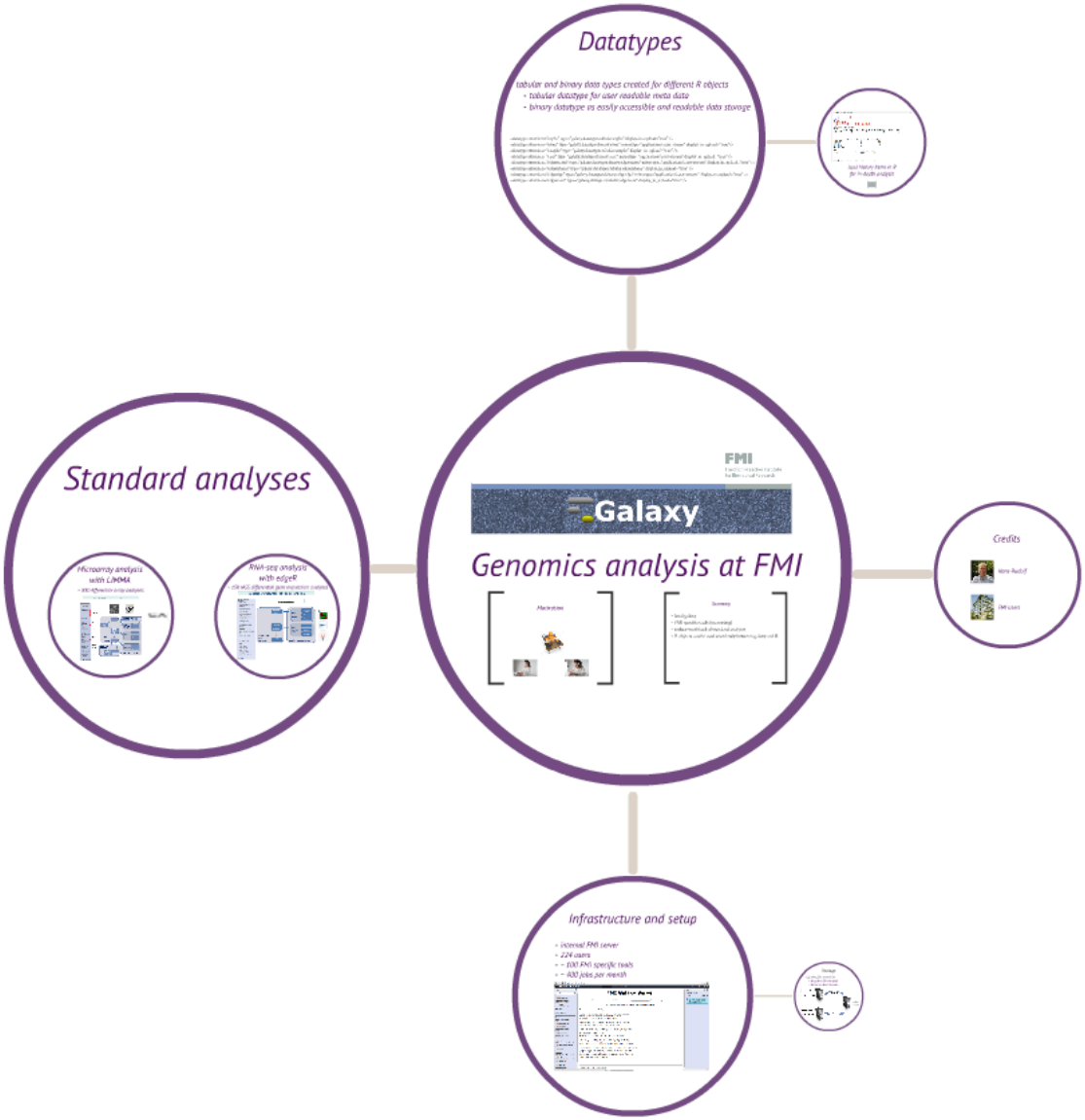
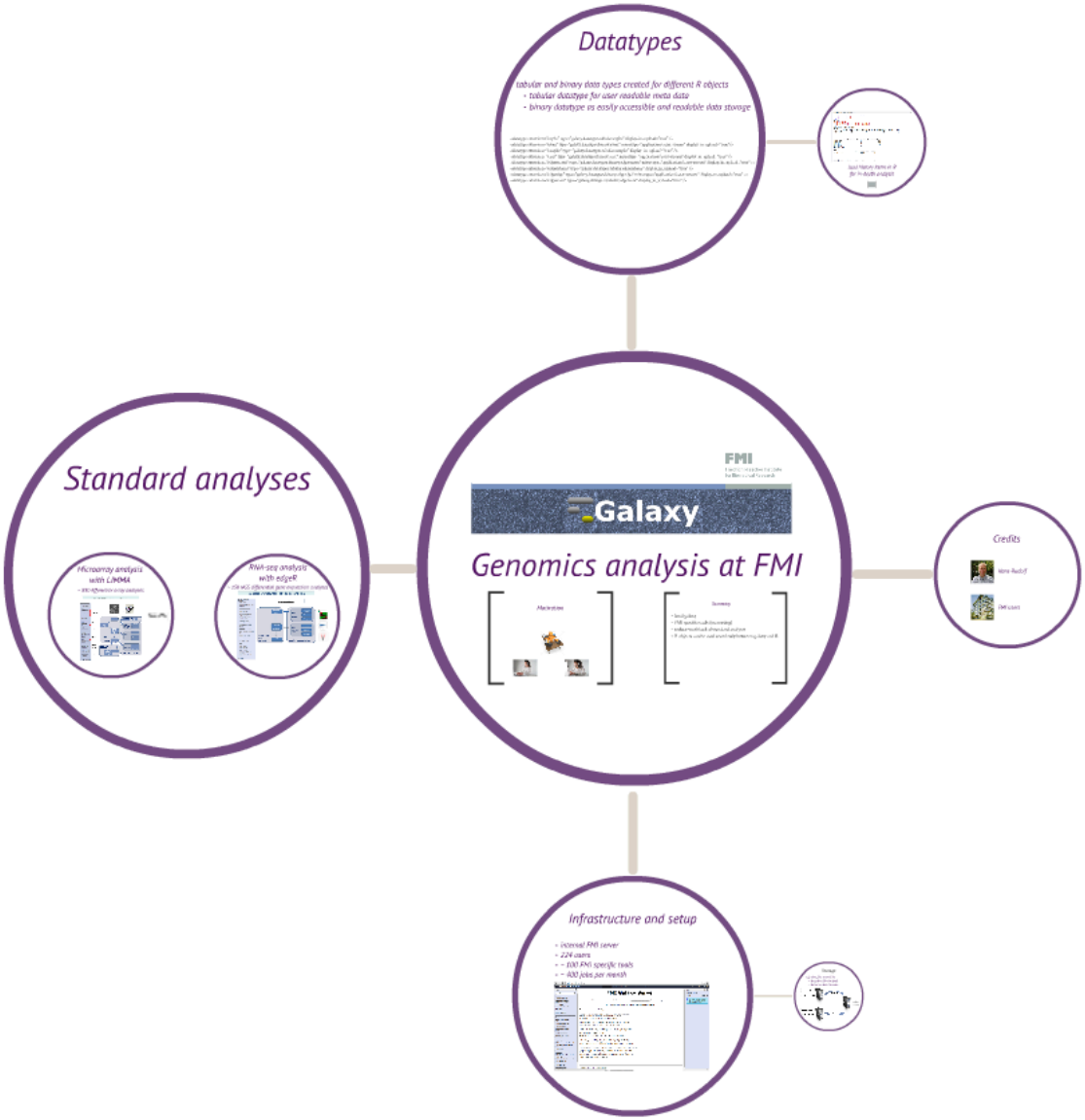

