A graphical user interface to DNA microarray data analysis using R and Bioconductor

Jarno Tuimala
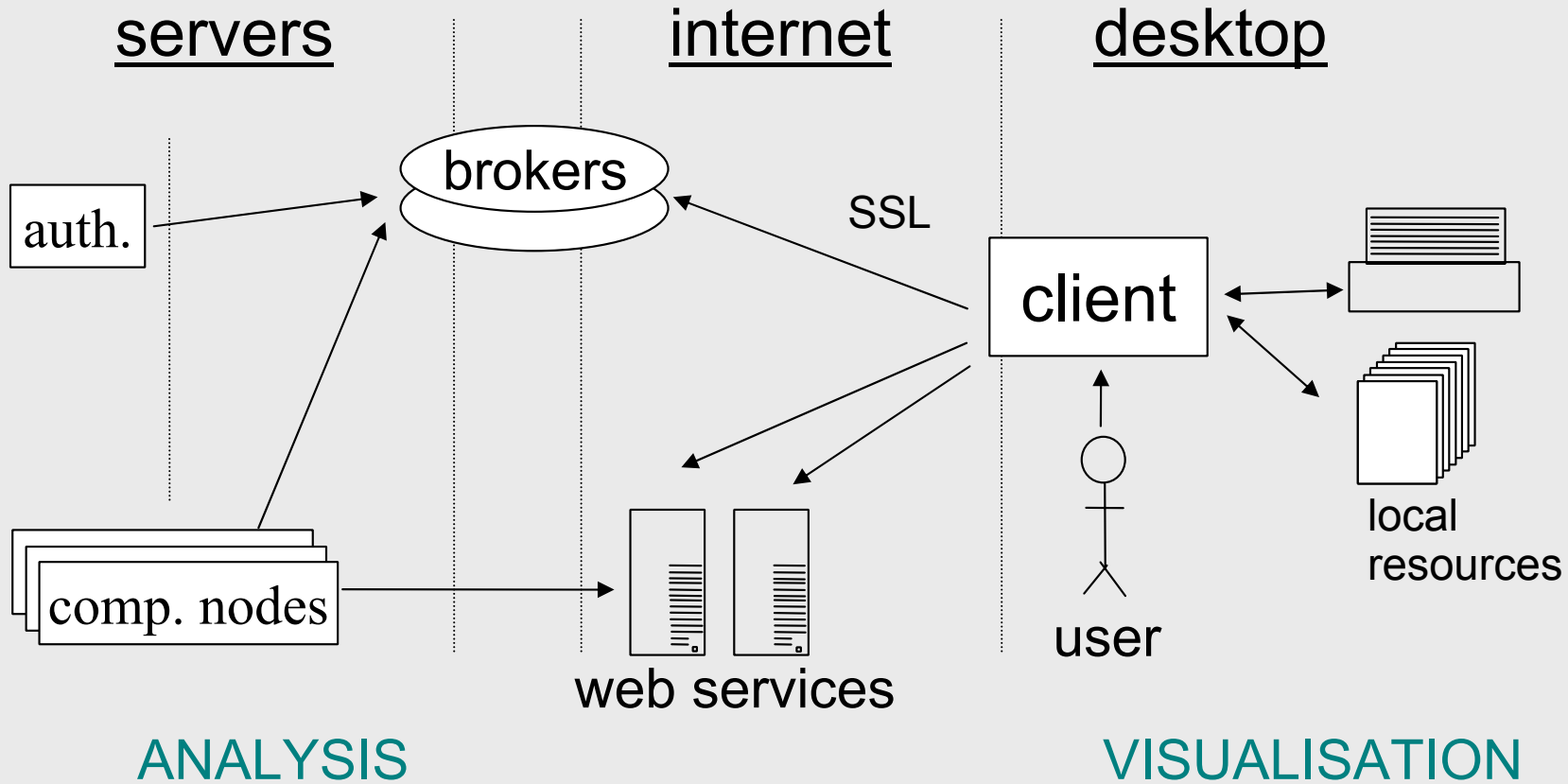
CSC, the Finnish IT Center for Science

jarno.tuimala@csc.fi

Chipster

➢ **Goal: Easy access to up-to-date analysis methods (R/Bioconductor)**

➢ **Features**
- Easy to use graphical interface
- Comprehensive selection of analysis tools and visualizations
- Support for different array types (Affymetrix, Agilent, Illumina, cDNA)
- Possibility to construct (and share) analysis workflows
- Compatible with Windows, Linux and Mac OS X
- Easy to install and update
- Graphical client program on user's computer sends analysis jobs to central computing servers

➢ **Availability**
- Guest accounts for evaluation
- Open source (http://chipster.sourceforge.net)

➢ **http://chipster.csc.fi**

System architecture

servers | internet | desktop

brokers

auth.

