

Arrays Analysis With GIANT

Julie Dubois-Chevalier & Jimmy Vandel

Version January 2019

Files required to start an analysis (datasets in your history)

Expression

Need to Upload

- CEL files
- Or Normalized Expressions file (tabular format)



Study Design

Upload or

Create in Galaxy*

- Conditions File (tabular format) !
- * Recommended to create in Galaxy to avoid mistakes in file (cf. Next Slide)



Precise the format when you upload your datasets

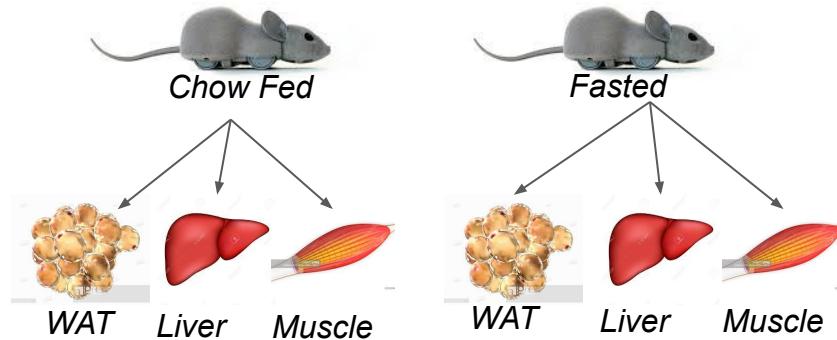
Example Dataset

Expression

GSE46495

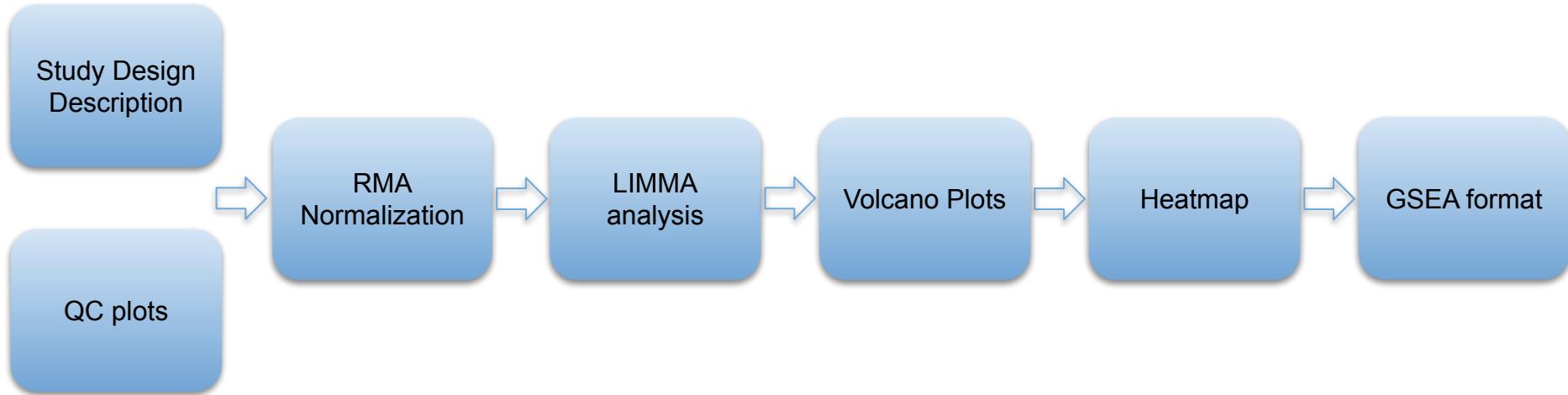
Transcriptome signature of white adipose tissue,
liver, and skeletal muscle in 24 hours fasted mice
(C57Bl/6J)

- CEL files



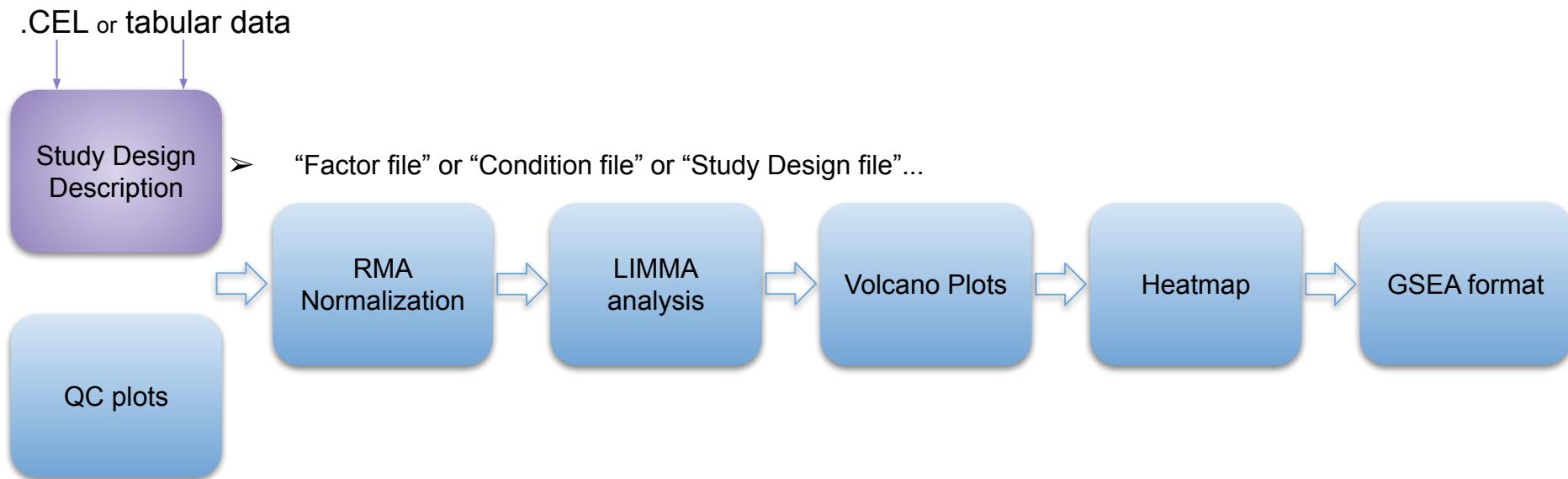
GIANT - Tools Suite

Microarray analysis pipeline: “ from .CEL to GSEA “



GIANT - Tools Suite

Microarray analysis pipeline: “ from .CEL to GSEA ”



Condition File Generator

GIANT

[GIANT-QC Plots](#) Descriptive plots of .CEL collections or normalized expression data

[GIANT-Heatmap and Hierarchical clustering](#) Run hierarchical clustering and plot heatmap from expression data and/or differential expression analysis

[GIANT-Normalization with APT](#)
[Summarize](#) Apply Affymetrix Power Tool summarize function to .CEL collection

[GIANT-GSEA Formating](#) Format input files for GSEA software

[GIANT-Differential Expression with LIMMA](#) Use LIMMA to detect differentially expressed genes

1 [GIANT-Factor file generator](#)
Generate factor file used by other GIANT tools

[GIANT-Plot volcanos](#) Plot volcano from tabular file

5 Add Value as much as useful for your first condition and concerned files + Add Conditions as much as useful
And execute !

2 Create a first condition

3 Choose 1st condition name (ex: Tissue ,Strain, Treatment...)

4 Click on each CEL file concerned by this condition & this value to add it in the list above

Choose a title
Expression tabular file
.CEL files

Choose to select
Normalized data or .CEL

Select file(s) of study

Value
1: Value
Value name
Liver

Choose 1st value of 1st condition (ex: dms0, gw, wt, ko, liver, white_adipose_tissue...)

Select sample sharing this value
Select/Unselect all

GSM1131302_3502_19485_fastedL5_MoGene1_1ST.CEL
GSM1131301_3502_19484_fastedL4_MoGene1_1ST.CEL

GSM1131307_3502_19490_fastedM5_MoGene1_1ST.CEL
GSM1131306_3502_19489_fastedM4_MoGene1_1ST.CEL
GSM1131305_3502_19488_fastedM3_MoGene1_1ST.CEL
GSM1131304_3502_19487_fastedM2_MoGene1_1ST.CEL
GSM1131303_3502_19486_fastedM1_MoGene1_1ST.CEL

GSM1131300_3502_19483_fastedL3_MoGene1_1ST.CEL

GSM1131299_3502_19482_fastedL2_MoGene1_1ST.CEL
GSM1131298_3502_19481_fastedL1_MoGene1_1ST.CEL

Parameters

Condition File Generator

2 results in history :

- log file
- Tabular file

The screenshot shows a software interface with a sidebar labeled "History" containing a search bar and a list of datasets. Below this is a section titled "Example_history" showing "32 shown" datasets with a total size of "337.7 MB". Two specific entries are highlighted with red boxes:

- 32: ConditionsGenerator toPersonalize Log**: This entry has a red box around its preview icon.
- 31: ConditionsGenerator or toPersonalize conditionsFile**: This entry also has a red box around its preview icon.

Below these, there are two more entries:

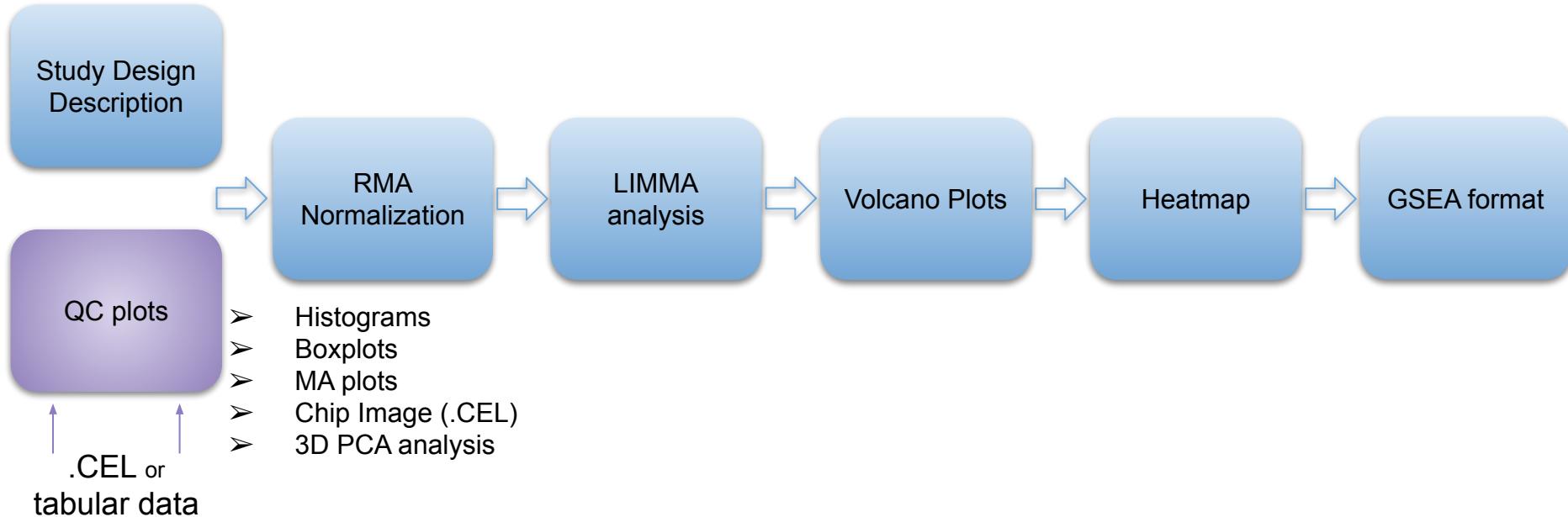
- 30: GSM1131278 3502_19461 fedF1 MoGene1_1ST.CEL**
- 29: GSM1131279 3502_19462 fedF2 MoGene1_1ST.CEL**

Each entry has a "View data" button at the bottom right.