Datatypes
tabular and binary data types created for different R objects
• tabular datatype for user readable meta data
• binary datatype as easily accessible and readable data storage
Standard analyses
Microarray analysis with LIMMA
RNA-seq analysis with edgeR
FMI
Galaxy
Genomics analysis at FMI
Credits
Infrastructure and setup
• internal FMI server
• 224 users
• ~ 100 FMI specific tools
• ~ 400 jobs per month


Datatypes
tabular and binary data types created for different R objects
• tabular datatype for user readable meta data
• binary datatype as easily accessible and readable data storage
Standard analyses
Microarray analysis with LIMMA
RNA-seq analysis with edgeR
FMI
Galaxy
Genomics analysis at FMI
Credits
Infrastructure and setup
• internal FMI server
• 224 users
• ~ 100 FMI specific tools
• ~ 400 jobs per month

FMI
Friedrich Miescher Institute
for Biomedical Research

Galaxy

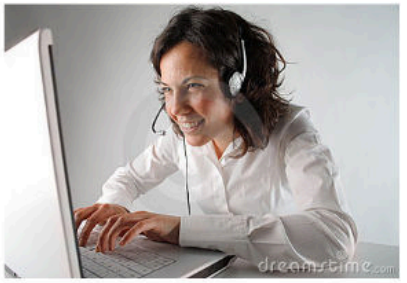
Genomics analysis at FMI

Motivation

Summary

- local galaxy
- FMI specific tools (increasing)
- reduce workload of standard analyses
- R objects can be used seamlessly between galaxy and R