SSL

client

comp. nodes

web services

user

local resources

ANALYSIS

VISUALISATION

Analysis tool content

➢ In addition to the usual analysis tools, such as, normalization, filtering, statistics, clustering and annotations, Chipster provides:

- Quality control methods for all supported chiptypes
- Linear (mixed) models for normalization and analysis
- Bootstrapping of hierarchical clustering results
- Alternative mappings of the probes on Affymetrix arrays (altCDF)
- "Gene set enrichment analysis"
- Promoter analysis

➢ Support for chiptypes:

- Affymetrix (3' expression, exon and SNP arrays)
- Agilent (1-color & 2-color systems)
- Illumina (expression arrays)
- Generic cDNA

Acknowledgements

- Aleksi Kallio
- Jarno Tuimala
- Taavi Hupponen
- Petri Klemelä, Mikko Koski, Janne Käki, Mika Rissanen

- All the pilot users
- Department of Computer Science, University of Helsinki
- Tekes, the Finnish Funding Agency for Technology and Innovation

http://chipster.csc.fi/
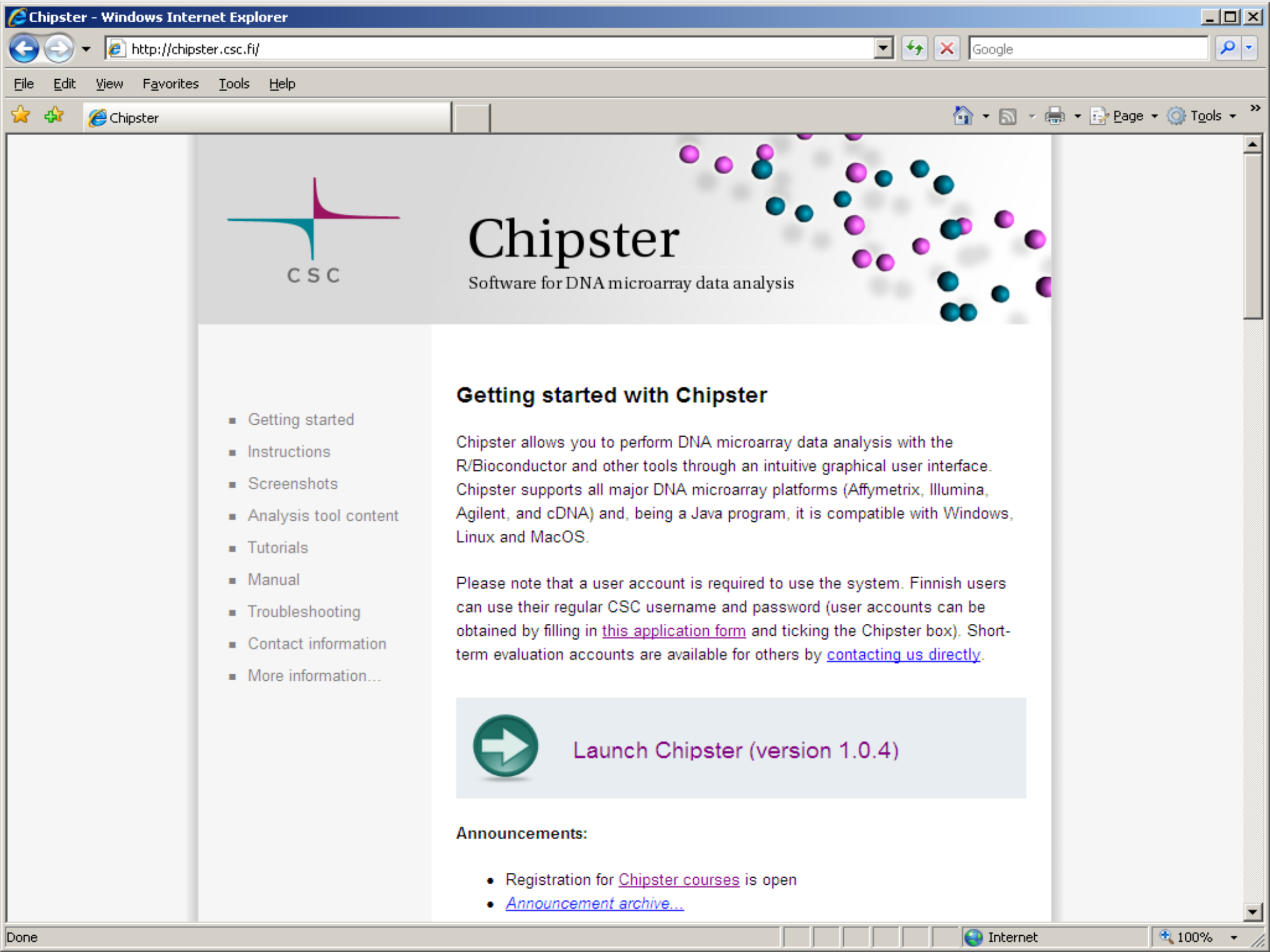
Chipster

Chipster
Software for DNA microarray data analysis

CSC

- Getting started
- Instructions
- Screenshots
- Analysis tool content
- Tutorials
- Manual
- Troubleshooting
- Contact information
- More information...

Getting started with Chipster

Chipster allows you to perform DNA microarray data analysis with the R/Bioconductor and other tools through an intuitive graphical user interface. Chipster supports all major DNA microarray platforms (Affymetrix, Illumina, Agilent, and cDNA) and, being a Java program, it is compatible with Windows, Linux and MacOS.

