

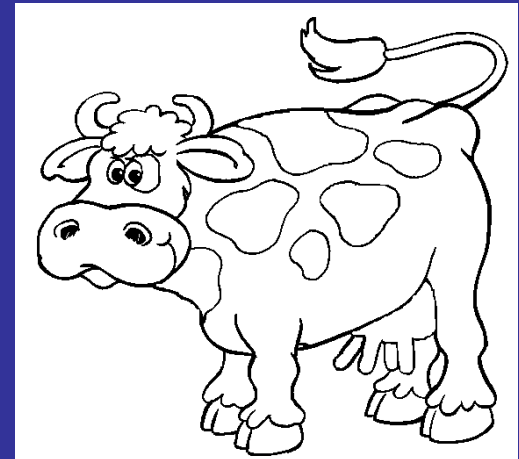
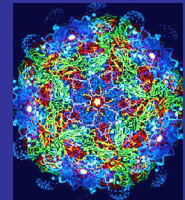
Bioinformatics at the Institute for Animal Health

Michael Watson

Research at the IAH

- Infectious diseases of farm animals
 - **Chicken**
 - Salmonella
 - Eimeria (coccidiosis)
 - Marek's disease
 - **Cow**
 - Foot and Mouth
 - Pestivirus
 - Streptococcal infection (mastitis)
 - **Pig**
 - African Swine Fever
 - Foot and Mouth
 - **Sheep**
 - TSEs (scrapie)
 - Orbivirus

Host-Pathogen Interactions



Ensembl

Ensembl Salmonella typhimurium Genome Server - Microsoft Internet Explorer provided by Institute for Animal Health

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media

Address http://iahc-linux03:8080/Salmonella_typhimurium/ Go Links

Salmonella typhimurium Genome Browser

Salmonella typhimurium Genome

The data on this site is a direct import of the [Salmonella typhimurium genome](#). No additional genebuild was carried out.

We have included *Salmonella typhimurium* into the Ensembl system to allow people to access the data through the Ensembl user interface (both for visualisation and data mining) and to provide cross-species integration through our comparative genomics resources (such as homologous gene links and protein family/domain pages).

Browse a Chromosome

Search for with

From To

[Advanced data](#) [EnsemblMart](#)

Done Local intranet

- Scripts to create an Ensembl database from any EMBL file

- Based on work of Keith James at Sanger Institute

- Have Ensembls for From To

- Salmonella typhi* and *typhimurium*

- Fowlpox

- Campylobacter jejuni*

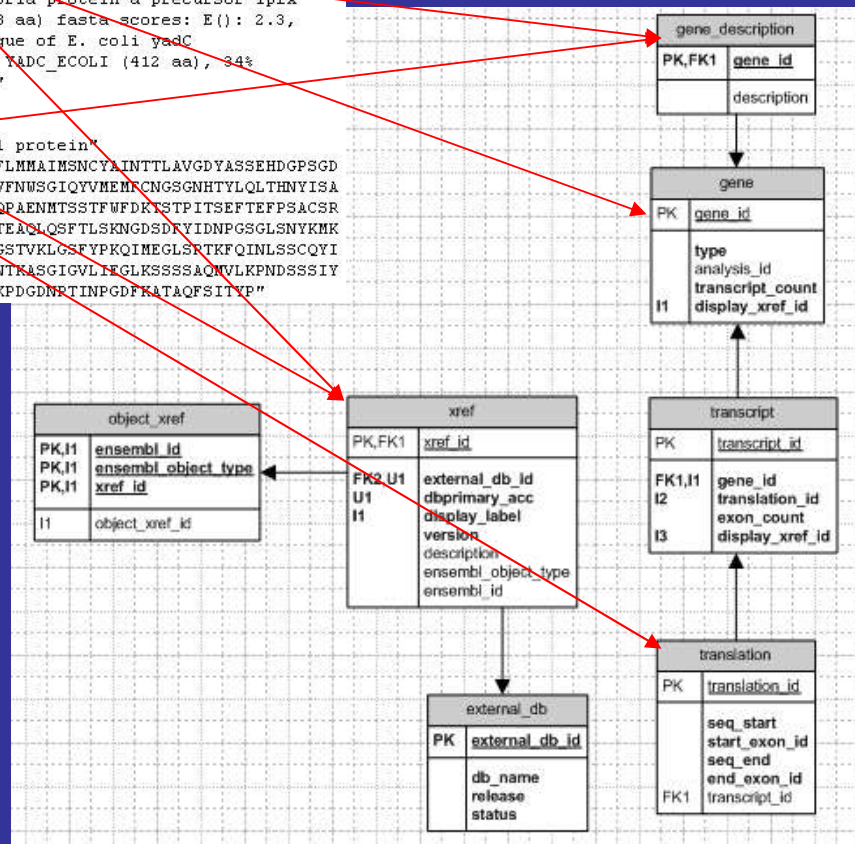
- Hepatitis B (collaboration w/Rostock)