Motivation



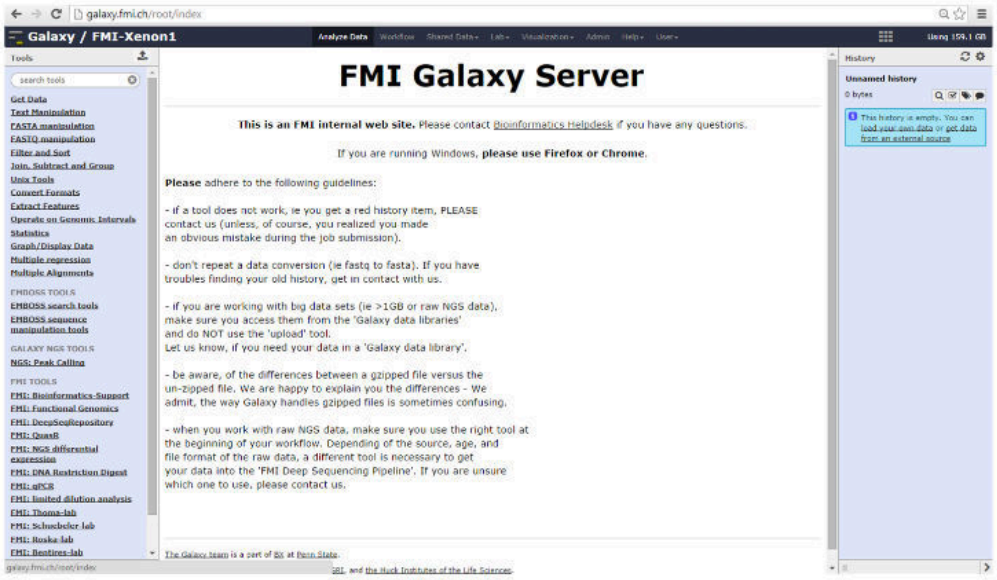

Graphic Leftovers


dreamstime.com

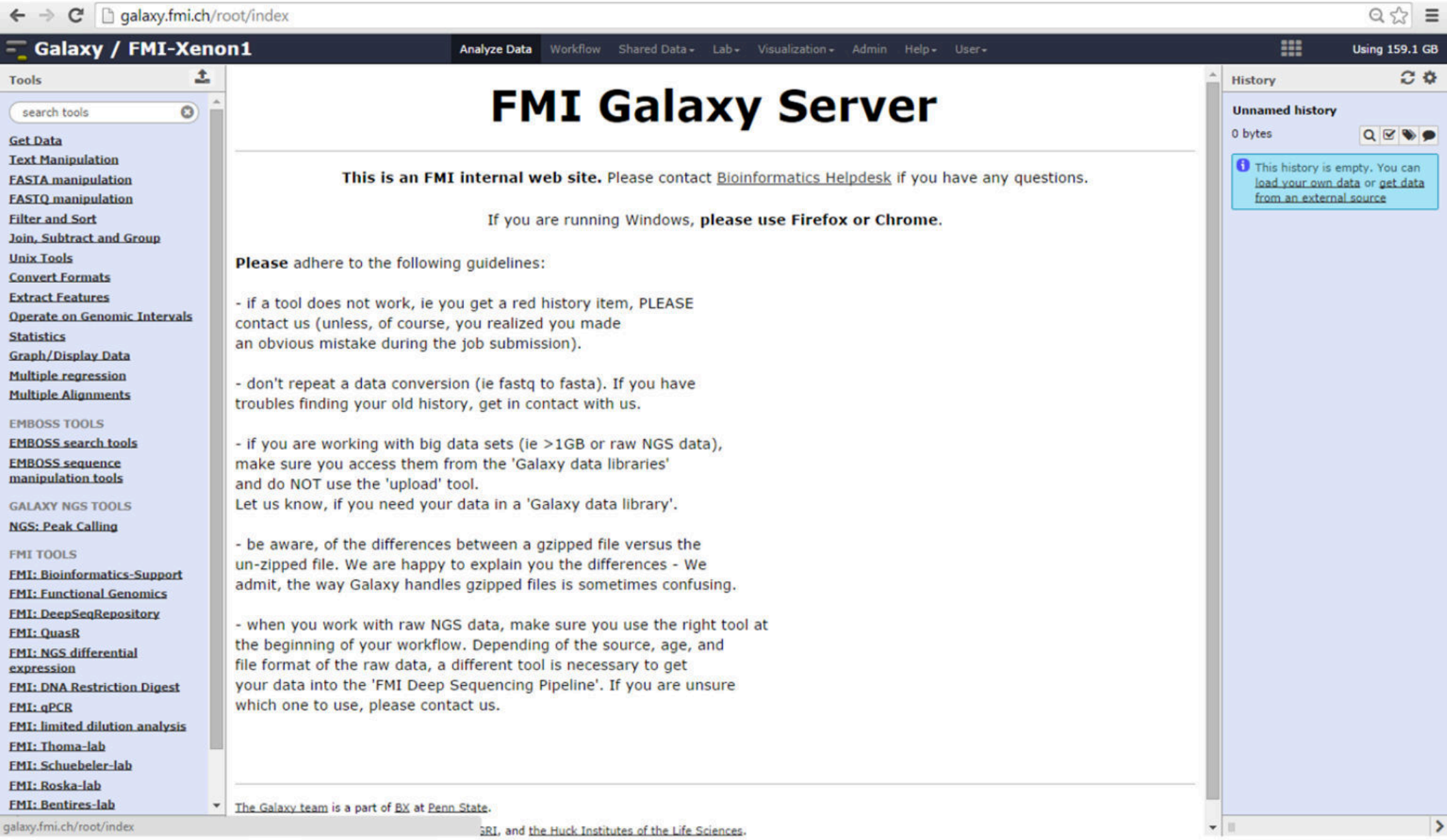

dreamstime.com

Infrastructure and setup

- *internal FMI server*
- *224 users*
- *~ 100 FMI specific tools*
- *~ 400 jobs per month*