Please note that a user account is required to use the system. Finnish users can use their regular CSC username and password (user accounts can be obtained by filling in this application form and ticking the Chipster box). Short-term evaluation accounts are available for others by contacting us directly.

→ Launch Chipster (version 1.0.4)

Announcements:

- Registration for Chipster courses is open
- *Announcement archive...*

Done   Internet   100%

File   Edit   View   Wizard   Workflow   Help

Datasets

Datasets
- My experiment
  - GSM11805.cel
  - GSM11814.cel
  - GSM11823.cel
  - GSM11830.cel
  - normalized.tsv
  - phenodata.tsv
  - sd-filter.tsv
  - two-sample.tsv
  - seqs.txt.wee
  - seqs.html
  - annotations.html

Analysis tools

● Normalisation	Single-slide methods
● Quality control	One sample tests
● Preprocessing	Two groups tests
● Statistics	Several groups tests
● Clustering	Correlate with phenodata
● Annotation	Time series
● Pathways	PCA
● Promoter Analysis	NMDS
● Visualisation	Sample size estimation
● Utilities	Linear modelling
	SAM

✓   Edit parameters   Run ▶

Tests for comparing the mean gene expression of two groups. Specify the test type, multiple testing correction method and a p-value cut-off for significance. All the test assume that the difference in group means is equal to be zero. LPE only works, if the whole data is used, i.e., the data should not be pre-filtered, if LPE is used.

More help   Show tool sourcecode

Workflow view

4 files
Norm   Phe   Qual   Qual
Prep
Stat   Path
Prom   Prom   Anno
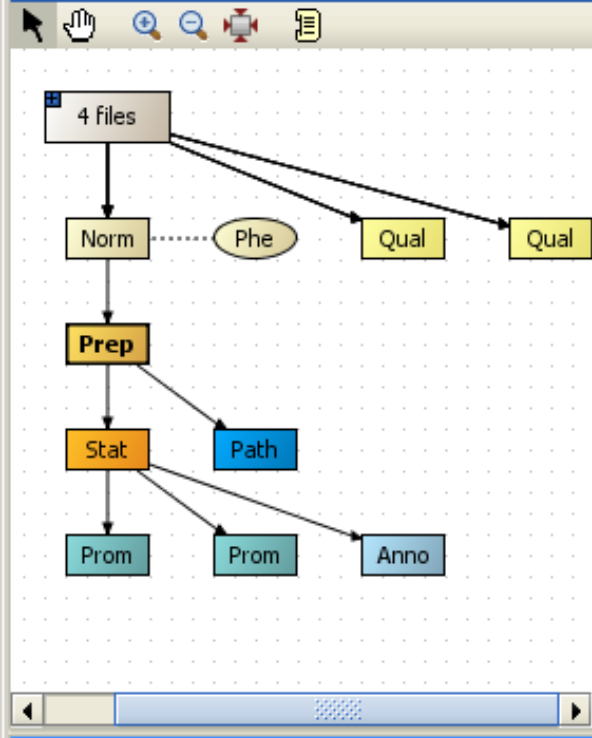
Data visualisation