- E.coli*

Basic Functionality

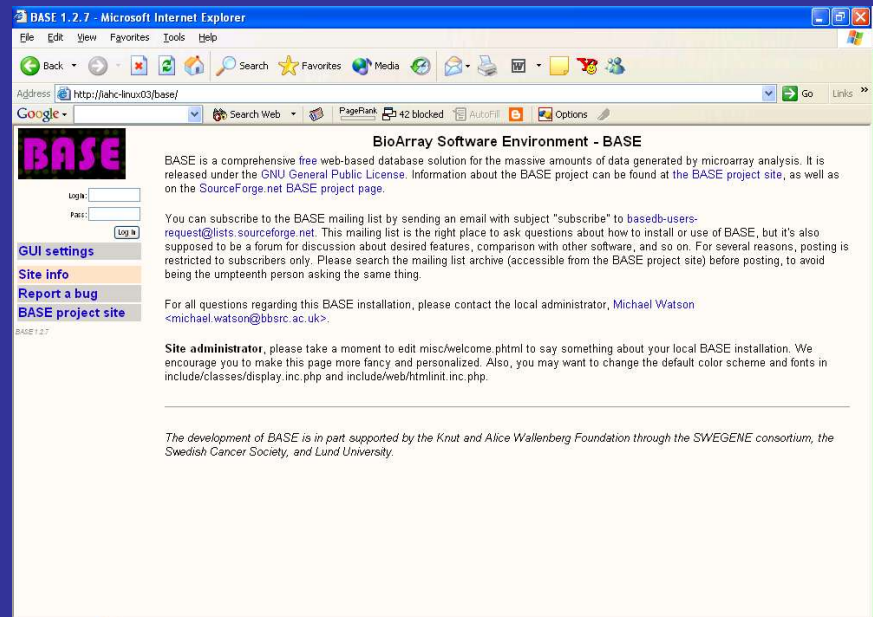
```

FT gene complement(210264..211439)
FT /locus_tag="STY0201"
FT /gene="staG"
FT /note="synonym: yadC"
FT CDS complement(210264..211439)
FT /locus_tag="STY0201"
FT /protein_id="NP_454792.1"
FT /gene="staG"
FT /transl_table=11
FT /note="Similar to Serratia marcescens fimbria a protein
FT precursor smfA SW:FMA_SERMA (P13421) (174 aa) fasta
FT scores: E(): 0.96, 24.6% id in 142 aa, and to Salmonella
FT typhimurium long polar fimbria protein a precursor lpfA
FT SW:LPFA_SALTY (P43660) (178 aa) fasta scores: E(): 2.3,
FT 26.3% id in 156 aa Orthologue of E. coli yadC
FT (YADC_ECOLI); Fasta hit to YADC_ECOLI (412 aa), 34%
FT identity in 407 aa overlap"
FT /db_xref="GI:16759175"
FT /codon_start=1
FT /product="putative fimbrial protein"
FT /translation="MLLRNTTF AAFLLMMAIMSNCYAINTTLAVGDYASSEHDGPSGD
FT SVFTDNSHNFQGTIAIHKETALRQITVFNWSGIQVVMETFCNGSGNHTYLQLTHNYISA
FT GKSYNHGPLYKTSIPGFYFTIEMTFDPAENMTSSTFWFDKISTPITSEFTEPPSACSR
FT TNVYSLGKLMYGLKIYAVVDSDFAPTEAQLQSFTLSKNGSDDFYIDNPGSGLSNYKMK
FT FNLAAATGLKAVWPTCSASTISGTVNGSTVVKLGSFYPKQIMEGLSRTKFQINLSSQYI
FT NNIEVKLASNNVGTKNTSLLTNNSTSNMNASGIGVLIKGLKSSSSAQNVLKPNDSSEIY
FT KDTTNNVTDGSPVGSATKSLFYQATLKPDDNRTINPGDFKATAQFSITTP"
  