- ~ 400 jobs per month



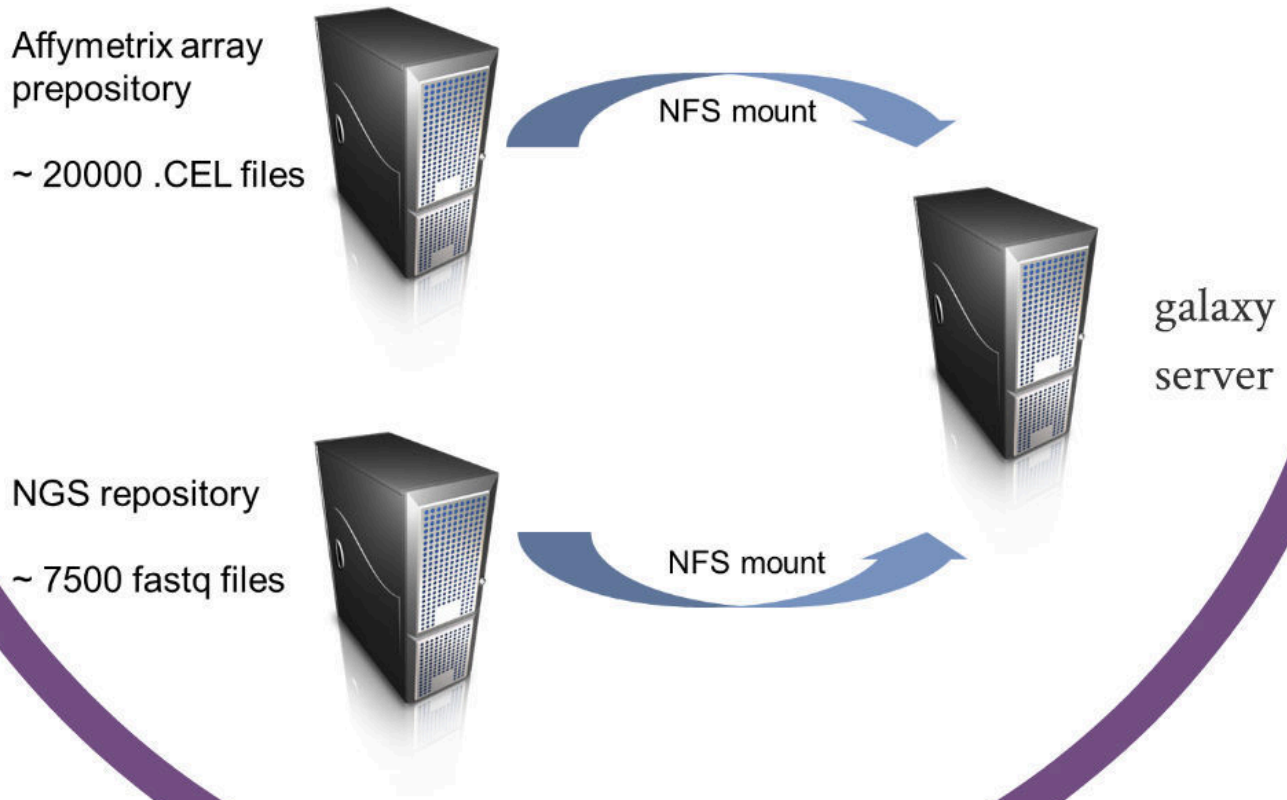
Infrastructure and setup

- *internal FMI server*
- *224 users*
- *~ 100 FMI specific tools*
- *~ 400 jobs per month*

Storage

big data files accessible

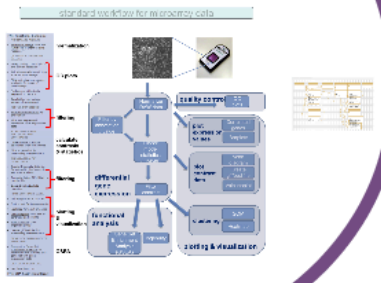
- *directly within in tools*
- *linked as data libraries*



Standard analyses

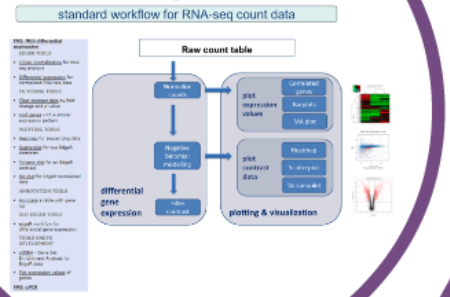
Microarray analysis with LIMMA

~ 850 Affymetrix array analyses



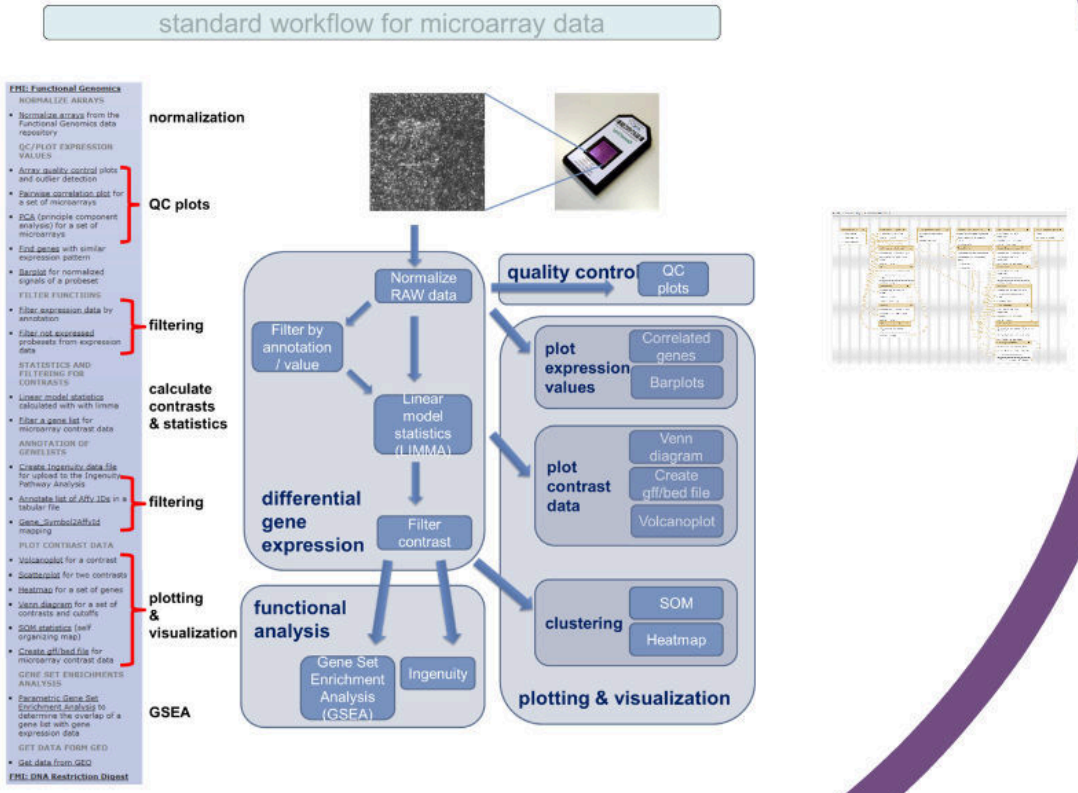
RNA-seq analysis with edgeR

~ 150 NGS differential gene expression analyses



Microarray analysis with LIMMA

~ 850 Affymetrix array analyses



standard workflow for microarray data

FMI: Functional Genomics

NORMALIZE ARRAYS

- Normalize arrays from the Functional Genomics data repository

QC/PLOT EXPRESSION VALUES

- Array quality control plots and outlier detection
- Pairwise correlation plot for a set of microarrays
- PCA (principle component analysis) for a set of microarrays
- Find genes with similar expression pattern
- Barplot for normalized signals of a probeset