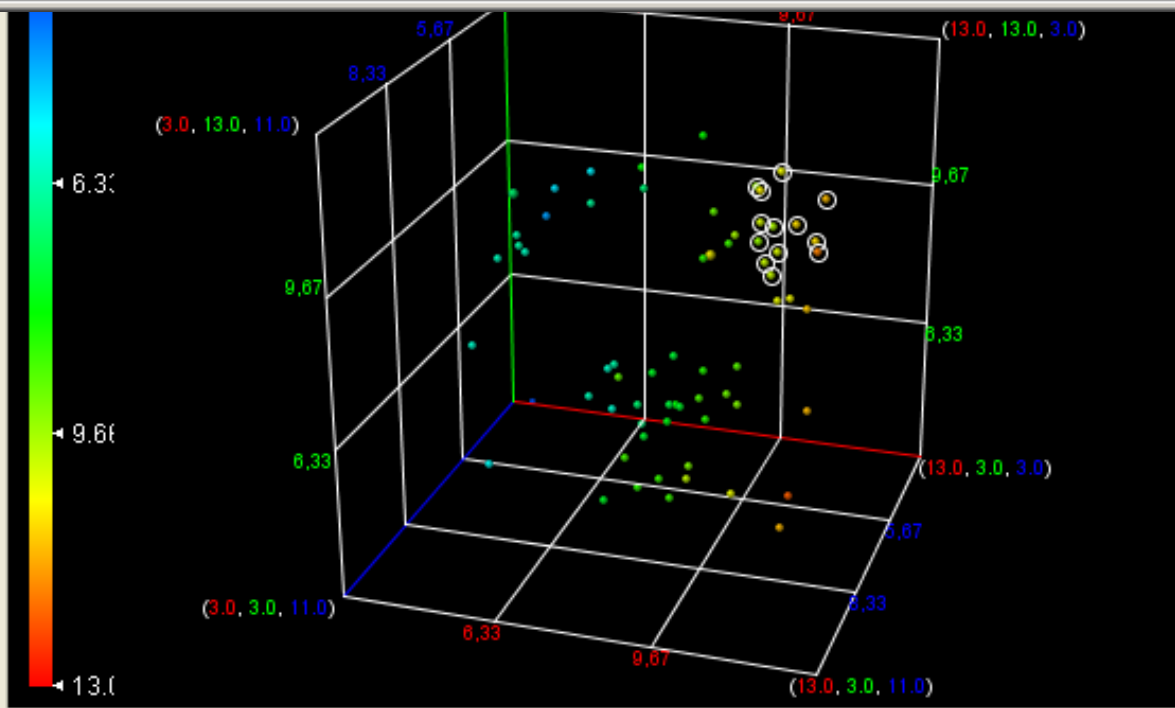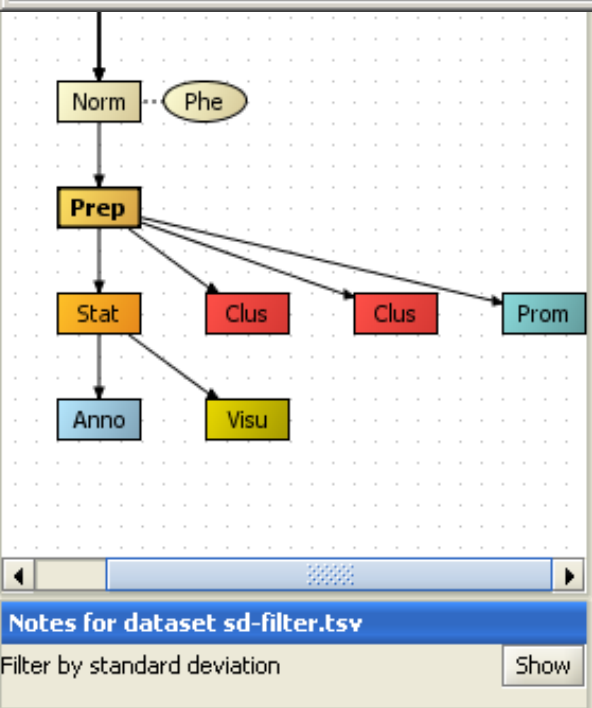Method: Spreadsheet ▾   Redraw   Maximize