```



Microarrays - BASE

- Open Source
- Free and Easy to set up
- Handles
 - Array Production (LIMS)
 - Experimental Annotation
 - Experimental Results
 - Experimental Analysis
 - MIAME



The screenshot shows a web browser window titled "BASE 1.2.7 - Microsoft Internet Explorer". The address bar displays "http://lshc-linux03/base/". The page content includes a "BASE" logo, a login form with "login:" and "pass:" fields, and a "Log" button. Below the login form are links for "GUI settings", "Site info", "Report a bug", and "BASE project site". The main text describes BASE as a comprehensive free web-based database solution for microarray analysis, released under the GNU General Public License. It provides instructions for subscribing to the mailing list and contacting the local administrator, Michael Watson, for questions. At the bottom, it mentions that the development of BASE is supported by the Knut and Alice Wallenberg Foundation through the SWEGENE consortium, the Swedish Cancer Society, and Lund University.

MAGE-ML

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<!DOCTYPE MAGE-ML (View Source for full doctype...)>
- <MAGE-ML identifier="BASE:dev:MAGE-ML">
- <BioMaterial_package>
- <Compound_assnlist>
  <Compound identifier="COMP:IAH-C-2" isSolvent="false" name="2" />
  <Compound identifier="COMP:IAH-C-1" isSolvent="false" name="1" />
</Compound_assnlist>
- <BioMaterial_assnlist>
- <LabeledExtract identifier="S:IAH-C-labeledextract-2" name="in vivo exp1 Cy3.e1.l1">
- <MaterialType_assn>
  <OntologyEntry category="MaterialType" value="unknown" />
</MaterialType_assn>
- <Treatments_assnlist>