1	2	3
Conditions	Tissue	FastFed
GSM1131307_3502_19490_fastedM5_MoGene1_1ST.CEL	Muscle	Fasted
GSM1131306_3502_19489_fastedM4_MoGene1_1ST.CEL	Muscle	Fasted
GSM1131305_3502_19488_fastedM3_MoGene1_1ST.CEL	Muscle	Fasted
GSM1131304_3502_19487_fastedM2_MoGene1_1ST.CEL	Muscle	Fasted
GSM1131303_3502_19486_fastedM1_MoGene1_1ST.CEL	Muscle	Fasted
GSM1131302_3502_19485_fastedL5_MoGene1_1ST.CEL	Liver	Fasted
GSM1131301_3502_19484_fastedL4_MoGene1_1ST.CEL	Liver	Fasted
GSM1131300_3502_19483_fastedL3_MoGene1_1ST.CEL	Liver	Fasted
GSM1131299_3502_19482_fastedL2_MoGene1_1ST.CEL	Liver	Fasted
GSM1131298_3502_19481_fastedL1_MoGene1_1ST.CEL	Liver	Fasted
GSM1131297_3502_19480_fastedF5_MoGene1_1ST.CEL	WAT	Fasted
GSM1131296_3502_19479_fastedF4_MoGene1_1ST.CEL	WAT	Fasted
GSM1131295_3502_19478_fastedF3_MoGene1_1ST.CEL	WAT	Fasted
GSM1131294_3502_19477_fastedF2_MoGene1_1ST.CEL	WAT	Fasted
GSM1131293_3502_19476_fastedF1_MoGene1_1ST.CEL	WAT	Fasted
GSM1131292_3502_19475_fedM5_MoGene1_1ST.CEL	Muscle	Fed
GSM1131291_3502_19474_fedM4_MoGene1_1ST.CEL	Muscle	Fed
GSM1131290_3502_19473_fedM3_MoGene1_1ST.CEL	Muscle	Fed
GSM1131289_3502_19472_fedM2_MoGene1_1ST.CEL	Muscle	Fed
GSM1131288_3502_19471_fedM1_MoGene1_1ST.CEL	Muscle	Fed
GSM1131287_3502_19470_fedL5_MoGene1_1ST.CEL	Liver	Fed
GSM1131286_3502_19469_fedL4_MoGene1_1ST.CEL	Liver	Fed
GSM1131285_3502_19468_fedL3_MoGene1_1ST.CEL	Liver	Fed
GSM1131284_3502_19467_fedL2_MoGene1_1ST.CEL	Liver	Fed
GSM1131283_3502_19466_fedL1_MoGene1_1ST.CEL	Liver	Fed
GSM1131282_3502_19465_fedF5_MoGene1_1ST.CEL	WAT	Fed
GSM1131281_3502_19464_fedF4_MoGene1_1ST.CEL	WAT	Fed
GSM1131280_3502_19463_fedF3_MoGene1_1ST.CEL	WAT	Fed
GSM1131279_3502_19462_fedF2_MoGene1_1ST.CEL	WAT	Fed
GSM1131278_3502_19461_fedF1_MoGene1_1ST.CEL	WAT	Fed

GIANT - Tools Suite

Microarray analysis pipeline: “ from .CEL to GSEA ”



QC plots

GIANT

[GIANT-QC Plots](#) Descriptive plots of .CEL collections or normalized expression data

[GIANT-Heatmap and Hierarchical clustering](#) Run hierarchical clustering and plot heatmap from expression data and/or differential expression analysis

[GIANT-Normalization with APT Summarize](#) Apply Affymetrix Power Tool summarize function to .CEL collection

[GIANT-GSEA Formating](#) Format input files for GSEA software

[GIANT-Differential Expression with LIMMA](#) Use LIMMA to detect differentially expressed genes

GIANT-QC Plots Descriptive plots of .CEL collections or normalized expression data (Galaxy Version 0.1) ▼ Options

Title for output
QCplot_BeforeNormalization

Select one .CEL collection or one tabular file

31: conditions.txt
30: GSM1131307_3502_19490_fastedM5_MoGene1_1ST.CEL
29: GSM1131306_3502_19489_fastedM4_MoGene1_1ST.CEL
28: GSM1131305_3502_19488_fastedM3_MoGene1_1ST.CEL
27: GSM1131304_3502_19487_fastedM2_MoGene1_1ST.CEL

Plots selection

Plot histograms
 Yes No
Plot intensity distribution for each condition (pm probes for .cel)

Plot MA plots
 Yes No
Plot MA plot for each condition, median value is used as reference

Plot boxplots
 Yes No
Plot intensity through boxplot for each condition (pm probes for .cel)

Display microarrays image (only for .CEL files)
 Yes No

PCA analysis

Plot 3D PCA
 Yes No
3D plot of conditions in the space defined by the 3 principal components

Factor information tabular file (optional)
 Nothing selected

Select factor informations to display (optional)

Advanced parameters

Execute

Choose a title

Select CEL files

Select desired plots



PCA analysis with conditions infos

And execute !

QC plots

2 results in history :

- log file
- HTML page

The screenshot shows the TestGalaxy GSE46495 history page. At the top is a search bar labeled "Rechercher des données". Below it, the title "TestGalaxy_GSE46495" is followed by "33 shown". A file size of "413.94 MB" is listed. To the right are three icons: a checkmark, a clipboard, and a speech bubble.

The history list contains the following items:

- 33: QCplot BeforeNorm alization Log.** (highlighted with a red box and a red circle around the eye icon)
- 32: QCplot BeforeNorm alization HTML.html** (highlighted with a red box and a red circle around the eye icon)
- 31: conditions.txt**
- 30: GSM1131307 3502_19490 fastedM5 MoGene1 1ST.CEL**
- 29: GSM1131306 3502_19489 fastedM4 MoGene1 1ST.CEL**

Html page

Histograms

[Histograms1](#)

Boxplots

[Boxplots1](#)

MA plots (show/hide)

[MAplot GSM1131287 3502_19470 fedL5 MoGene1 1ST.CEL](#)

[MAplot GSM1131288 3502_19471 fedM1 MoGene1 1ST.CEL](#)

[MAplot GSM1131289 3502_19472 fedM2 MoGene1 1ST.CEL](#)

[MAplot GSM1131290 3502_19473 fedM3 MoGene1 1ST.CEL](#)

[MAplot GSM1131291 3502_19474 fedM4 MoGene1 1ST.CEL](#)

[MAplot GSM1131292 3502_19475 fedM5 MoGene1 1ST.CEL](#)

[MAplot GSM1131293 3502_19476 fastedF1 MoGene1 1ST.CEL](#)

[MAplot GSM1131294 3502_19477 fastedF2 MoGene1 1ST.CEL](#)

[MAplot GSM1131295 3502_19478 fastedF3 MoGene1 1ST.CEL](#)

Microarray

[Microarray GSM1131287 3502_19470 fedL5 MoGene1 1ST.CEL](#)

[Microarray GSM1131288 3502_19471 fedM1 MoGene1 1ST.CEL](#)

[Microarray GSM1131289 3502_19472 fedM2 MoGene1 1ST.CEL](#)

[Microarray GSM1131290 3502_19473 fedM3 MoGene1 1ST.CEL](#)

[Microarray GSM1131291 3502_19474 fedM4 MoGene1 1ST.CEL](#)

[Microarray GSM1131292 3502_19475 fedM5 MoGene1 1ST.CEL](#)

[Microarray GSM1131293 3502_19476 fastedF1 MoGene1 1ST.CEL](#)

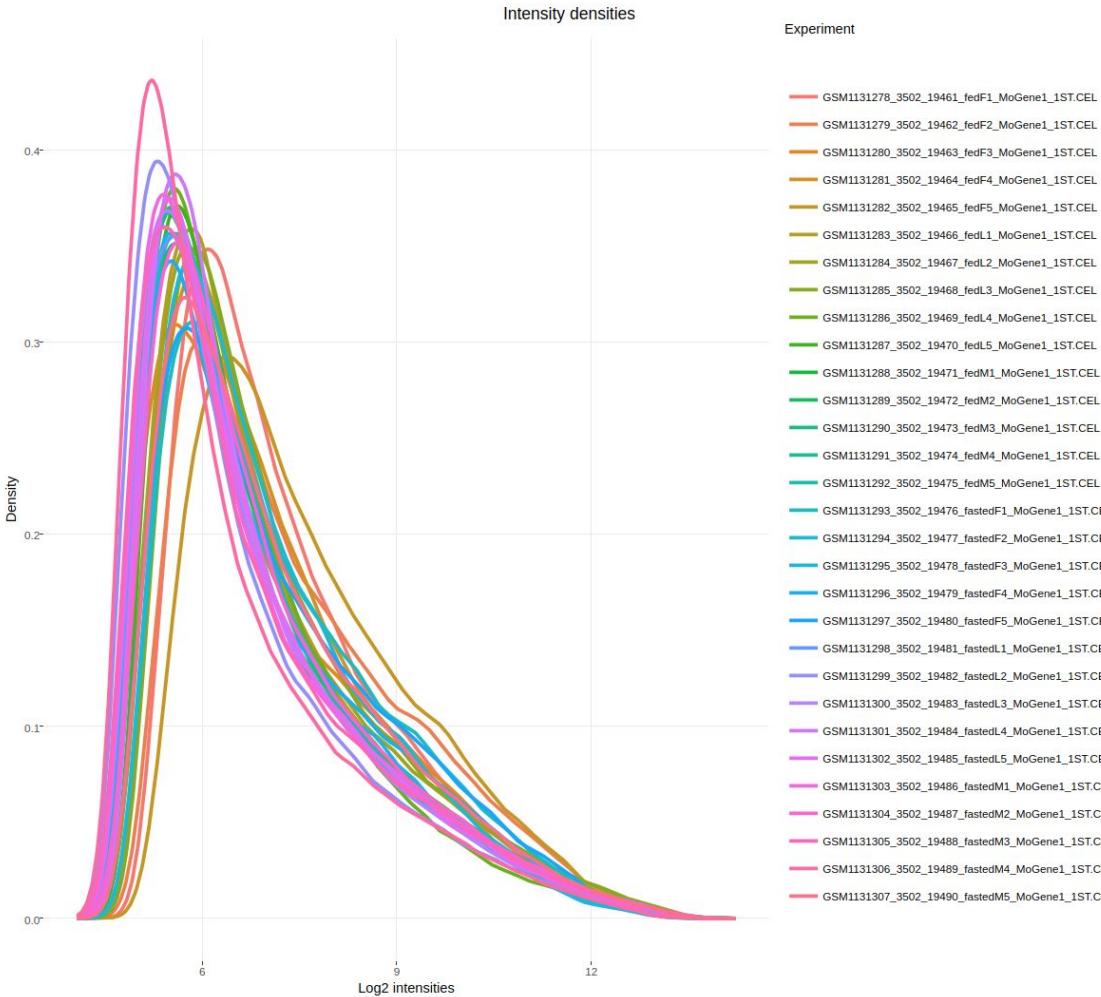
[Microarray GSM1131294 3502_19477 fastedF2 MoGene1 1ST.CEL](#)

[Microarray GSM1131295 3502_19478 fastedF3 MoGene1 1ST.CEL](#)