FILTER FUNCTIONS

- Filter expression data by annotation
- Filter not expressed probesets from expression data

STATISTICS AND FILTERING FOR CONTRASTS

- Linear model statistics calculated with with limma
- Filter a gene list for microarray contrast data

ANNOTATION OF GENELISTS

- Create Ingenuity data file for upload to the Ingenuity Pathway Analysis
- Annotate list of Affy IDs in a tabular file
- Gene_Symbol2AffyId mapping

PLOT CONTRAST DATA

- Volcanoplot for a contrast
- Scatterplot for two contrasts
- Heatmap for a set of genes
- Venn diagram for a set of contrasts and cutoffs
- SOM statistics (self organizing map)
- Create gff/bed file for microarray contrast data

GENE SET ENRICHMENTS ANALYSIS

- Parametric Gene Set Enrichment Analysis to determine the overlap of a gene list with gene expression data

GET DATA FORM GEO

- Get data from GEO

FMI: DNA Restriction Digest

normalization

QC plots

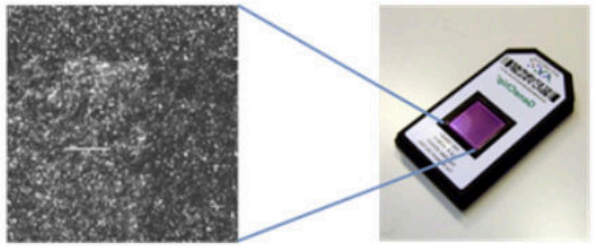
filtering

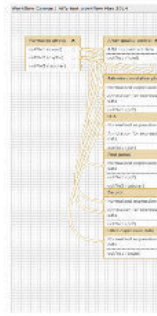
calculate contrasts & statistics

filtering

plotting & visualization

GSEA





Normalize RAW data

Filter by annotation / value

Linear model statistics (LIMMA)