- <Treatment identifier="T:IAH-C-le2" order="0">
+ <ProtocolApplications_assnlist>
- <Action_assn>
  <OntologyEntry category="unknown" value="Labeling" />
</Action_assn>
+ <SourceBioMaterialMeasurements_assnlist>
</Treatment>
</Treatments_assnlist>
+ <Labels_assnreflist>
</LabeledExtract>
- <BioSample identifier="S:IAH-C-biosample-2" name="in vivo exp1 Cy3.e1">
- <MaterialType_assn>
  <OntologyEntry category="MaterialType" value="unknown" />
</MaterialType_assn>
- <Treatments_assnlist>
- <Treatment identifier="T:IAH-C-ex2" order="0">
  <ProtocolApplications_assnlist>
- <Action_assn>
  <OntologyEntry category="unknown" value="Extraction" />
```

ArrayPipe: Normalisation Pipeline

- Based on R and Bioconductor
- Setup:

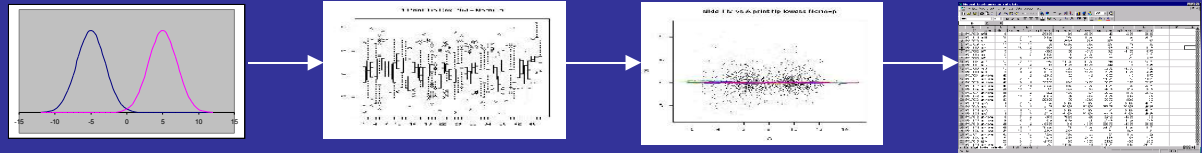
- Apache

- R

- Perl

- Xvfb

- Implements a Flexible user-driven pipeline



Pipeline

http://iahc-linux03/cgi-bin/array/start_pipeline.cgi?dir=28492 - Microsoft Internet Explorer

File Edit View Favorites Tools Help

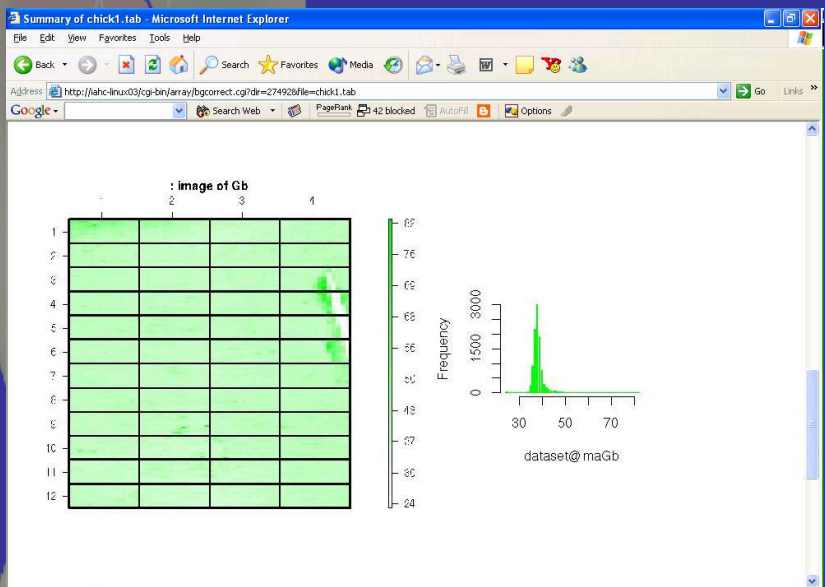
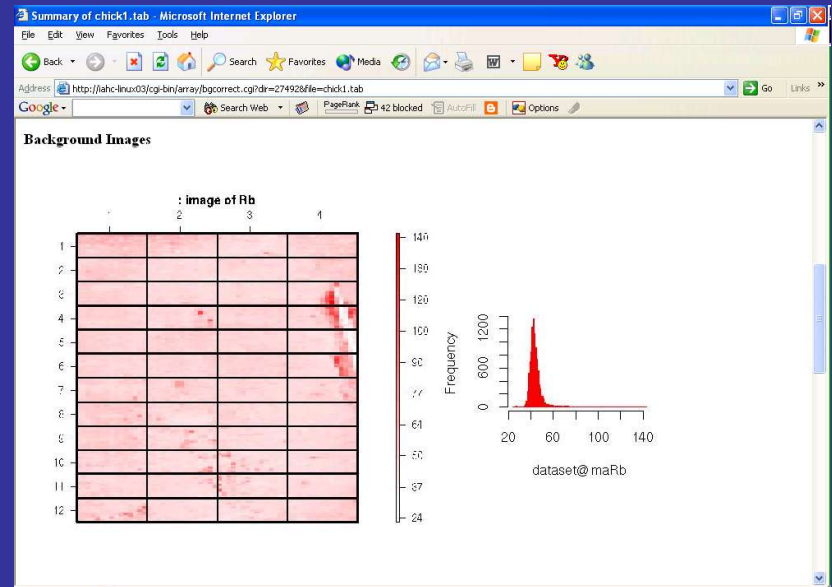
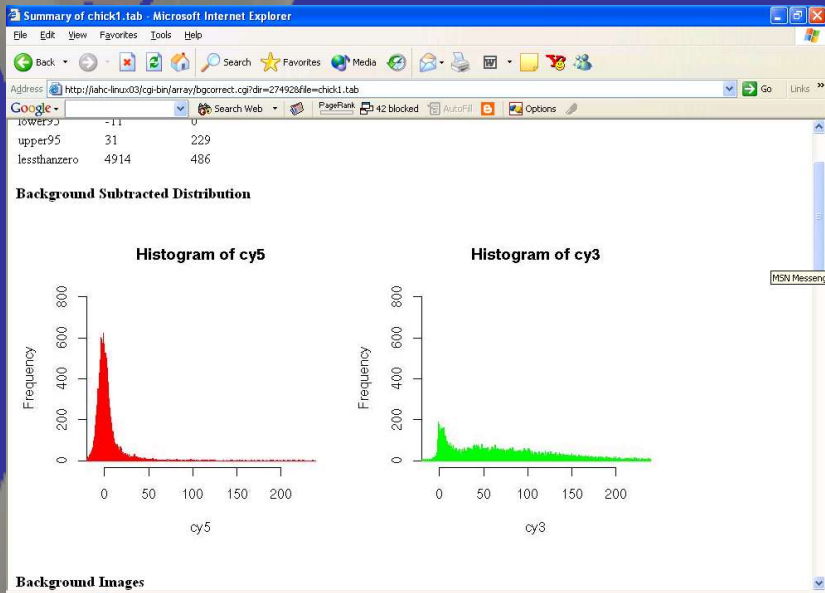
Back Forward Stop Home Search Favorites Media Mail Print Word Pad Options

Address http://iahc-linux03/cgi-bin/array/start_pipeline.cgi?dir=28492 Go Links

Google Search Web PageRank 42 blocked AutoFill Options

Root	Dataset	BG Corrected	Normalised	Averaged
[D]	chick1a.tab [D,BG Correct]	chick1a.14276.bgc [D,H,Visualise,Normalise]	chick1a.14276.14358.norm [D,H,Avg Replicates]	chick1a.14276.14358.14530.avg [D]
		chick1a.1560.bgc [D,H,Visualise,Normalise]	chick1a.1560.17384.norm [D,H,Avg Replicates]	chick1a.1560.17384.8809.avg [D]