QC plots

Histograms

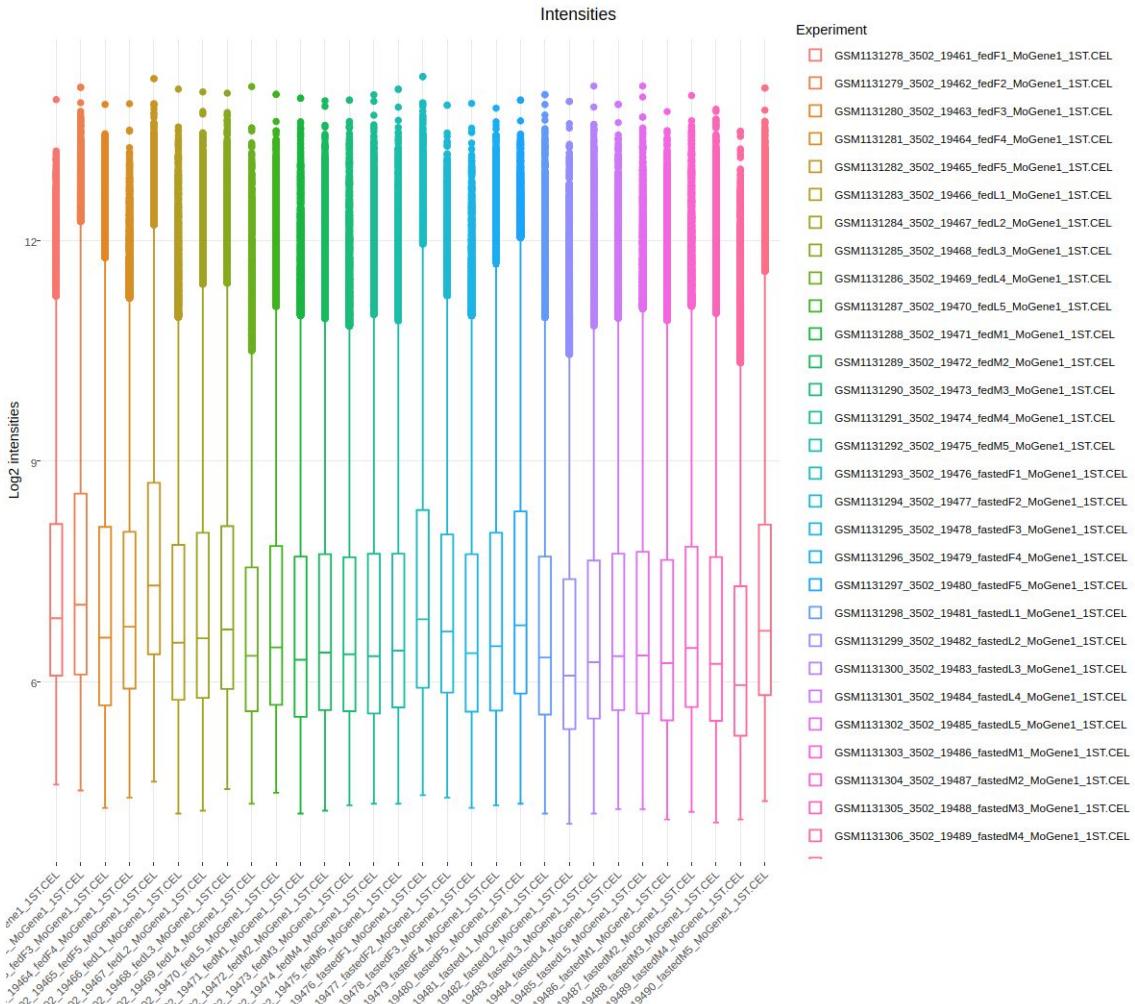
Histograms1



QC plots

Boxplots

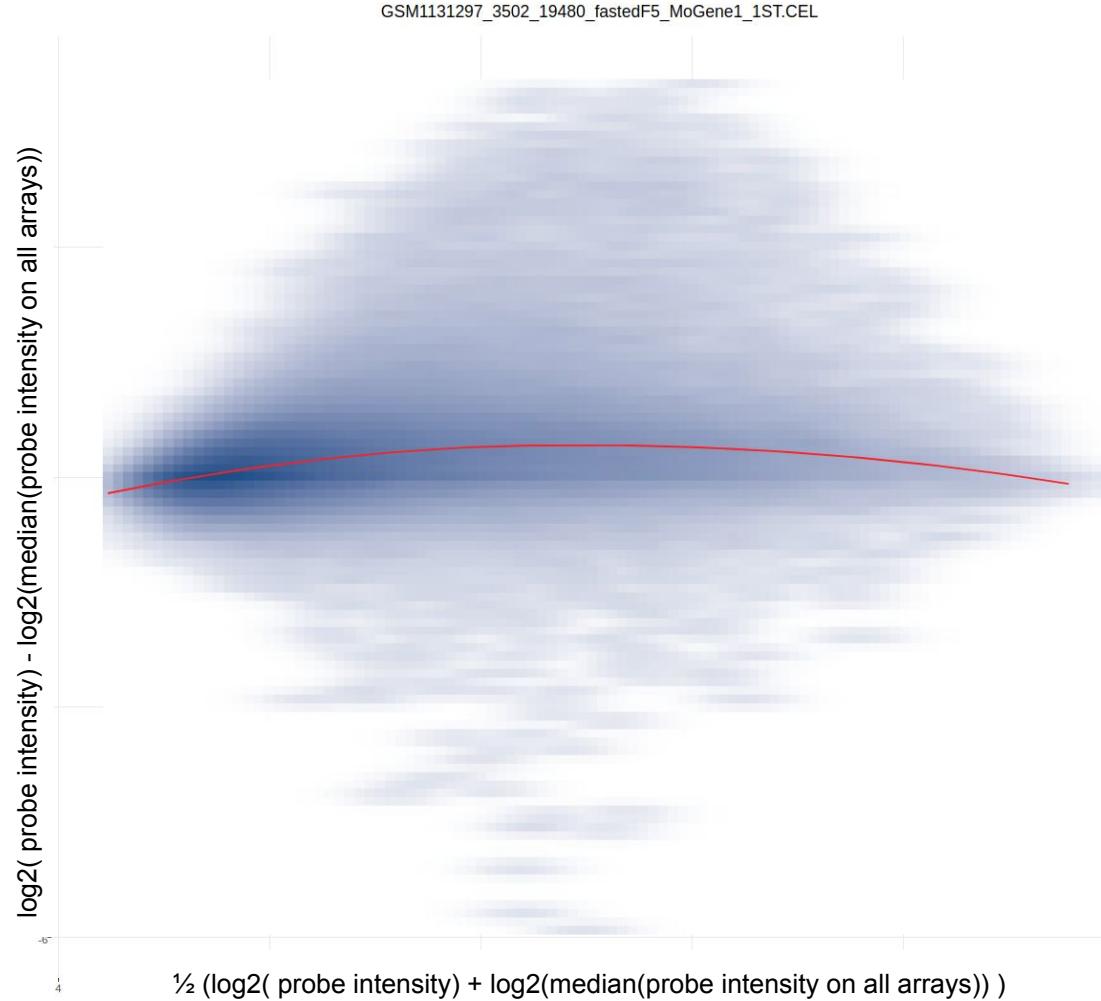
Boxplots1



QC plots

MA plots (show/hide)

[MAplot GSM1131287_3502_19470_fedL5_MoGene1_1ST.CEL](#)
[MAplot GSM1131288_3502_19471_fedM1_MoGene1_1ST.CEL](#)
[MAplot GSM1131289_3502_19472_fedM2_MoGene1_1ST.CEL](#)
[MAplot GSM1131290_3502_19473_fedM3_MoGene1_1ST.CEL](#)
[MAplot GSM1131291_3502_19474_fedM4_MoGene1_1ST.CEL](#)
[MAplot GSM1131292_3502_19475_fedM5_MoGene1_1ST.CEL](#)



QC plots

GIANT

[GIANT-QC Plots](#) Descriptive plots of .CEL collections or normalized expression data

[GIANT-Heatmap and Hierarchical clustering](#) Run hierarchical clustering and plot heatmap from expression data and/or differential expression analysis

[GIANT-Normalization with APT Summarize](#) Apply Affymetrix Power Tool summarize function to .CEL collection

[GIANT-GSEA Formating](#) Format input files for GSEA software

[GIANT-Differential Expression with LIMMA](#) Use LIMMA to detect differentially expressed genes

GIANT-QC Plots Descriptive plots of .CEL collections or normalized expression data (Galaxy Version 0.1) ▼ Options

Title for output
QCplot_AfterNormalization

Select one .CEL collection or one tabular file
 34: APT_GcSstRMA_NormalizedData
31: conditions.txt
30: GSM1131307_3502_19490_fastedM5_MoGene1_1ST.CEL
29: GSM1131306_3502_19489_fastedM4_MoGene1_1ST.CEL
28: GSM1131305_3502_19488_fastedM3_MoGene1_1ST.CEL

Plots selection

Plot histograms
 Yes No
Plot intensity distribution for each condition (pm probes for .cel)

Plot MA plots
 Yes No
Plot MA plot for each condition, median value is used as reference

Plot boxplots
 Yes No
Plot intensity through boxplot for each condition (pm probes for .cel)

Display microarrays image (only for .CEL files)
 Yes No

PCA analysis

Plot 3D PCA
 Yes No
3D plot of conditions in the space defined by the 3 principal components

Factor information tabular file (optional)
 31: conditions.txt

Select factor informations to display (optional)
 Select/Unselect all
 Diet Tissue