Filter contrast

differential gene expression

quality control

QC plots

plot expression values

Correlated genes

Barplots

plot contrast data

Venn diagram

Create gff/bed file

Volcanoplot

clustering

SOM

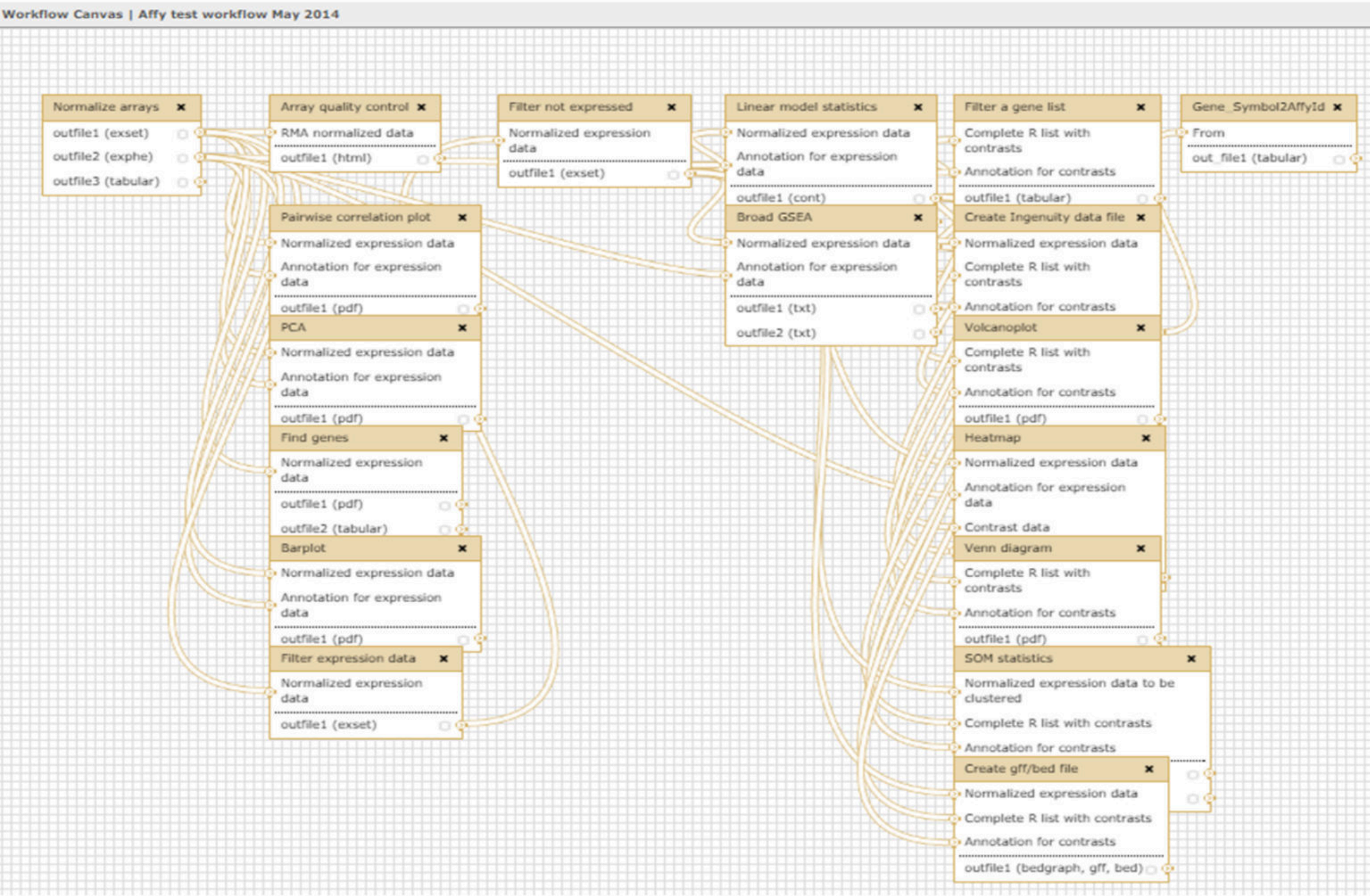
Heatmap

plotting & visualization

functional analysis

Gene Set Enrichment Analysis (GSEA)

Ingenuity


Workflow Canvas | Affy test workflow May 2014
Normalize arrays
outfile1 (exset)
outfile2 (exphe)
outfile3 (tabular)
Array quality control
RMA normalized data
outfile1 (html)
Filter not expressed
Normalized expression data
outfile1 (exset)
Linear model statistics
Normalized expression data
Annotation for expression data
outfile1 (cont)
Filter a gene list
Complete R list with contrasts
Annotation for contrasts
outfile1 (tabular)
Gene_Symbol2AffyId
From
out_file1 (tabular)
Pairwise correlation plot
Normalized expression data
Annotation for expression data
outfile1 (pdf)
Broad GSEA
Normalized expression data
Annotation for expression data
outfile1 (txt)
outfile2 (txt)
Create Ingenuity data file
Normalized expression data
Complete R list with contrasts
Annotation for contrasts
PCA
Normalized expression data
Annotation for expression data
outfile1 (pdf)
Volcanoplot
Complete R list with contrasts
Annotation for contrasts
outfile1 (pdf)
Find genes
Normalized expression data
outfile1 (pdf)
outfile2 (tabular)
Heatmap
Normalized expression data
Annotation for expression data
Contrast data
Barplot
Normalized expression data
Annotation for expression data
outfile1 (pdf)
Venn diagram
Complete R list with contrasts
Annotation for contrasts
outfile1 (pdf)
Filter expression data
Normalized expression data
outfile1 (exset)
SOM statistics
Normalized expression data to be clustered
Complete R list with contrasts
Annotation for contrasts
Create gff/bed file
Normalized expression data
Complete R list with contrasts
Annotation for contrasts
outfile1 (bedgraph, gff, bed)

RNA-seq analysis with edgeR

~ 150 NGS differential gene expression analyses

standard workflow for RNA-seq count data

FMI: NGS differential expression

EDGER TOOLS

- EdgeR normalization for RNA-seq analysis
- Differential expression for normalized RNA-seq data

FILTERING TOOLS

- Filter contrast data by fold-change and p-value
- Find genes with a similar expression pattern

PLOTTING TOOLS

- Heatmap for sequencing data
- Scatterplot for two EdgeR contrasts
- Volcano plot for an EdgeR contrast
- MA plot for EdgeR normalized data

ANNOTATION TOOLS

- Annotate a table with gene ids

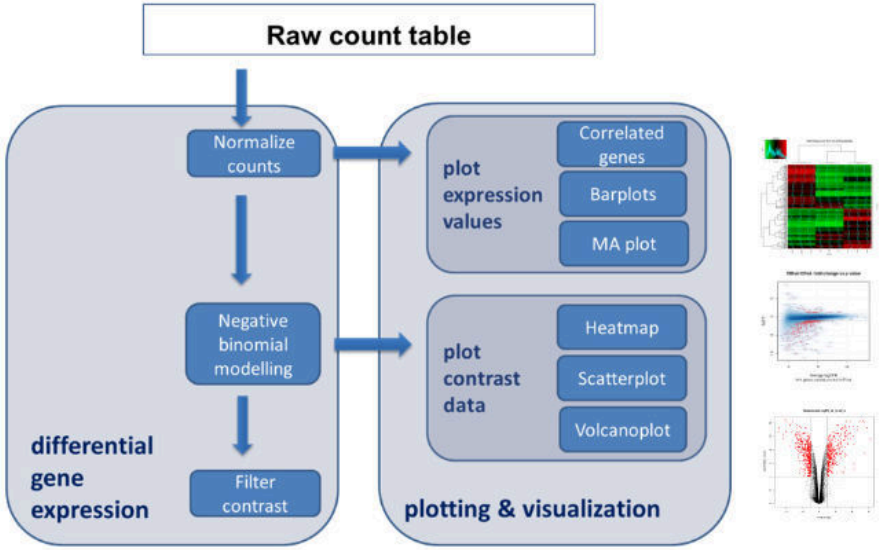
OLD EDGER TOOLS

- edgeR workflow for differential gene expression

TOOLS UNDER DEVELOPMENT

- pGSEA - Gene Set Enrichtment Analysis for EdgeR data
- Plot expression values of genes

FMI: qPCR



standard workflow for RNA-seq count data

FMI: NGS differential expression

EDGER TOOLS

- EdgeR normalization for RNA-seq analysis
- Differential expression for normalized RNA-seq data