		chick1a.15813.bgc [D,H,Visualise,Normalise]	chick1a.15813.15938.norm [D,H,Avg Replicates]	chick1a.15813.15938.16077.avg [D]
[D]	chick3.tab [D,BG Correct]	chick3.23167.bgc [D,H,Visualise,Normalise]	chick3.23167.23182.norm [D,H,Avg Replicates]	chick3.23167.23182.8810.avg [D]
		chick3.14284.bgc [D,H,Visualise,Normalise]	chick3.14284.14373.norm [D,H,Avg Replicates]	chick3.14284.14373.14532.avg [D]
		chick3.15882.bgc [D,H,Visualise,Normalise]	chick3.15882.15953.norm [D,H,Avg Replicates]	chick3.15882.15953.16081.avg [D]
[D]	chick4.tab [D,BG Correct]	chick4.15890.bgc [D,H,Visualise,Normalise]	chick4.15890.15968.norm [D,H,Avg Replicates]	chick4.15890.15968.16083.avg [D]
		chick4.7920.bgc [D,H,Visualise,Normalise]	chick4.7920.7935.norm [D,H,Avg Replicates]	chick4.7920.7935.8811.avg [D]
		chick4.14292.bgc [D,H,Visualise,Normalise]	chick4.14292.14388.norm [D,H,Avg Replicates]	chick4.14292.14388.14540.avg [D]
		chick5.7953.bgc [D,H,Visualise,Normalise]	chick5.7953.7961.norm [D,H,Avg Replicates]	chick5.7953.7961.8812.avg [D]
		chick5.14318.bgc [D,H,Visualise,Normalise]	chick5.14318.14403.norm [D,H,Avg Replicates]	chick5.14318.14403.14545.avg [D]
		chick5.15898.bgc [D,H,Visualise,Normalise]	chick5.15898.15983.norm [D,H,Avg Replicates]	chick5.15898.15983.16085.avg [D]

Background Subtraction



Summary of chick1.tab - Microsoft Internet Explorer

Address: <http://lhc-linux03/cg-bin/array/bgcorrect.cgi?dr=274928file=chick1.tab>

Google Search Web PageRank 42 blocked AutoFill Options

Adjust

Cy5

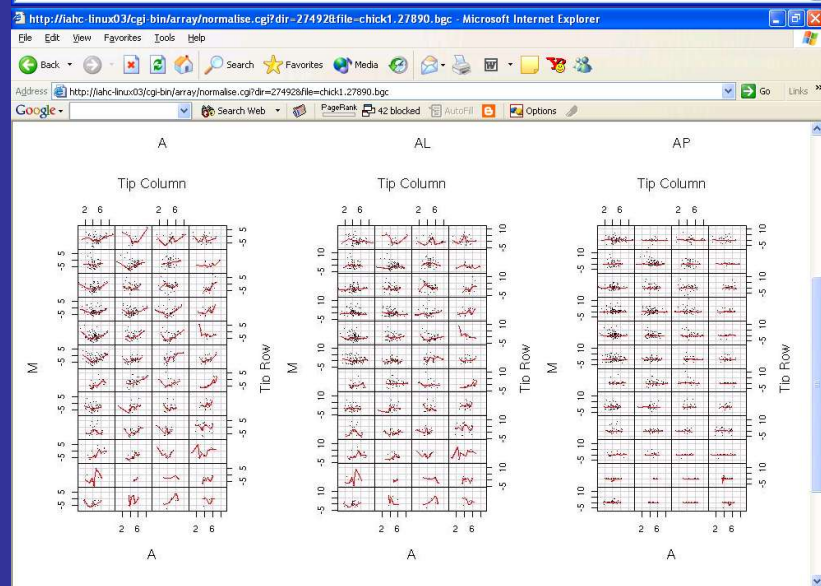
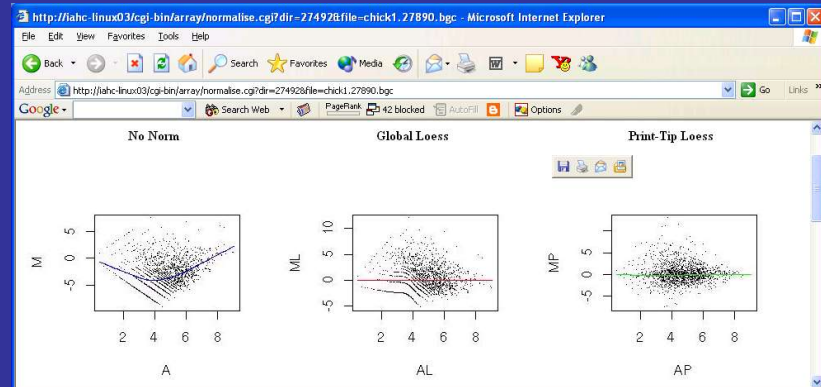