Advanced parameters

Execute

Choose a title

Select normalized data

Select desired plots

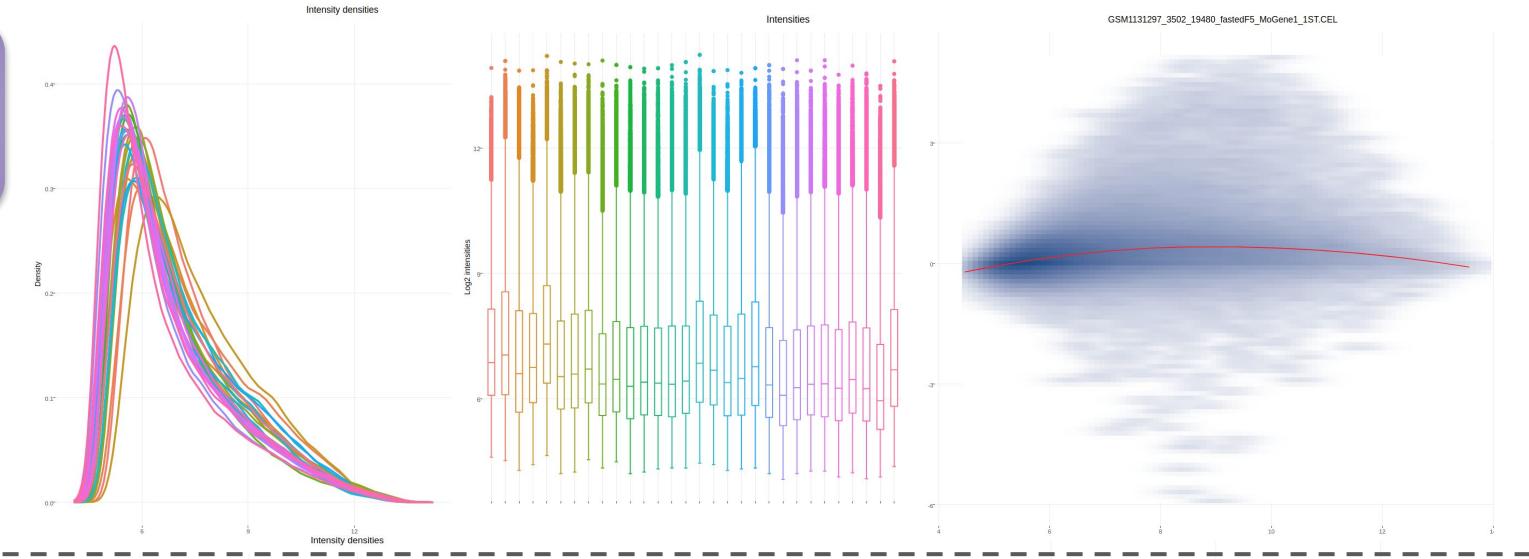
⚠ Microarrays images

PCA analysis with conditions infos

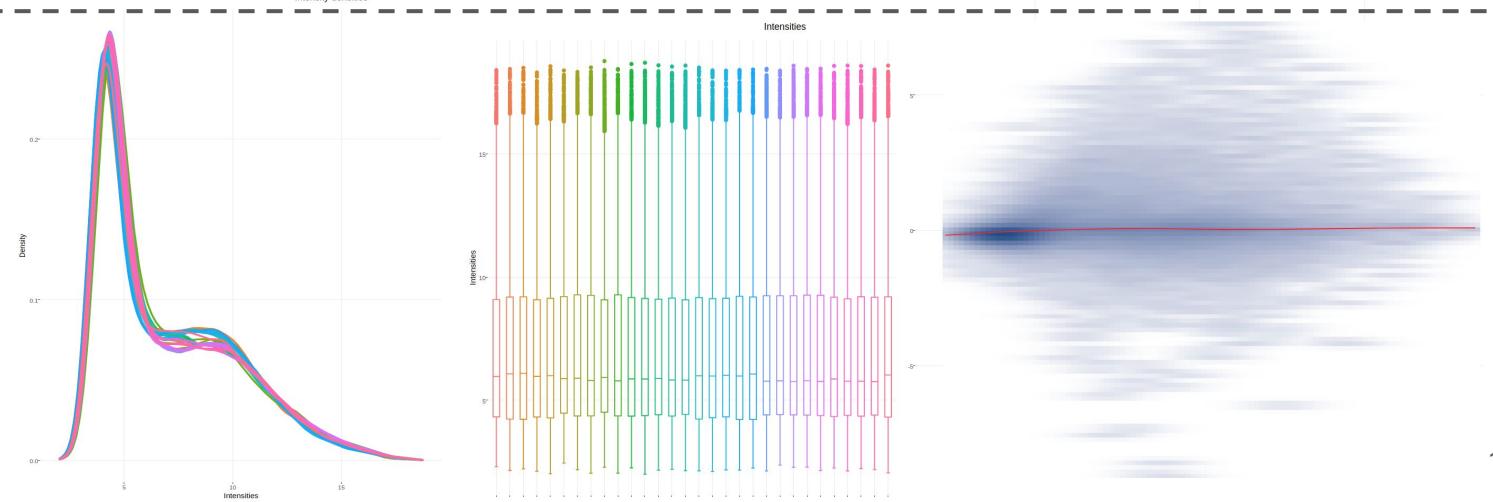
And execute !

QC plots

RAW data



Normalized data



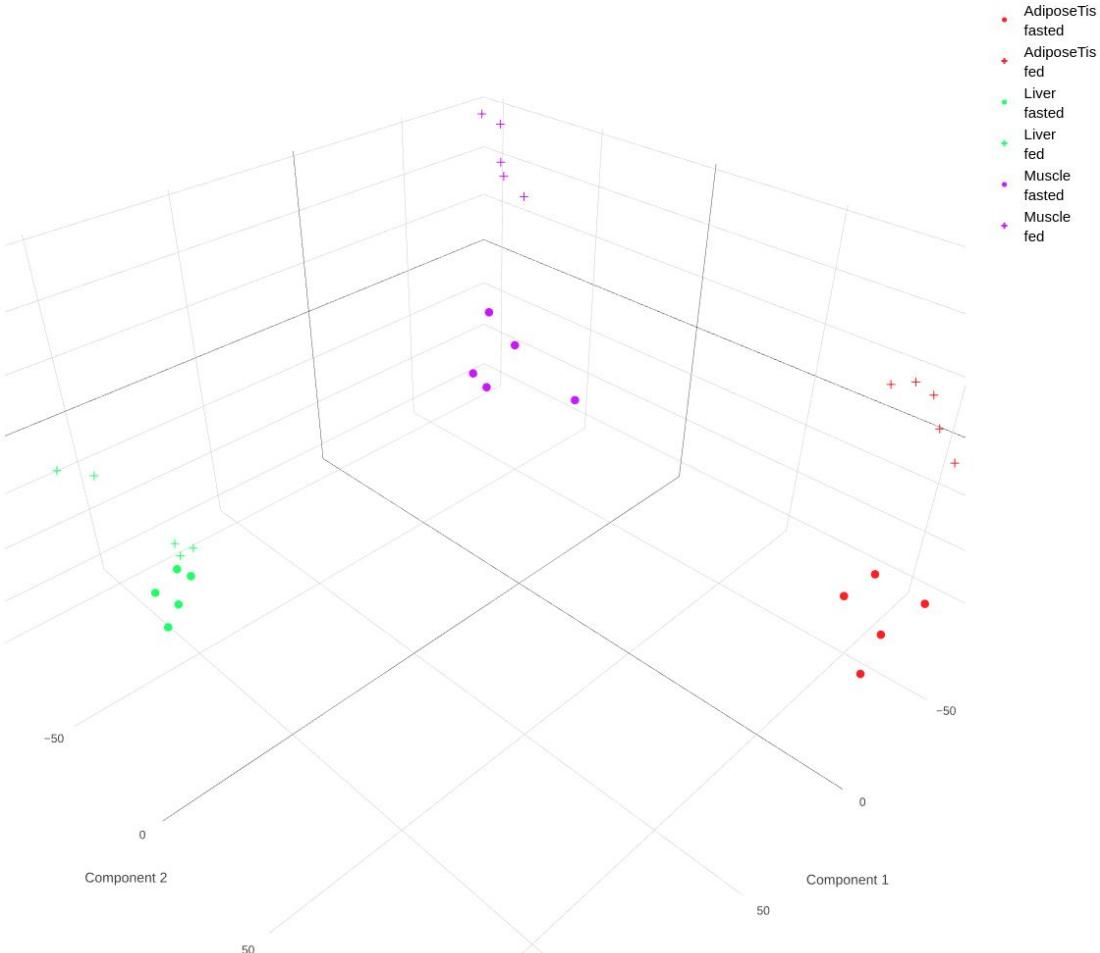
QC plots

PCA

PCA Tissue * Diet

Scree plot

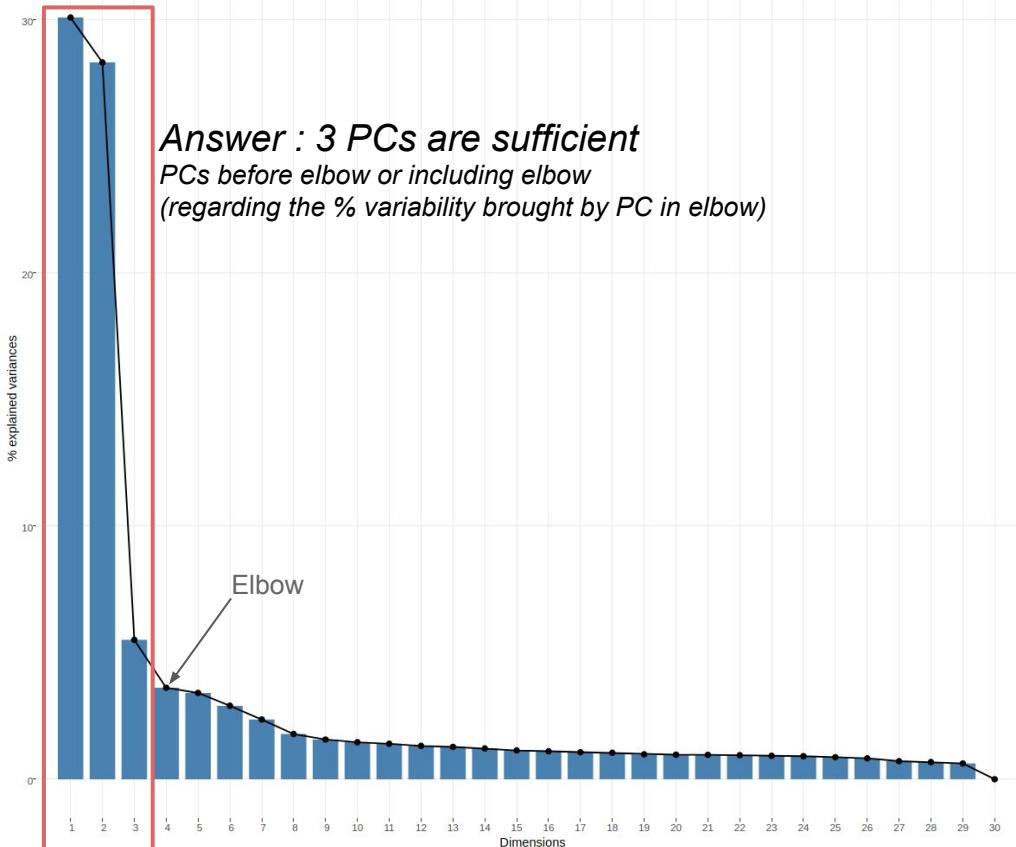
Principal Component Analysis



QC plots

How many Principal Components are sufficient to explain your data ?

Scree plot



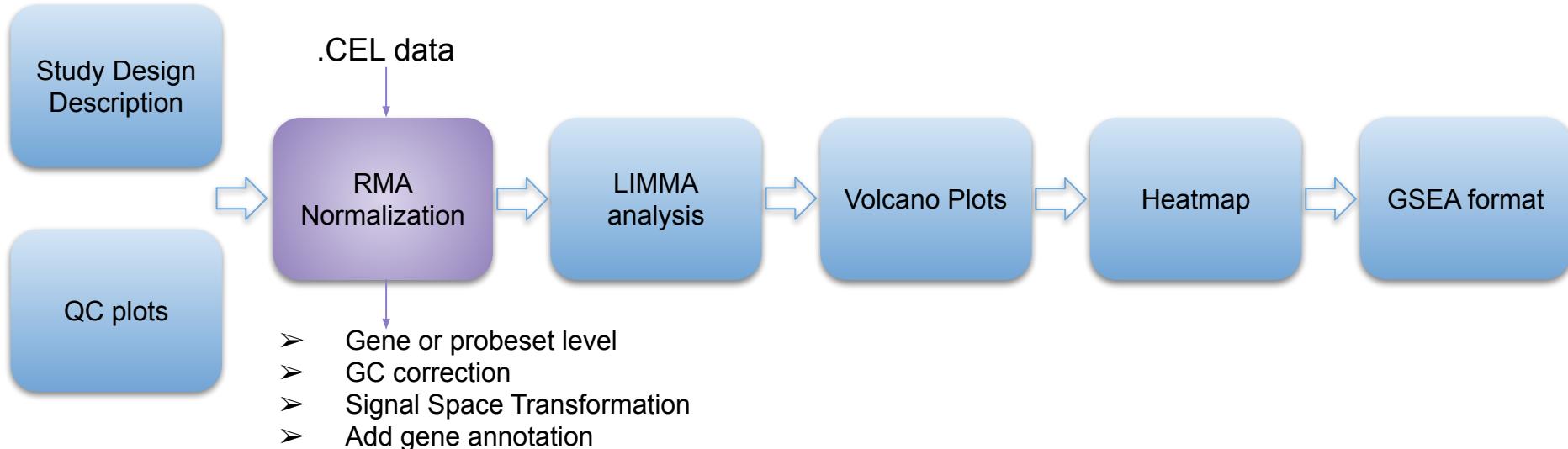
PCA

[PCA Tissue * Diet](#)

[Scree plot](#)

GIANT - Tools Suite

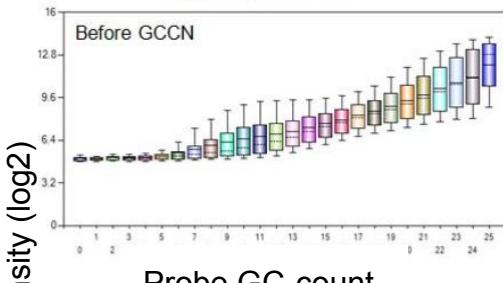
Microarray analysis pipeline: “ from .CEL to GSEA “



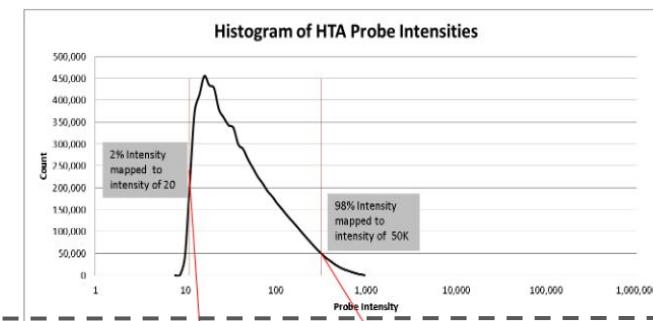
RMA Normalization

What is a Normalization in GIANT ?