Showing 67 rows of 67

	chip.microarray1.cel	chip.microarray2.cel	chip.microarray3.cel	chip.microarray4.cel	flag.microarray1.cel	flag.microarray2.cel	fla
201650_at	8.06	10.63	6.12	3.83	P	P	P
201744_s_at	9.74	8.73	7.25	4.0	P	P	P
201842_s_at	10.18	8.71	8.66	5.11	P	P	P
201909_at	10.38	10.75	5.39	5.58	P	P	A
202018_s_at	10.44	7.57	5.88	4.95	P	P	A
202036_s_at	9.97	7.97	9.11	4.77	P	P	P
202112_at	8.96	8.68	5.57	10.89	P	P	A
202237_at	10.54	11.14	6.4	10.33	P	P	P
202286_s_at	10.79	8.6	9.15	4.24	P	P	P
202310_s_at	10.59	11.18	5.23	7.49	P	P	A
202404_s_at	10.13	9.84	4.54	6.14	P	P	A
203021_at	10.8	9.65	6.47	5.71	P	P	P
203453_at	9.99	8.52	9.83	5.28	P	P	P
204259_at	11.14	9.93	8.53	4.87	P	P	P
204268_at	11.01	7.67	9.59	5.49	P	P	P
204304_s_at	9.49	8.48	5.82	4.25	P	P	P
204607_at	5.59	5.52	9.01	9.49	A	A	P
204704_s_at	6.72	3.56	9.17	4.11	P	A	P

Notes for dataset

Filter by standard deviation   Show

Annotations by GeneCruiser (Broad Institute / MIT)

Probeset ID	UniProt Gene Name	UniProt	RefSeq	Entrez Gene	Entrez Gene Symbol	Entrez Gene Description	UniProt Description	GO Term	GO ID
201744_s_at	TACSTD2 (synonyms: GA733-1, M1S1, TROP2)	P09758	NM_002353	4070	TACSTD2	tumor-associated calcium signal transducer 2	Tumor-associated calcium signal transducer 2 precursor (Pancreatic carcinoma marker protein GA733-1) (Cell surface glycoprotein Trop-2)	cell proliferation, cell surface receptor linked signal transduction, cytosol, integral to plasma membrane, receptor activity, visual perception	0008283, 0007166, 0005829, 0005887, 0004872, 0007601
201842_s_at	SFRP1 (synonyms: FRP, FRP1, SARP2)	Q8N474	NM_003012	6422	SFRP1	secreted frizzled-related protein 1	Secreted frizzled-related protein 1 precursor (sFRP-1) (Frizzled-related protein 1) (FRP-1) (Secreted apoptosis-related protein 2) (SARP-2)	Wnt receptor signaling pathway, anti-apoptosis, extracellular space, membrane, morphogenesis, signal transduction, transmembrane receptor activity	0016055, 0006916, 0005615, 0016020, 0009653, 0007165, 0004888

