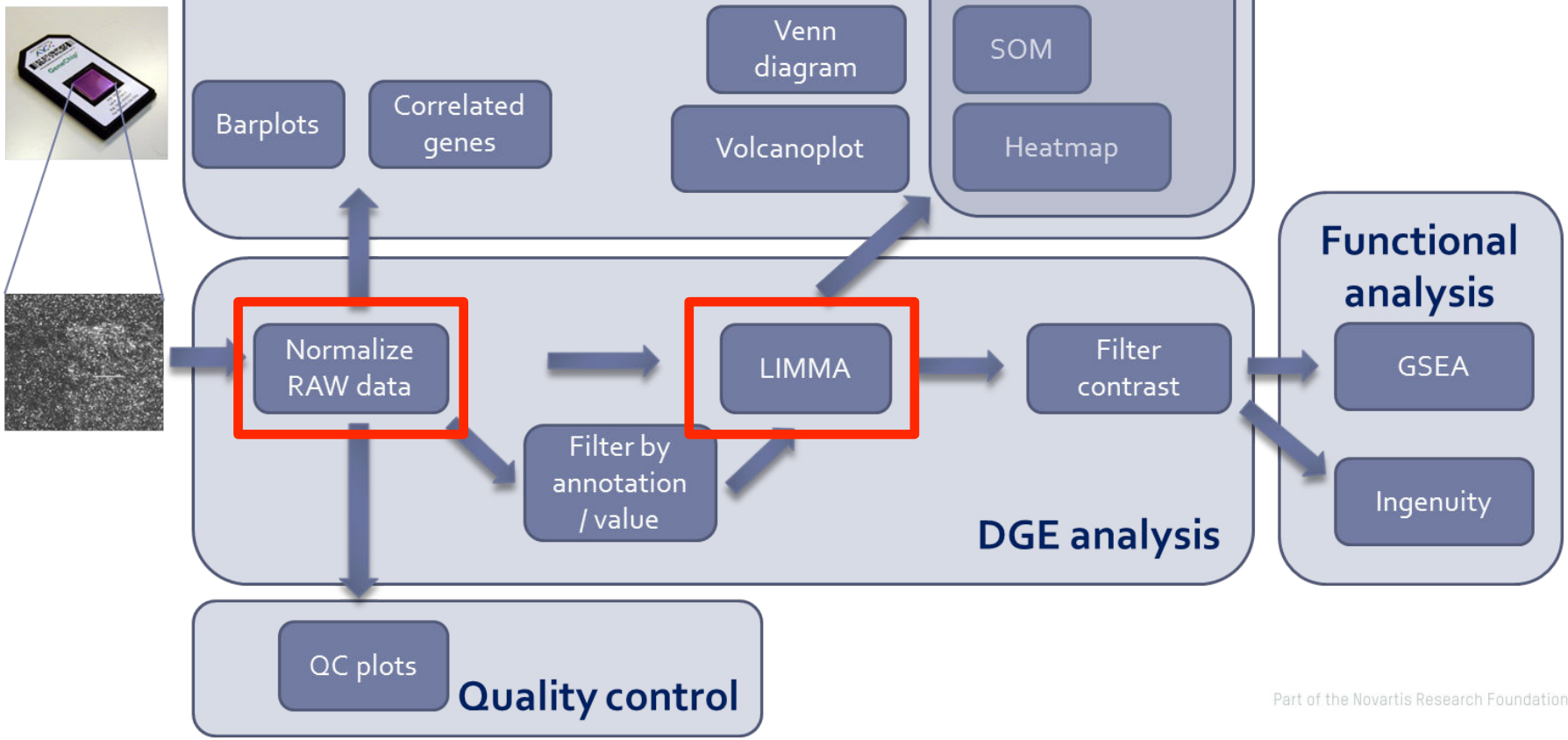
Galaxy workshop Bern
Oct 3, 2012

FMI
Friedrich Miescher Institute
for Biomedical Research

Reduce standard analyses for bioinformaticians

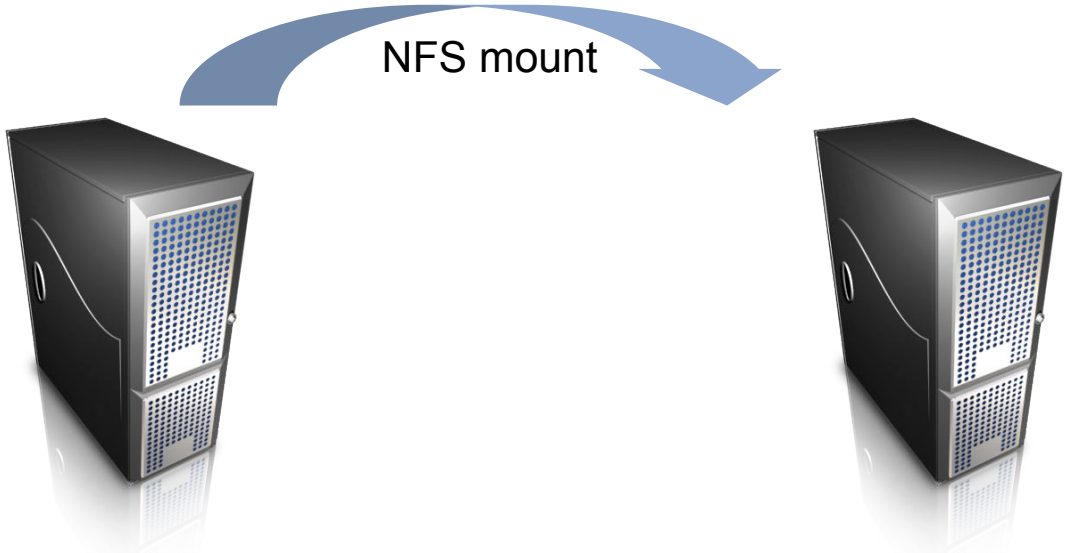
Reduce standard analyses for bioinformaticians

- Wrap R/Bioconductor scripts for Affymetrix microarray analyses in Galaxy



FMI
Friedrich Miescher Institute
for Biomedical Research

Raw data stored outside galaxy



Data repository server

~ 20000 .CEL files

Dedicated Galaxy server

Friedrich Miescher Institute
for Biomedical Research

File paths passed as input parameter to tool