GCCN GC correction

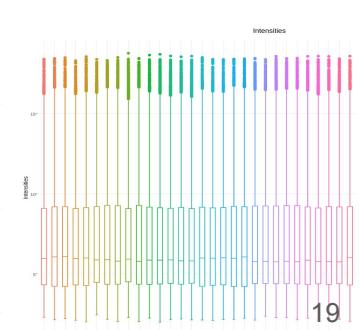
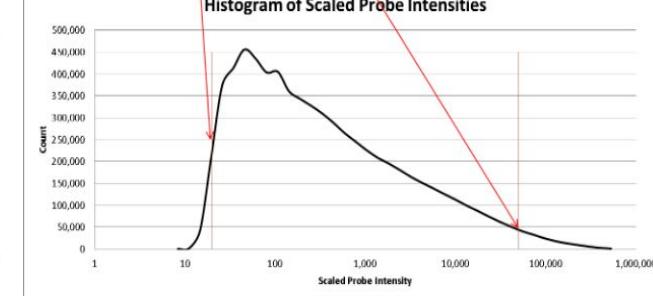
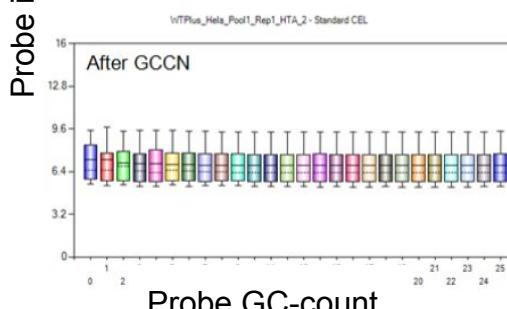
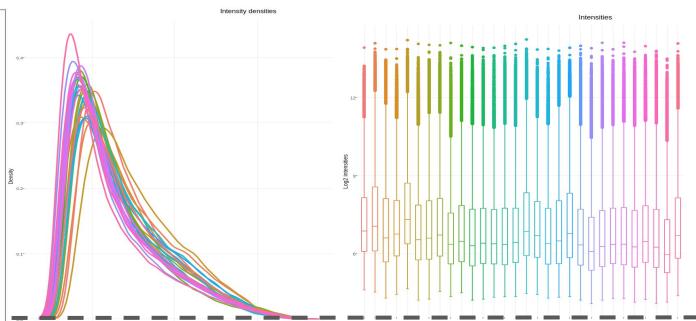


SST scaling



RMA Normalization

- RMA background
- Quantile Normalization
- Median Polish



Normalization in GIANT - recommendations

1. HTA/MTA & Hu/MoGene : use GCcorrection+Scaling+RMA
 - a. Advantages : scaled data are comparable to QPCR & RNA-seq signal level
 - b. Warning for Hu/MoGene, to be similar with an old analysis, you should use RMA only
2. MOE 430 : use Scaling+RMA
 - a. Advantages : scaled data are comparable to QPCR & RNA-seq signal level
 - b. Warning : no available information about GC content, so GC correction option has no effect
3. No Affymetrix data : use a dataset already normalized to perform further analysis steps (differential analysis...)

RMA Normalization

GIANT

[GIANT-QC Plots](#) Descriptive plots of .CEL collections or normalized expression data

[GIANT-Heatmap and Hierarchical clustering](#) Run hierarchical clustering and plot heatmap from expression data and/or differential expression analysis

[GIANT-Normalization with APT](#)
[Summarize Apply Affymetrix Power Tool summarize function to .CEL collection](#)

[GIANT-GSEA Formating](#) Format input files for GSEA software

[GIANT-Differential Expression with LIMMA](#) Use LIMMA to detect differentially expressed genes

For Post-Normalization Processing

GIANT-Normalization with APT Summarize Apply Affymetrix Power Tool summarize function to .CEL collection (Galaxy Version 0.1) Options

Title for output
APT_GcStRMA

.cel collection file

30: GSM1131307_3502_19490_fastedM5_MoGene1_1ST.CEL
29: GSM1131306_3502_19489_fastedM4_MoGene1_1ST.CEL
28: GSM1131305_3502_19488_fastedM3_MoGene1_1ST.CEL
27: GSM1131304_3502_19487_fastedM2_MoGene1_1ST.CEL
26: GSM1131303_3502_19486_fastedM1_MoGene1_1ST.CEL

Normalization to perform
 gc correction + scale intensity + rma
 scale intensity + rma
 gc correction + rma
 rma
For more details go to APT webpage

Normalization level
 Core genes
 Probe set
'Core genes' option is not available for all arrays

Select GeneChip array kind
common arrays (HTA, HuGene, MoGene)
Name
Mouse Gene 1.1 ST arrays

Add gene annotation
Yes No

Discard probe set without gene annotation
Yes No

Merging approach for probe set with same gene annotation
 No merging
 Mean between probes [recommended]
 Keep probe with higher variance
 Keep probe with lower variance

Execute

Choose a title

Select CEL files

Select normalization options

 GC correction

Select GeneChip kind

Add gene annotation

And execute !

RMA Normalization

- 2 results in history :
- log file
 - normalized data

Rechercher des données

TestGalaxy_GSE46495

35 shown

419.92 MB

35: APT_GcSstRMA_Log

34: APT_GcSstRMA_NormalizedData

33: QCplot_BeforeNormalization_Log.

32: QCplot_BeforeNormalization_HTML.html.

35: APT_GcSstRMA_Log

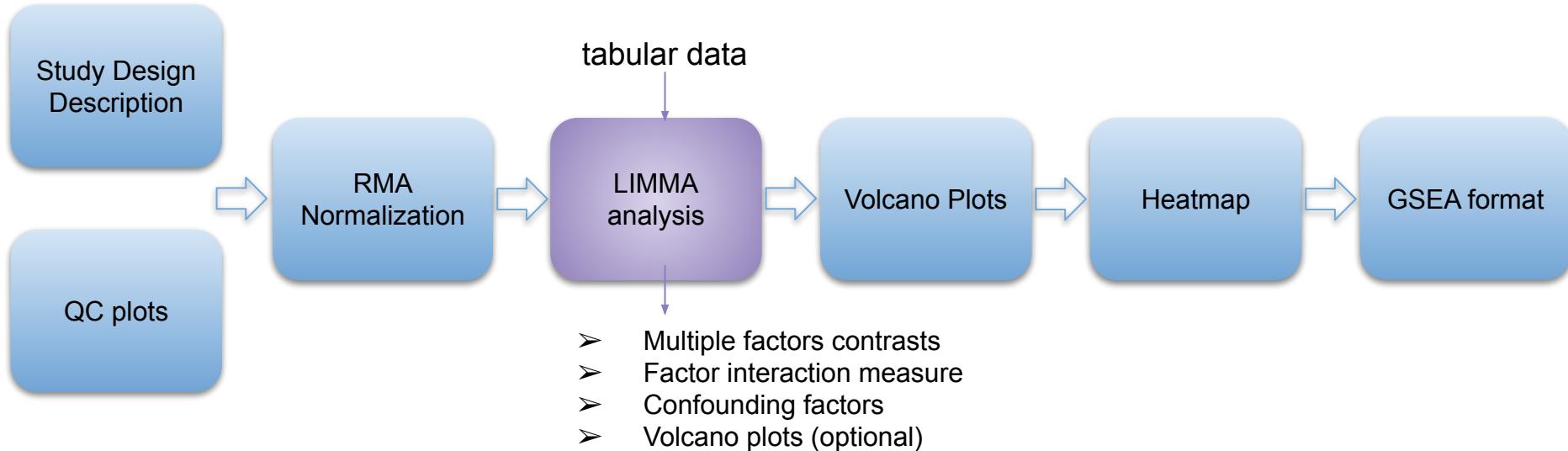
```
Running ProbesetSummarizeEngine...
Opening clf file: MoGene1.1ST.clf
Opening pgf file: MoGene1.1ST.pgf
Reading 241576 probesets.....Done. (0.06 min)
Setting analysis info.
Reading and pre-processing 30 cel files.....Done. (0.08 min)
Processing 1 chipstream
Applying GCCorrection to 30 cel datasets.....Done. (0.13 min)
Applying ScaleIntensities to 30 cel datasets.....Done. (0.16 min)
Applying RMA background transformation 30 cel datasets.....Done. (0.23 min)
Computing sketch normalization for 30 cel datasets.....Done. (0.03 min)
Applying sketch normalization to 30 cel datasets.....Done. (0.13 min)
Finalizing 1 chipstream.
Processing Probesets.....Done. (0.21 min)
Flushing output reporters. Finalizing output.
Done.
```

34: APT_GcSstRMA_NormalizedData

1	2	3	4
Conditions	GSM1131278_3502_19461_fedF1_MoGene1_1ST.CEL	GSM1131279_3502_19462_fedF2_MoGene1_1ST.CEL	GSM1131280_3502_19463_fedF3_MoGene1_1ST.CEL
Cep72	5.79442	6.13562	5.84982
Unc93b1	10.344	10.3944	10.2048
n-R5s217	4.32256	4.14826	4.35658
Gm28388	4.16515	4.16639	4.18478
C920025E04Rik	4.71314	5.56514	5.28329
Fdxr	5.41385	5.48893	5.36192
Cep76	6.91781	6.92847	6.72883
Uba1	13.7473	13.8369	13.596
Cep78	8.03659	8.1785	8.2553

GIANT - Tools Suite

Microarray analysis pipeline: “ from .CEL to GSEA ”



Differential Analysis - Confounding factors

1. What is a confounding factor ?
 - a. Paired-analysis (same individual for 2 or more samples)
 - b. Batch effect : scan dates for example
 - c. Any “blocking” effect of samples not confounded with principal factors
2. How deal with them ?
 - a. In differential analysis : add a column in conditionFile.txt and use the form “confounding factor” option. Possibility to use multiple confounding factors.
 - b. Why we doesn't use existing removeBatch functions or Combat package ?
 - i. It's better to make a global modelisation for the differential analysis. So to take into account all potential effects (principal & confounding) in the model.
 - ii. This kind of data pre-treatment is only dedicated for unbiased data exploratory (as clustering/network inference/co-module expression).

LIMMA analysis

GIANT

[GIANT-QC Plots](#) Descriptive plots of .CEL collections or normalized expression data

[GIANT-Heatmap and Hierarchical clustering](#) Run hierarchical clustering and plot heatmap from expression data and/or differential expression analysis

[GIANT-Normalization with APT Summarize](#) Apply Affymetrix Power Tool summarize function to .CEL collection

[GIANT-GSEA Formating](#) Format input files for GSEA software

[GIANT-Differential Expression with LIMMA](#) Use LIMMA to detect differentially expressed genes

GIANT-Differential Expression with LIMMA Use LIMMA to detect differentially expressed genes (Galaxy Version 0.1) ▼ Options

Input files

Title for output
LIMMA_FedVsFasted

Normalized expression tabular file
34: APT_GcSstRMA_NormalizedData

Factor information tabular file
38: conditions.txt

Contrast definition

Select all factors to include in the global model (excepting confounding factors)

Select/Unselect all

Diet
 Tissue
 MouseID

Confounding factors are selected in the corresponding section below.