Norm ···· Phe

Prep

Stat   Clus   Clus   Prom

Anno   Visu

Notes for dataset sd-filter.tsv
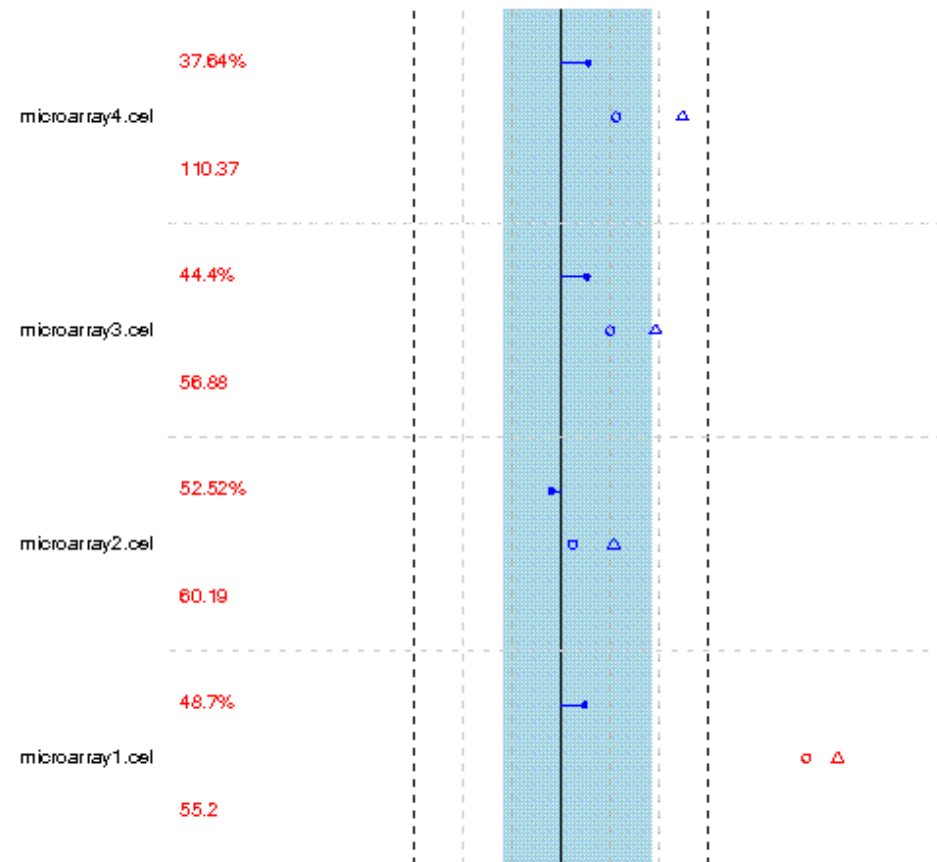
Filter by standard deviation   Show

Settings | Selected

13 Points selected

201744_s_at
201842_s_at
201909_at
202036_s_at
202286_s_at
202310_s_at
202404_s_at
203021_at
203453_at
204259_at
204304_s_at
213629_x_at
AFFX-M27830_5_at