xml file created by cron job for each group in galaxy_dist/tool-data

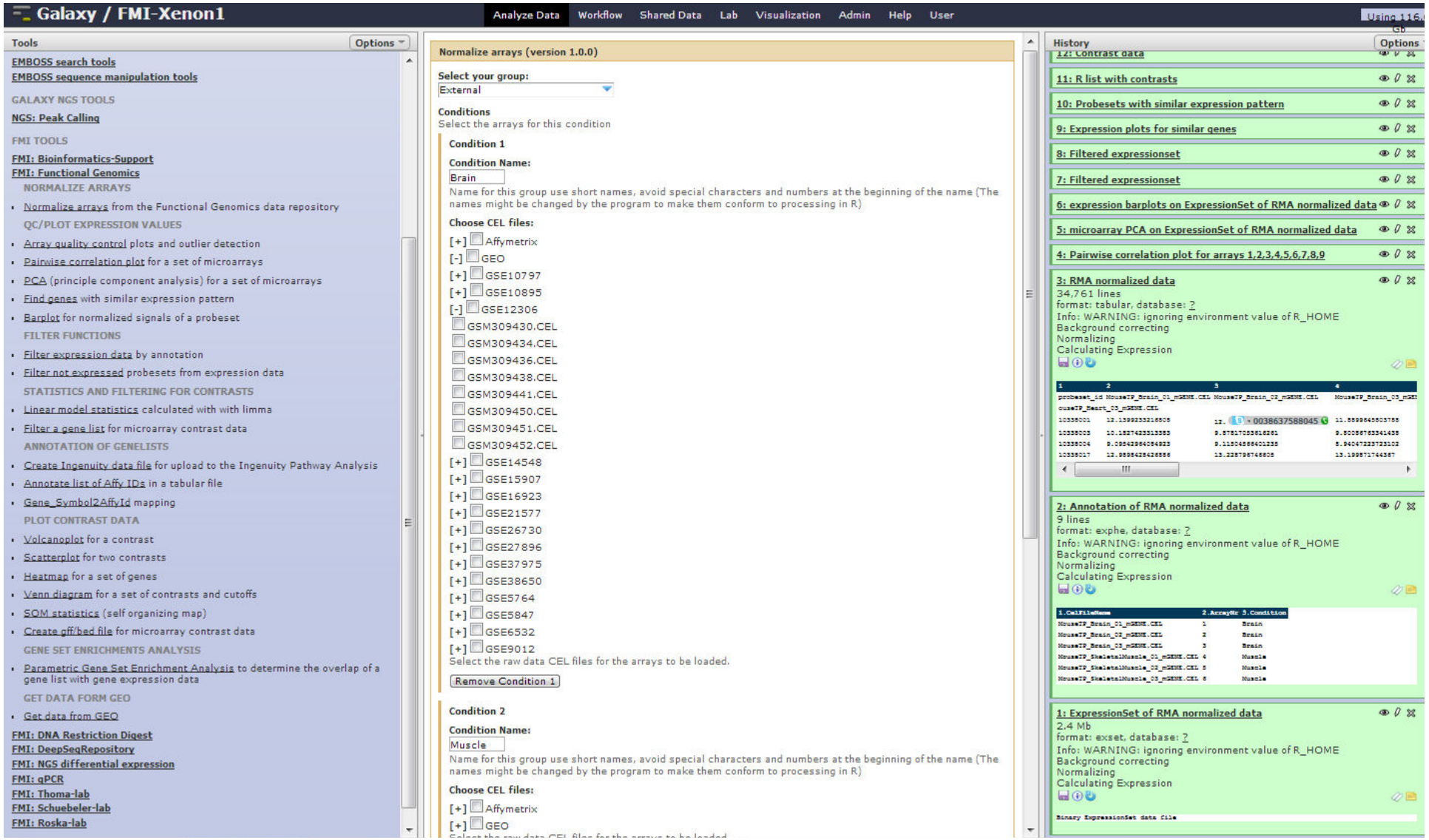
```
<option name="GSE10895" value="GSE10895">
   <option type="m eta_key" name="GSM276175.CEL" value="/work/affy/external/GEO/GSE10895/GSM276175.CEL.gz"/>
   <option type="meta_key" name="GSM276176.CEL" value="/work/affy/external/GEO/GSE10895/GSM276176.CEL.gz"/>
   <option type="meta_key" name="GSM276177.CEL" value="/work/affy/external/GEO/GSE10895/GSM276177.CEL.gz"/>
   <option type="meta_key" name="GSM276178.CEL" value="/work/affy/external/GEO/GSE10895/GSM276178.CEL.gz"/>
</option>
```