Contrast

1: Contrast

Contrast name
FedVsFasted_Liver

Select factor levels of 1st group

Select/Unselect all

fed*AdiposeTis
 fasted*AdiposeTis
 fed*Liver
 fasted*Liver
 fed*Muscle
 fasted*Muscle

Select factor levels of 2nd group

Select/Unselect all

fed*AdiposeTis
 fasted*AdiposeTis
 fed*Liver
 fasted*Liver
 fed*Muscle
 fasted*Muscle

} Choose a title

} Select normalized data and factor information

} Select factors needed for contrasts

⚠ No confounding

} Define a first contrast :
- name
- first group members
- second group members

LIMMA analysis

GIANT

[GIANT-QC Plots](#) Descriptive plots of .CEL collections or normalized expression data

[GIANT-Heatmap and Hierarchical clustering](#) Run hierarchical clustering and plot heatmap from expression data and/or differential expression analysis

[GIANT-Normalization with APT Summarize](#) Apply Affymetrix Power Tool summarize function to .CEL collection

[GIANT-GSEA Formating](#) Format input files for GSEA software

[GIANT-Differential Expression with LIMMA](#) Use LIMMA to detect differentially expressed genes

2: Contrast

Contrast name
FedVsFasted_AllTissues

Select factor levels of 1st group

Select/Unselect all

fed*AdiposeTis
 fasted*AdiposeTis
 fed*Liver
 fasted*Liver
 fed*Muscle
 fasted*Muscle

Select factor levels of 2nd group

Select/Unselect all

fed*AdiposeTis
 fasted*AdiposeTis
 fed*Liver
 fasted*Liver
 fed*Muscle
 fasted*Muscle

+ Insert Contrast

Add interaction contrasts

If you have selected two factors at least.

Select one control group for each factor (and only one)

Select/Unselect all

Diet:fed
 Diet:fasted
 Tissue:AdiposeTis
 Tissue:Liver
 Tissue:Muscle

Paired analysis/confounding factor

Add confounding factors

To control factors producing spurious association as batch effects or to analyze paired data

Select confounding factors

Select/Unselect all

MouseID

Define a second contrast :

- name
- first group members
- second group members

As many as you need !

To compute interaction contrast you need control group **for each factor**

Precise potential confounding factors as paired analysis...

LIMMA analysis

GIANT

[GIANT-QC Plots](#) Descriptive plots of .CEL collections or normalized expression data

[GIANT-Heatmap and Hierarchical clustering](#) Run hierarchical clustering and plot heatmap from expression data and/or differential expression analysis

[GIANT-Normalization with APT Summarize](#) Apply Affymetrix Power Tool summarize function to .CEL collection

[GIANT-GSEA Formating](#) Format input files for GSEA software

[GIANT-Differential Expression with LIMMA](#) Use LIMMA to detect differentially expressed genes

Output section

Output FDR p-val threshold
0.05

Plot histograms
 Yes No
Plot nominal p-val distribution for each comparison.

Plot volcanos
 Yes No
Plot volcano for each comparison.

Fold change threshold for volcanos (both 'log2(threshold)' and 'log2(1/threshold)' values will be used)
2.0

Add gene/probe information
 Yes No

Organism
Mouse genes (GRCm38.p6)

Nature of row names
Gene name

HTML snapshot format
 PNG format SVG format

Advanced parameters

Execute

Define graphic and output filtering options

 **FDR p-val / FC thres.**

Select studied organism to add gene infos in output files

And execute !

LIMMA analysis

4 results in history :

- log file
- HTML page
- **Tabular LIMMAdetailed** For future developments, don't use it
- Tabular LIMMAstatistics

The screenshot shows the Galaxy history interface with the following details:

- Search bar: Rechercher des données
- Project: TestGalaxy_GSE46495
- Statistics: 41 shown, 9 deleted, 481.43 MB
- Actions: Checkmark, Eye icon (highlighted with a red box), Pen icon, X icon.
- Results:
 - 50: LIMMA FedVsFasted d_Log (Eye icon highlighted with a red box)
 - 49: LIMMA FedVsFasted d_HTML.html (Eye icon highlighted with a red box)
 - 48: LIMMA FedVsFasted d_LIMMAstatistics
 - 38: conditions.txt

Html page

LIMMA statistics (p.val, FC)

[LIMMA results](#)

P-val histograms

[Histogram Diet_fasted:Tissue_AdiposeTis](#)

[Histogram Diet_fasted:Tissue_Muscle](#)

[Histogram FedVsFasted_AllTissues](#)

[Histogram FedVsFasted_Liver](#)

Source of variation

[F-ratio barplot](#)

Volcanos

[Volcano Diet_fasted:Tissue_AdiposeTis](#)

[Volcano Diet_fasted:Tissue_Muscle](#)

[Volcano FedVsFasted_AllTissues](#)

[Volcano FedVsFasted_Liver](#)

LIMMA analysis

Can I continue my analysis ?

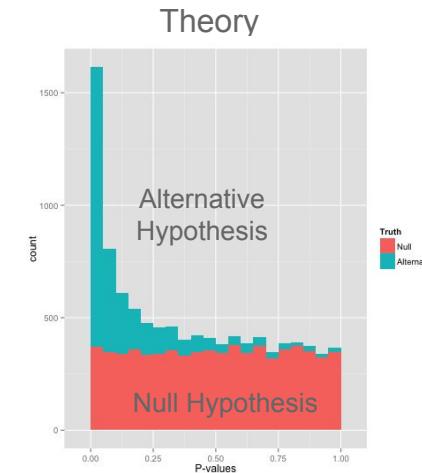
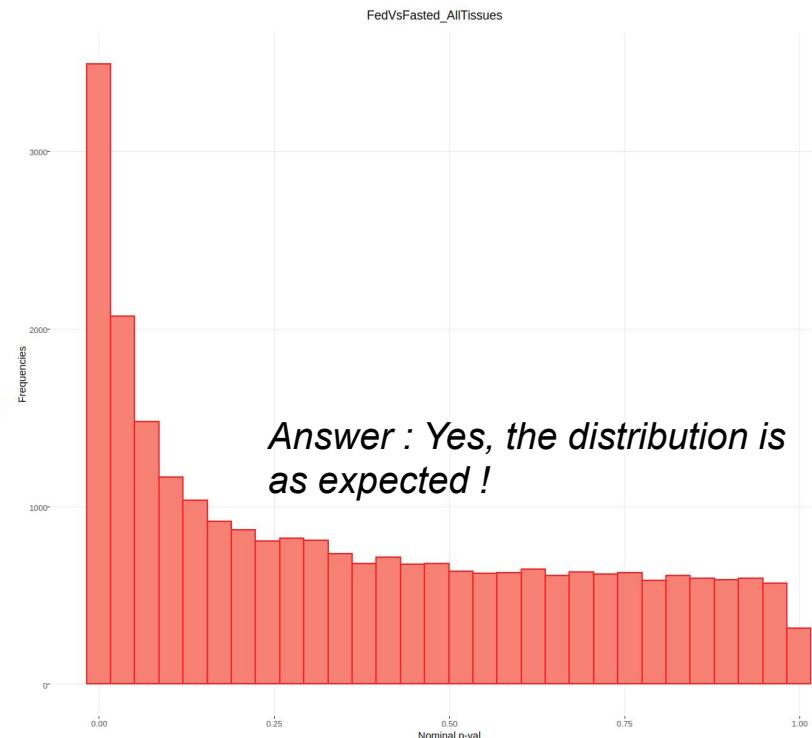
P-val histograms

[Histogram Diet_fasted:Tissue_AdiposeTis](#)

[Histogram Diet_fasted:Tissue_Muscle](#)

[Histogram FedVsFasted_AllTissues](#)

[Histogram FedVsFasted_Liver](#)



LIMMA analysis

Can I continue my analysis ?

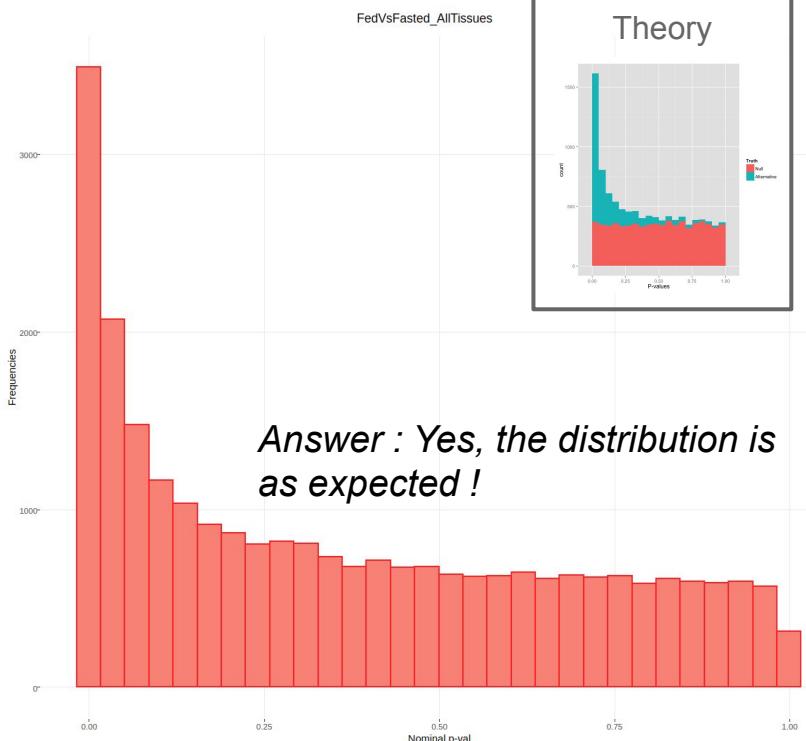
P-val histograms

[Histogram Diet_fasted:Tissue_AdiposeTis](#)

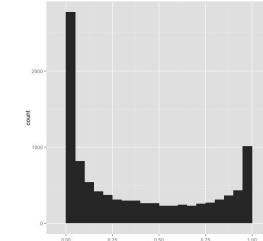
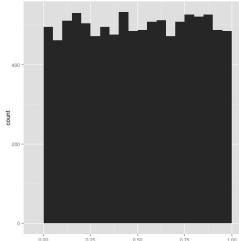
[Histogram Diet_fasted:Tissue_Muscle](#)

[Histogram FedVsFasted_AllTissues](#)

[Histogram FedVsFasted_Liver](#)

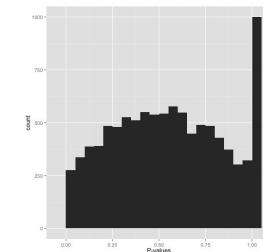
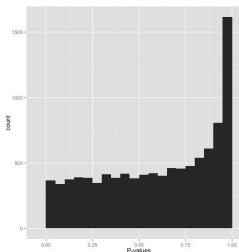


BAD examples & solutions... or not



*Small % hypothesis Not Null, apply FDR to find them

* DON'T :"Accept everything with p-values < 0.05"



*Something is wrong with your test (assumption on signal distribution ?)

* Find a statistician

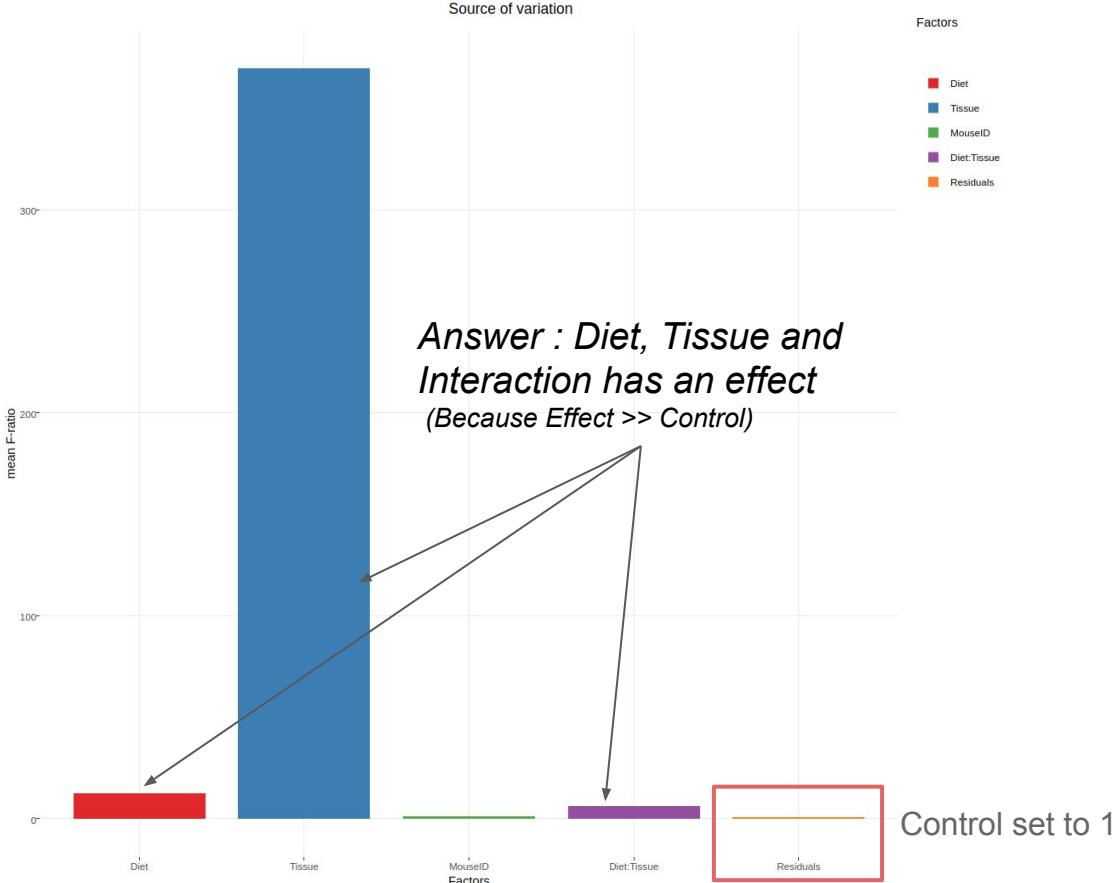
*“What the...?!?”