Annotate

RNA digestion plot
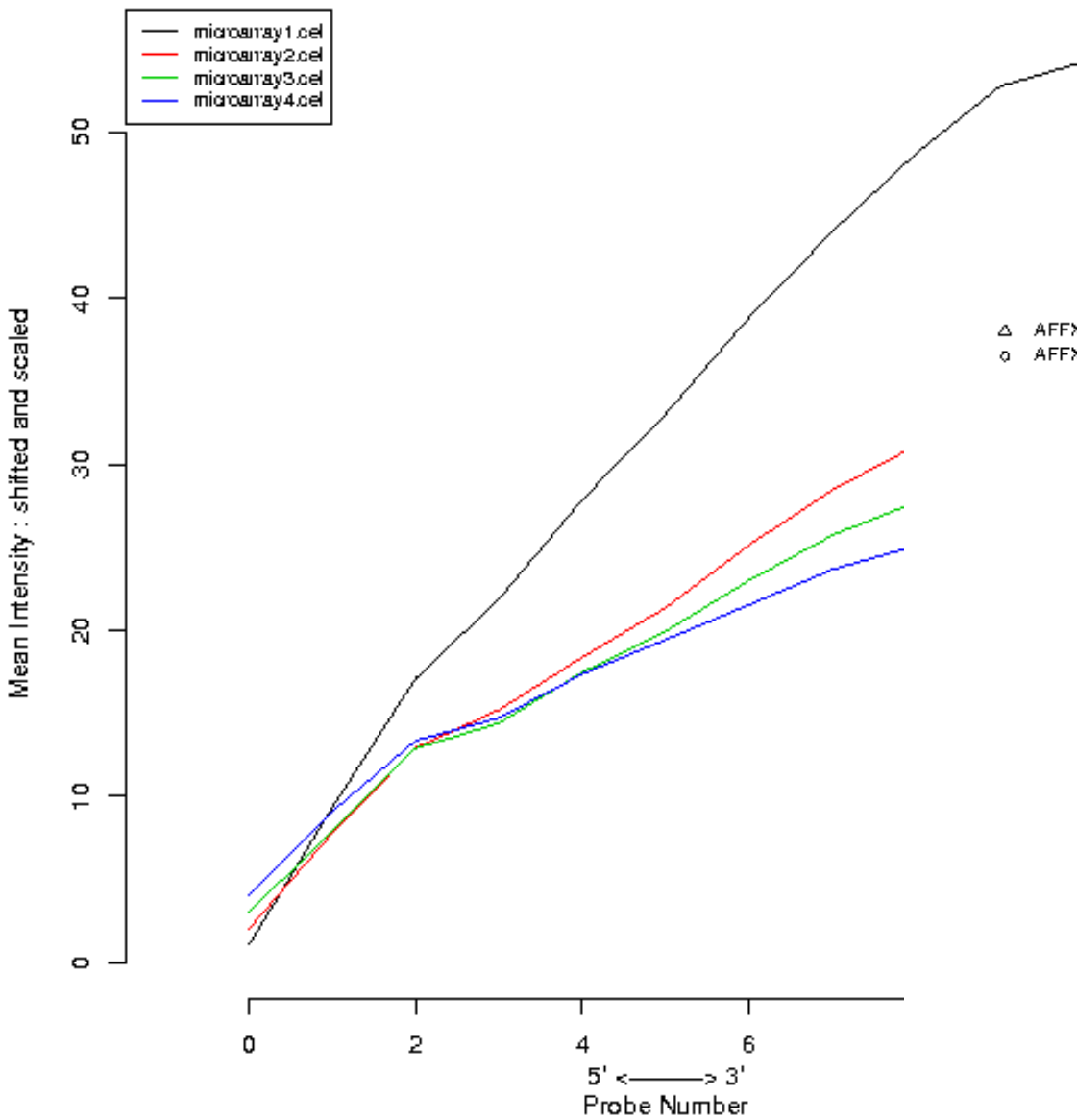
QC Stats

Chipster v1.0.2 (build 349)

File Edit View Wizard Workflow Help

Datasets

- Datasets
  - My experiment
    - GSM11805.cel
    - GSM11814.cel
    - GSM11823.cel
    - GSM11830.cel
    - normalized.tsv
    - phenodata.tsv
    - sd-filter.tsv
    - two-sample.tsv
    - seqs.txt.wee
    - seqs.html
    - annotations.html

History

Show for Datasets:

☑ Step title   ☑ Applied Operation   ☐ User Notes
☑ Dataset Name   ☑ Parameters
☐ Creation Date   ☐ Operation Source Code

```
Step 3
------

Dataset name: two-sample.tsv
Created with operation: Two groups tests
Parameter column: group
Parameter test: t-test
Parameter p.value.adjustment.method: none
Parameter p.value.threshold: 0.05

Step 4


Dataset name: seqs.txt.wee
Created with operation: Weeder
Parameter species: human
Parameter promoter.size: short
```

Save...   Close

If run on a prenormalized file, generates a blank phenodata for it.

More help   Show tool sourcecode

Maximize

Workflow view

4 files

Norm - - - - - Phe   Qual

Prep

Stat   Path

Prom   Prom   Anno
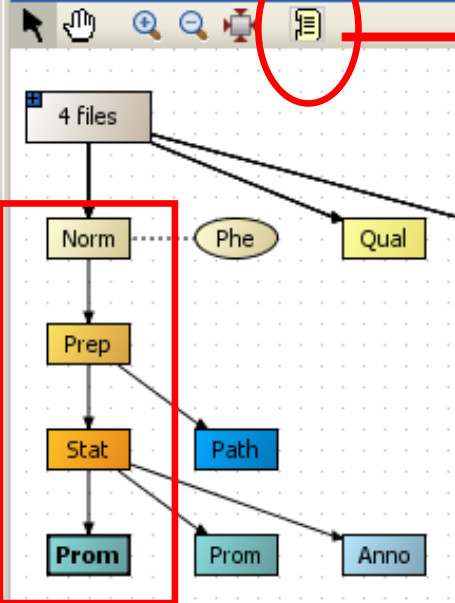
Notes for dataset

Weeder   Show

Connected to broker at nami.csc.fi   Ready   31M / 508M

Workflow – reusing your analysis pipeline

- ➢ **Creates a "macro" that can be applied to another normalized dataset and phenodata**

- ➢ **Choose a dataset, and workflow records the analysis steps starting from it**

- ➢ **Workflows can be shared with other users**

Wizard for Affymetrix data

➢ **Ready-made workflow to find differentially expressed genes**

- Normalization
- Phenodata creation
- Statistical test
- Hierarchical clustering