FILTERING TOOLS

- Filter contrast data by fold-change and p-value
- Find genes with a similar expression pattern

PLOTTING TOOLS

- Heatmap for sequencing data
- Scatterplot for two EdgeR contrasts
- Volcano plot for an EdgeR contrast
- MA plot for EdgeR normalized data

ANNOTATION TOOLS

- Annotate a table with gene ids

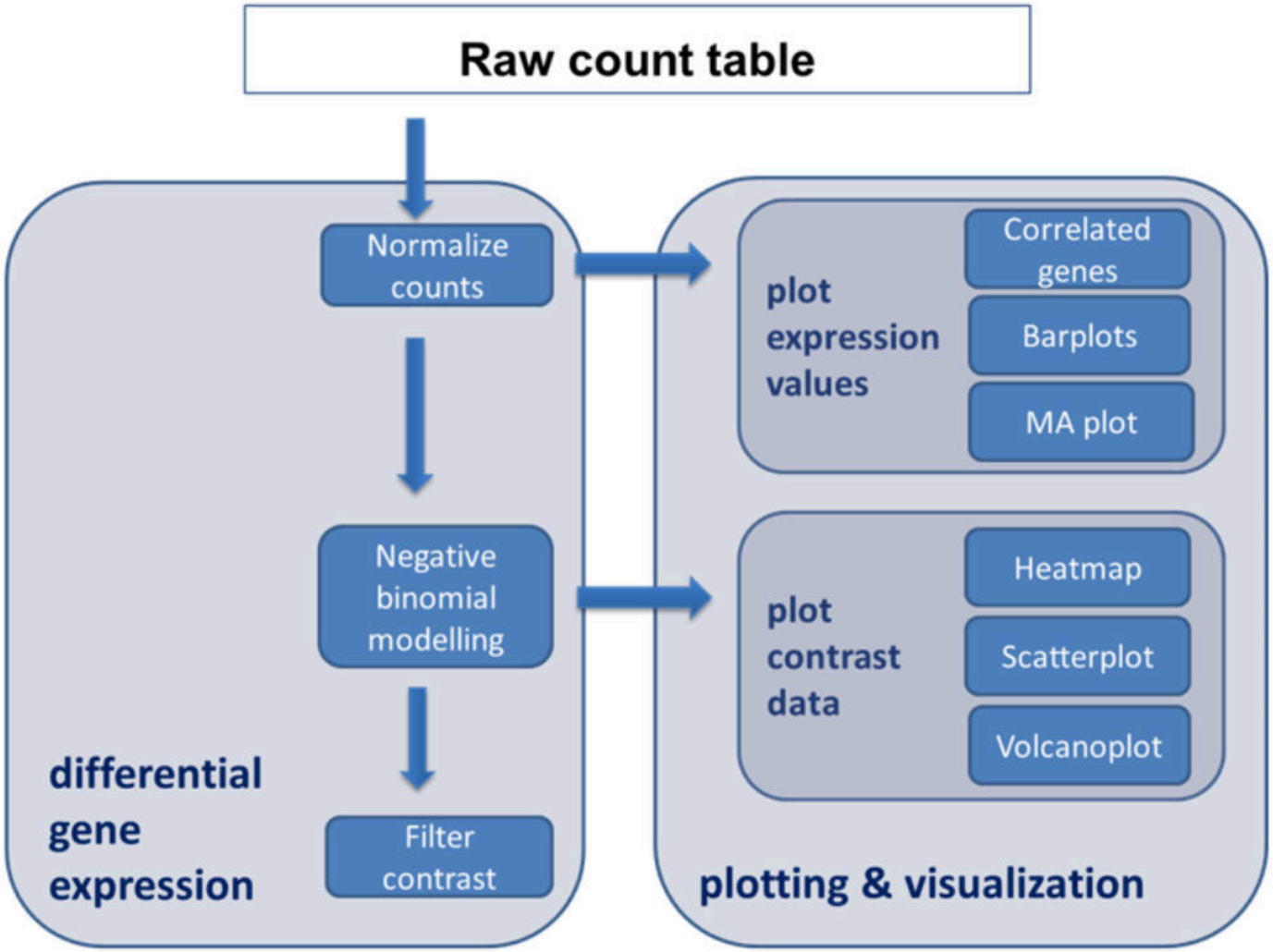
OLD EDGER TOOLS

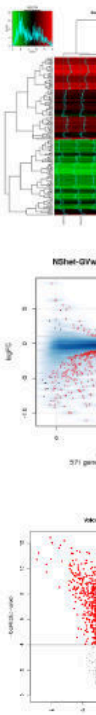
- edgeR workflow for differential gene expression

TOOLS UNDER DEVELOPMENT

- pGSEA - Gene Set Enrichtment Analysis for EdgeR data
- Plot expression values of genes