- Subtract Local Background and then add arbitrary value
- Subtract Local Background and then adjust so 95% > 0 (95% lowest is -11)
- Do not subtract background...
- Use this value as background and subtract from all spots...
- Use this value as background and subtract from all spots...**

Cy3

- Subtract Local Background and then add arbitrary value
- Subtract Local Background and then adjust so 95% > 0 (95% lowest is 0)
- Do not subtract background...
- Use this value as background and subtract from all spots...
- Use this value as background and subtract from all spots...**

Normalisation

- M v A Plots
- Global
- Print-tip
- Normalisation
- None
- Median
- Loess
- Print-tip Loess
- Print-tip Loess + MAD



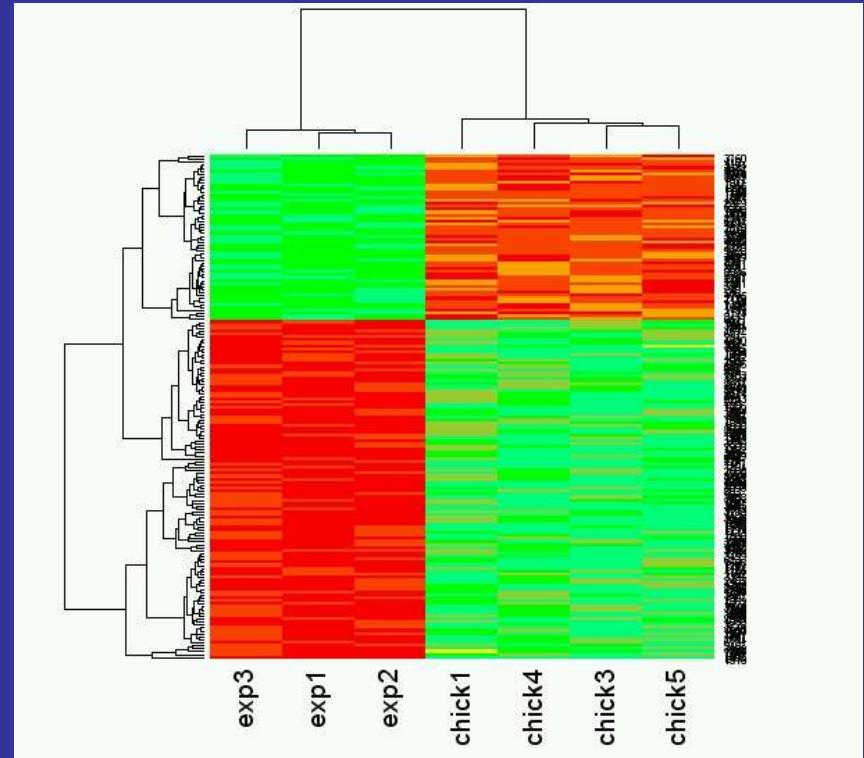
The screenshot shows a 'Normalise' form with the following options:

- No Normalisation
- Median Normalisation
- Loess Normalisation
- Print-tip Loess Normalisation
- Print-tip Loess Normalisation followed by print-tip MAD scale normalisation

The 'normalise' input field is currently empty.

Microarray Analysis

- Use R/Bioconductor
- Limma
 - Linear models
 - Bayesian Statistics
- Cluster Analysis
- PCA



```
d <- read.table("myData.txt", header=TRUE, sep="\t", quote="")
smalld <- d[d$p<0.001,]
dd <- dist(smalld[,3:9], method="euclidean")
h <- hclust(dd, method="complete")
td <- t(smalld[,3:9])
tdd <- dist(td, method="euclidean")
th <- hclust(tdd, method="complete")
col = c("green", "springgreen", "yellowgreen", "yellow", "orange", "orangered", "red")
heatmap(as.matrix(smalld[,3:9]), as.dendrogram(h), as.dendrogram(th), col = mycol)
```



Open Source Proteomics Bioinformatics

???