Tool xml file

```
<conditional>
  </when>
    <when value="external">
      <repeat name="conditions" title="Condition" help="Select the arrays for this condition">
       <param name="condName" type="text" label="Condition Name" help="Name for this group">
          <validator type="length" min="1" max="12" message="Please enter a short name for each condition!" />
       </param>
       <param name="cels" type="drill_down" display="checkbox" hierarchy="recurse" multiple="true" label="CEL files“ from_file="fungen/external.xml">
          <validator type="no_options" message="Select at least 2 .CEL files for each condition!" />
        </param>
    </repeat>
  </when>
</conditional>
```

From CEL files to ExpressionSet

Friedrich Miescher Institute for Biomedical Research



Custom datatypes for R objects

FMI
Friedrich Miescher Institute
for Biomedical Research

Normalization

History item	Comment
ExpressionSet of normalized data	Binary dataset, used as input for following tools
Annotation of normalized data	Tabular file listing the normalized CEL files and conditions
Normalized data	Large tabular file with normalized expression values for each probeset on each array

Linear modeling (LIMMA)

History item	Comment
Contrast data (R list)	Binary dataset with contrast data; to be used as input for subsequent tools
Contrast data	Tabular dataset with normalized expression values for all arrays, logFCs, p-values, adj p-values for each contrast and annotation
Contrast Information	Tabular dataset with contrast names

Additional concepts used

- Access rights controlled via $userEmail and lookup table
- In R:
  - R objects stored as binary files using save(): faster reading and writing
    - Data
    - Annotation
  - suppressMessages() and suppressWarnings() to avoid output to standard error
- External data (GEO, ArrayExpress, Affymetrix) downloaded to data repository

FMI
Friedrich Miescher Institute
for Biomedical Research

Status quo:

- 21 tools implemented
- > 500 analyses run

We are currently working on:

- QC for raw data
- Plot probesets for genomic regions
- Meta analyses
- Time-course analyses
- Data preparation for external tools / webservices
  - David
  - Broad GSEA
  - Mara
- Data preparation for GEO submission

FMI
Friedrich Miescher Institute
for Biomedical Research

Thanks to

- **Hans-Rudolf Hotz**
- **FMI IT support**
- **FMI Galaxy users (155)**