FMI: qPCR





genes

Barplots

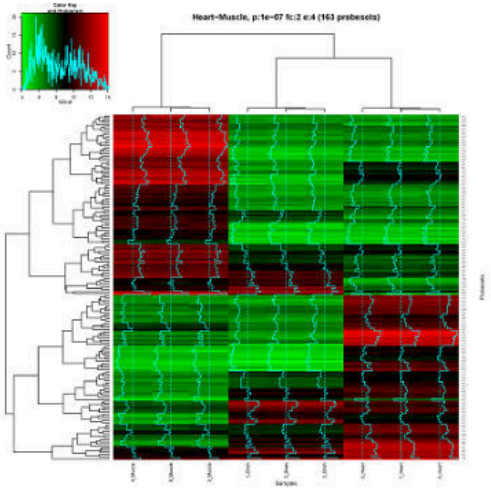
MA plot

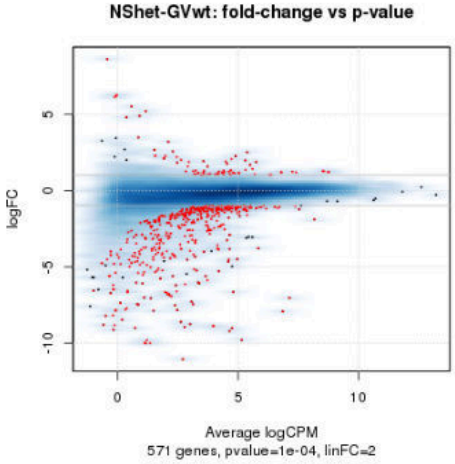
Heatmap

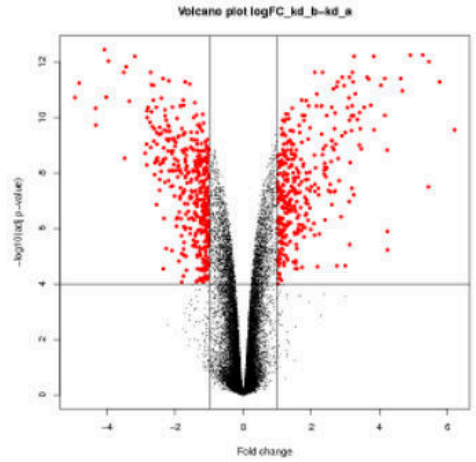
Scatterplot

Volcanoplot

alization







Datatypes

tabular and binary data types created for different R objects

- *tabular datatype for user readable meta data*
- *binary datatype as easily accessible and readable data storage*

```
<datatype extension="exphe" type="galaxy.datatypes.tabular:exphe" display_in_upload="true" />
<datatype extension="exset" type="galaxy.datatypes.binary:exset" mimetype="application/octet-stream" display_in_upload="true"/>
<datatype extension="conphe" type="galaxy.datatypes.tabular:conphe" display_in_upload="true" />
<datatype extension="cont" type="galaxy.datatypes.binary:cont" mimetype="application/octet-stream" display_in_upload="true" />
<datatype extension="edgernorm" type="galaxy.datatypes.binary:edgernorm" mimetype="application/octet-stream" display_in_upload="true" />
<datatype extension="edgerpheno" type="galaxy.datatypes.tabular:edgerpheno" display_in_upload="true" />
<datatype extension="edgerdge" type="galaxy.datatypes.binary:edgerdge" mimetype="application/octet-stream" display_in_upload="true" />
<datatype extension="edgercont" type="galaxy.datatypes.tabular:edgercont" display_in_upload="true" />
```

tabular and binary data types created for different R objects

- *tabular datatype for user readable meta data*
- *binary datatype as easily accessible and readable data storage*

```
<datatype extension="exphe" type="galaxy.datatypes.tabular:exphe" display_in_upload="true" />
<datatype extension="exset" type="galaxy.datatypes.binary:exset" mimetype="application/octet-stream" display_in_upload="true"/>
<datatype extension="conphe" type="galaxy.datatypes.tabular:conphe" display_in_upload="true" />
<datatype extension="cont" type="galaxy.datatypes.binary:cont" mimetype="application/octet-stream" display_in_upload="true" />
<datatype extension="edgernorm" type="galaxy.datatypes.binary:edgernorm" mimetype="application/octet-stream" display_in_upload="true" />
<datatype extension="edgerpheno" type="galaxy.datatypes.tabular:edgerpheno" display_in_upload="true" />
<datatype extension="edgerdge" type="galaxy.datatypes.binary:edgerdge" mimetype="application/octet-stream" display_in_upload="true" />
<datatype extension="edgercont" type="galaxy.datatypes.tabular:edgercont" display_in_upload="true" />
```

```
Console C:/Users/tiroloff/Downloads/
> load("Galaxy120-[Normalized_DGE_object].edgernorm")
Loading required package: edgeR
Loading required package: limma
> ls()
[1] "dge"
Warning messages:
1: package 'edgeR' was built under R version 3.0.2
2: package 'limma' was built under R version 3.0.2
> dge
An object of class "DGEList"
$counts
             HR1.R2.1_DMSO MBA.PR_HR1.P.2_BYL MBA.PR_HR1.P.3_BYL719 MBA.PR_HR1.P.3_DMSO
NM_001001130           303                451                   449                  437
NM_001001144          1072               1321                  1362                  968
NM_001001152           122                319                   269                  146
NM_001001160            15                 10                    12                   15
NM_001001176             9                 13                     9                   45
             MBA.PR_HR1.R2.1.ON_BYL719 MBA.PR_HR1.R2.1.ON_DMSO
NM_001001130                       430                     396
NM_001001144                      1313                    1300
NM_001001152                       183                     201
NM_001001160                        20                       8
NM_001001176                        32                      25
28842 more rows ...

$samples
                          group lib.size norm.factors
HR1.R2.1_DMSO           control 61531434            1
MBA.PR_HR1.P.2_BYL      control 51500844            1
MBA.PR_HR1.P.3_BYL719   control 54682219            1
MBA.PR_HR1.P.3_DMSO        test 63461504            1
MBA.PR_HR1.R2.1.ON_BYL719  test 50446360            1
MBA.PR_HR1.R2.1.ON_DMSO    test 57703484            1
```

load history items in R
for in-depth analysis


dreamstime.com

Summary

- local galaxy
- FMI specific tools (increasing)
- reduce workload of standard analyses
- R objects can be used seamlessly between galaxy and R

Credits

Hans-Rudolf

FMI users