* Find a statistician

We expect a uniform distribution of p-value except for low p-values. If it's not the case, there is a problem with your dataset. <http://varianceexplained.org/statistics/interpreting-pvalue-histogram/>

LIMMA analysis

Which factor has an effect in my study ?



Inspect the graph and see if there is an “effect”. Generally, the effect here is retrieve in the number of differential expressed genes

LIMMA analysis

LIMMA statistics (p.val, FC)

LIMMA results

Fed vs Fasted inLiver												Fed vs Fasted Global			
Gene	Info	p-val	FDR.p-val	FC	log2(FC)	t-stat	p-val	FDR.p-val	FC	log2(FC)	t-stat				
0610007P14Rik	NA	0.1767	0.5372	1.258	0.3315	1.4	0.4394	0.72	1.116	0.1578	0.7889				
0610009B22Rik	RIKEN cDNA 0610009B22 gene	0.198	0.5611	0.77	-0.3771	-1.332	0.1894	0.481	1.253	0.3252	1.359				
	RIKEN cDNA 0610009L18 gene	0.5914	0.835	1.063	0.08881	0.5456	0.08914	0.3214	1.186	0.2458	1.787				
0610009O20Rik	NA	0.2534	0.6133	1.122	0.1656	1.176	0.08749	0.318	1.16	0.2138	1.797				
0610010B08Rik	NA	0.5333	0.8042	0.889	-0.1698	-0.634	0.172	0.456	1.249	0.3207	1.417				
	RIKEN cDNA 0610010F05 gene	0.9759	0.9926	1.003	0.004895	0.03063	0.4799	0.7483	1.07	0.09725	0.72				
0610010K14Rik	RIKEN cDNA 0610010K14 gene	0.2977	0.6523	0.8151	-0.295	-1.069	0.6805	0.8689	1.07	0.09743	0.4179				
	NA	0.1188	0.4544	1.242	0.3128	1.63	0.08007	0.302	1.23	0.2991	1.844				
0610012G03Rik	RIKEN cDNA 0610012G03 gene	0.6663	0.8732	1.043	0.06077	0.4376	0.003222	0.04594	1.313	0.3927	3.346				
	RIKEN cDNA 0610025J13 gene	0.1016	0.4249	1.248	0.3198	1.716	0.1001	0.3418	1.207	0.2716	1.724				
Gene	Info	p-val	FDR.p-val	FC	log2(FC)	t-stat	p-val	FDR.p-val	FC	log2(FC)	t-stat				

LIMMA analysis

Rechercher des données

TestGalaxy_GSE46495
41 shown, 9 deleted

481.43 MB

50: LIMMA_FedVsFasted_Log.

49: LIMMA_FedVsFasted_HTML.html

48: LIMMA_FedVsFasted_d_LIMMAstatistics eye

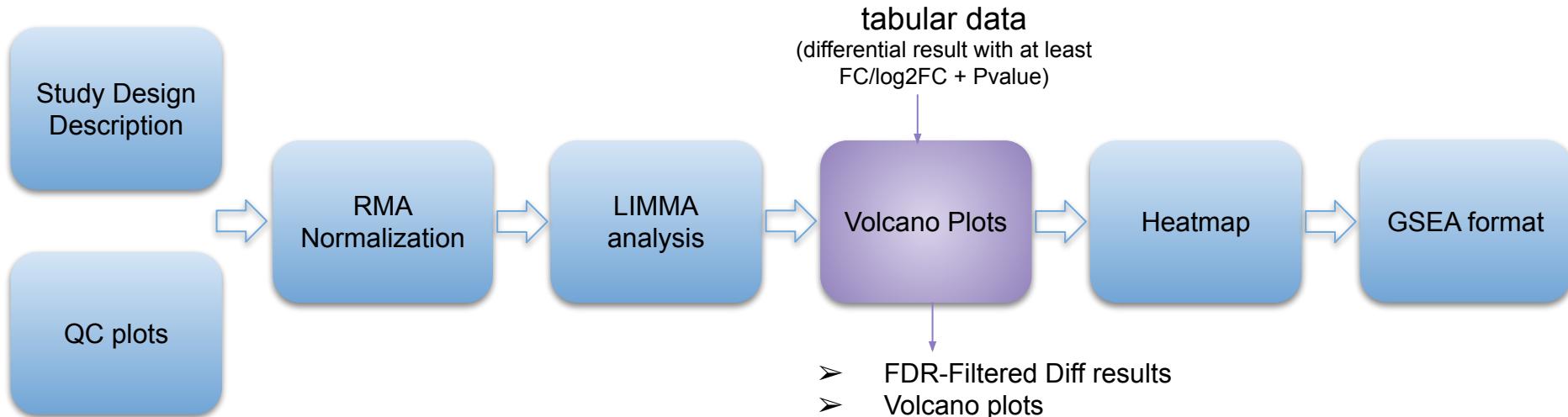
38: conditions.txt

1	2	3	4	5	6	7
LIMMA	Comparison	FastVSFed_inLiver	FastVSFed_inLiver	FastVSFed_inLiver	FastVSFed_inLiver	FastVSFed_inLiver
Gene	Info	p-val	FDR.p-val	FC	log2(FC)	t-stat
Gm15998	na	3.534e-24	9.091e-20	0.0371	-4.752	-33.71
Serpina7	na	3.088e-21	3.972e-17	0.04525	-4.466	-26.26
Pcsk9	na	1.444e-20	1.238e-16	0.09787	-3.353	-24.79
Hsd3b5	na	1.105e-19	7.108e-16	0.01318	-6.246	-22.97
Cdkn1a	na	1.671e-18	8.596e-15	26.95	4.752	20.73
Lss	na	1.607e-17	6.89e-14	0.1578	-2.663	-19.01
Hsd3b4	na	6.914e-17	2.307e-13	0.09852	-3.344	-17.97
Rgn	na	7.176e-17	2.307e-13	0.3125	-1.678	-17.94
Slc16a5	na	8.689e-17	2.483e-13	8.548	3.096	17.81
Mfsd2a	na	1.024e-16	2.634e-13	18.19	4.185	17.7
Zbtb16	na	1.163e-16	2.719e-13	8.234	3.042	17.61
Insig1	na	1.296e-16	2.779e-13	0.1683	-2.571	-17.53
Apex2	na	2.46e-16	4.868e-13	4.662	2.221	17.1
Usp2	na	2.				16.99
Acot1	na	3.				16.78
Rdh11	na	4.				-16.65
Sds1	na	5.				16.61
Cyp17a1	na	6.				16.45
Cyp4a31	na	1.				15.91
Plin5	na	2.				15.62
Asl	na	3.				15.49
Id3	na	4.				-15.26
Fkbp5	na	5.				15.15
Slc27a1	na	5.				15.17
Slc22a5	na	5.08e-15	5.477e-12	3.588	1.843	15.17
Gm15441	na	7e-15	6.925e-12	17.55	4.133	14.98
Gm38481	na	8e-15	7.622e-12	0.1291	-2.953	-14.9
Aqp8	na	8.654e-15	7.95e-12	0.1235	-3.017	-14.85
Hmgcr	na	1.042e-14	9.242e-12	0.2544	-1.975	-14.74
Keg1	na	1.199e-14	1.028e-11	0.1545	-2.694	-14.66
1810055G02Rik	na	1.239e-14	1.028e-11	4.194	2.068	14.64
Adgrf1	na	1.651e-14	1.282e-11	0.149	-2.747	-14.47
Mvk	na	1.695e-14	1.282e-11	0.2556	-1.968	-14.46
Cyp2c70	na	1.672e-14	1.282e-11	0.137	-2.867	-14.46
Nsdhl	na	1.815e-14	1.334e-11	0.1722	-2.538	-14.42
Gstm3	na	2.106e-14	1.435e-11	0.1146	-3.126	-14.33
Eif4ebp3	na	2.105e-14	1.435e-11	4.46	2.157	14.33
Nudt7	na	2.142e-14	1.435e-11	0.1682	-2.572	-14.32
Fdps	na	2.176e-14	1.435e-11	0.1778	-2.492	-14.31
Mvr	na	2.737e-14	1.717e-11	0.1435	-2.801	-14.18

Column
Useful only for GSEA
pre-ranking.
Don't use for results
interpretation

GIANT - Tools Suite

Microarray analysis pipeline: “ from .CEL to GSEA “



Volcano Plot

GIANT

[GIANT-QC Plots](#) Descriptive plots of .CEL collections or normalized expression data

[GIANT-Heatmap and Hierarchical clustering](#) Run hierarchical clustering and plot heatmap from expression data and/or differential expression analysis

[GIANT-Normalization with APT](#)
[Summarize](#) Apply Affymetrix Power Tool summarize function to .CEL collection

[GIANT-GSEA](#) Formating Format input files for GSEA software

[GIANT-Differential Expression with LIMMA](#) Use LIMMA to detect differentially expressed genes

[GIANT-Factor file generator](#)
Generate factor file used by other GIANT tools

[GIANT-Plot volcanos](#) Plot volcano from tabular file

GIANT-Plot volcanos Plot volcano from tabular file (Galaxy Version 0.2.0)

Input files

Title for output
Volcano_toPersonalize

Differential results file
35: LIMMA_toPersonalize_LIMMAstatistics

Select number of header lines in file
2

Volcano definition

Volcano
+ Insert Volcano

Output section

Select FC values kind selected as in
 log2(FC)
 FC

Info: log2(FC) will be displayed in volcano

Threshold FDR p-val threshold
0.05

Fold change threshold for volcanos
2.0

Add gene/probe information
Yes No

HTML snapshot format
 PNG format
 SVG format

Execute

1

2 Insert Contrast as much as you want to plot (1plot/contrast)

3 Pval, NOT FDR

4

5 Add Volcano as much as contrasts you want to plot **And execute !**

Choose a title

Select LIMMA results file OR Differential table (ex: RNA-seq)

= 2 if it's a LIMMA result

Choose a Plot Title

Select column of differential file containing p-val for contrast of interest

Pval, NOT FDR

Select column of differential file containing FC or log2(FC) for contrast of interest

Column selected for plot is FC or Log2(FC) ?

Threshold FDR for Volcano & for output table

Threshold FC, only for Volcano

COMPATIBILITY with ANY Differential Results (from GIANT or not, From microarrays or RNA-seq ...) in tabular format

Volcano Plot

3 results in history :

- log file
- Tabular file (with selected genes regarding FDR threshold)
- Html page

History

search datasets

Example_history
41 shown
366.04 MB

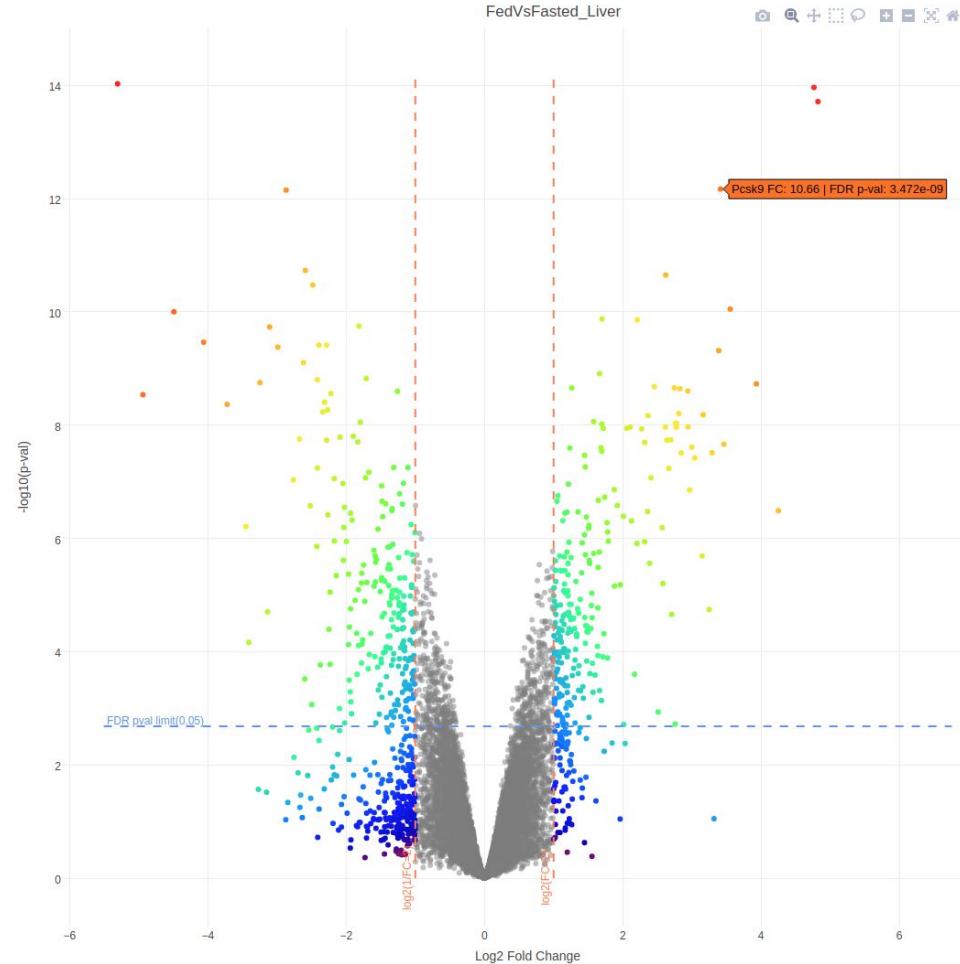
41: Volcano toPersonalize Log

40: Volcano toPersonalize HTML.html

39: Volcano toPersonalize LIMMAstatistics

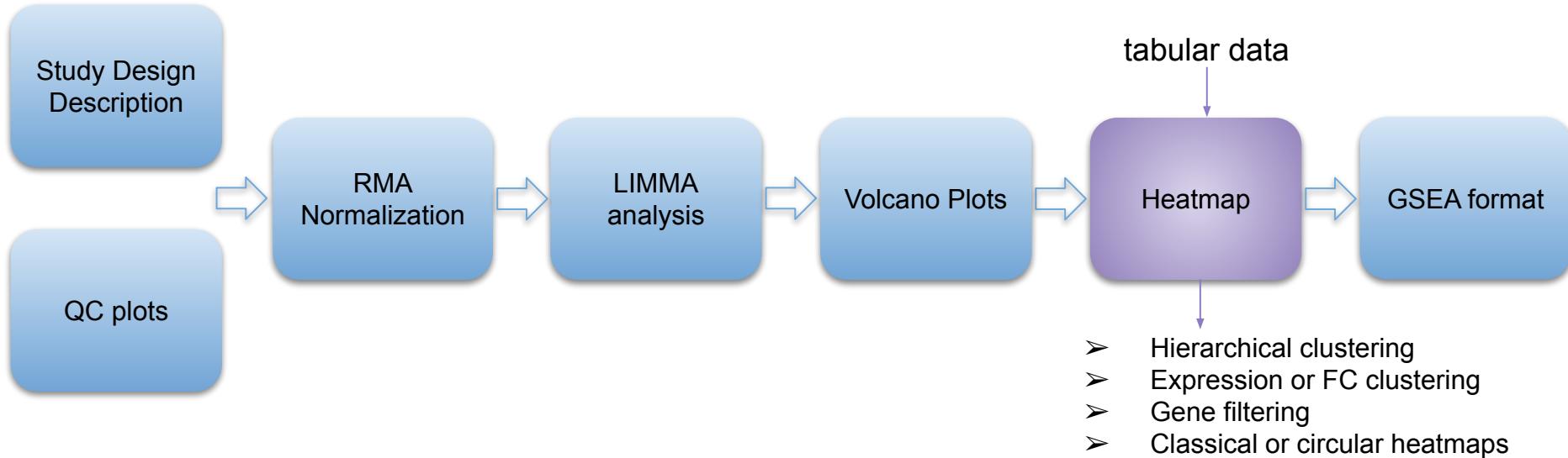
Volcanos

- [Volcano Diet_fasted:Tissue_AdiposeTis](#)
[Volcano Diet_fasted:Tissue_Muscle](#)
[Volcano FedVsFasted_AllTissues](#)
[Volcano FedVsFasted_Liver](#)



GIANT - Tools Suite

Microarray analysis pipeline: “ from .CEL to GSEA “



Heatmap

GIANT

[GIANT-QC Plots](#) Descriptive plots of .CEL collections or normalized expression data

[GIANT-Heatmap and Hierarchical clustering](#) Run hierarchical clustering and plot heatmap from expression data and/or differential expression analysis

[GIANT-Normalization with APT Summarize](#) Apply Affymetrix Power Tool summarize function to .CEL collection

[GIANT-GSEA Formating](#) Format input files for GSEA software

[GIANT-Differential Expression with LIMMA](#) Use LIMMA to detect differentially expressed genes

GIANT-Heatmap and Hierarchical clustering Run hierarchical clustering and plot heatmap from expression data and/or differential expression analysis (Galaxy Version 0.1) ▼ Options

Title for output
Heatmap_basedOnExpression

Data to cluster
Expression data

Normalized expression tabular file
 34: APT_GcSstRMA_NormalizedData

Probes/genes filtering
Filter input probes/genes before clustering

Filter
Based on differential expression results (FC and p-val)

Differential analysis tabular file (as given by LIMMA diff.exp. tool)
 48: LIMMA_FedVsFasted_LIMMAStatistics

Select comparisons to use for filtering
 Select/Unselect all
 FedVsFasted_Liver

Fold change threshold for input (both 'threshold' and '1/threshold' values will be used)
2.0
Minimum value is 1 (ie. all probes/genes are kept)

FDR p-val threshold for input
0.05
When several comparisons are selected a conservative rule is applied (see details below)

[Advanced parameters](#) ▼

Choose a title

Choose data to cluster, here expression data

Select optional filtering policy

Select filtering options corresponding to chosen policy (here contrast to use and thresholds on FC and FDR p-val)

And execute !

Heatmap

3 results in history :

- log file
- HTML page
- tabular results

Rechercher des données

TestGalaxy_GSE46495
44 shown, 12 deleted

493.9 MB

56: Heatmap_BasedOnExpression_Log	<input type="button" value="eye"/> <input type="button" value="edit"/> <input type="button" value="X"/>
55: Heatmap_BasedOnExpression_HTML.html	<input style="outline: 2px solid red;" type="button" value="eye"/> <input type="button" value="edit"/> <input type="button" value="X"/>
54: Heatmap_BasedOnExpression_ClusteringResults	<input type="button" value="eye"/> <input type="button" value="edit"/> <input type="button" value="X"/>
50: LIMMA_FedVsFastestLog	<input type="button" value="eye"/> <input type="button" value="edit"/> <input type="button" value="X"/>

Html page

Clustering tabular

[Clustering results](#)

Heatmap plot

[Heatmap](#)

Scree plot

[Scree plot](#)

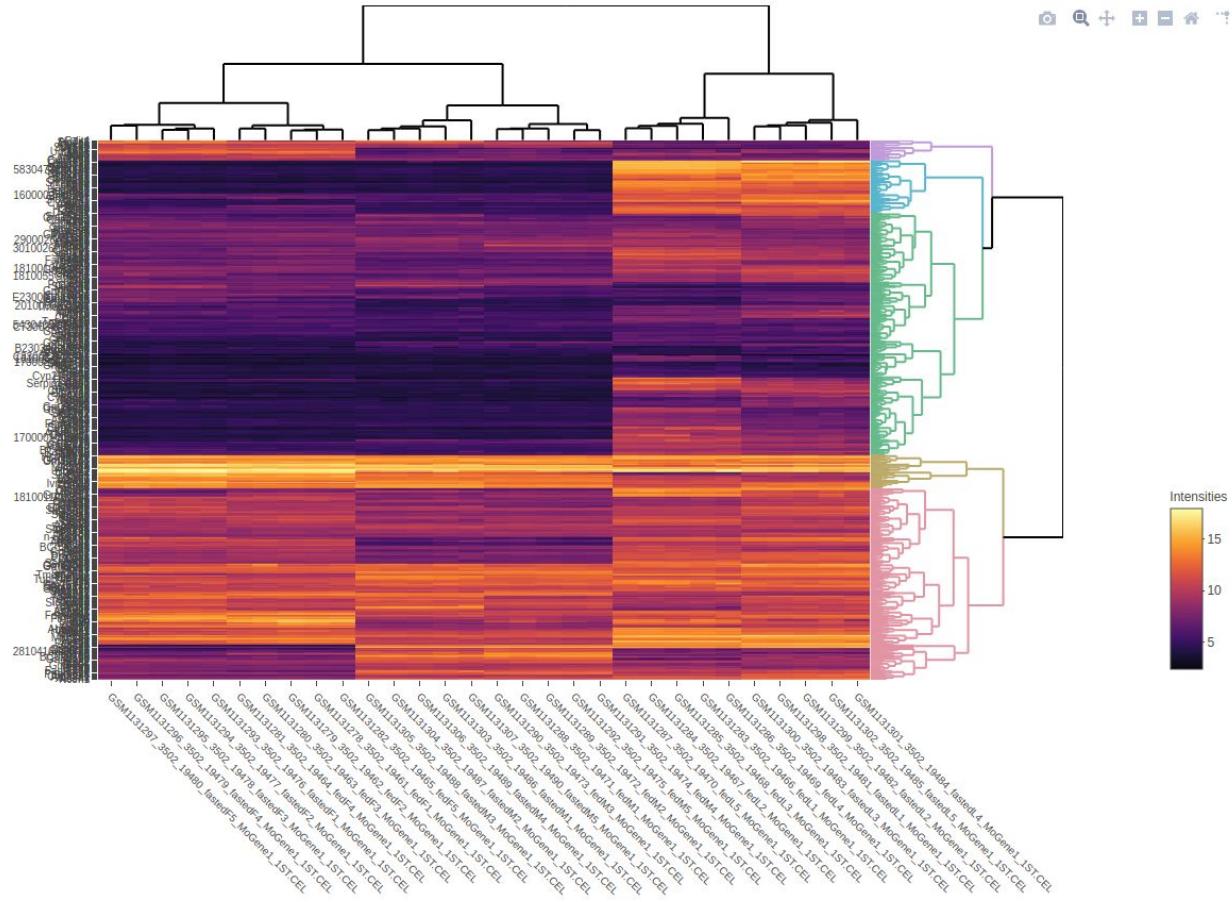
Circular plot

[Circular plot](#)

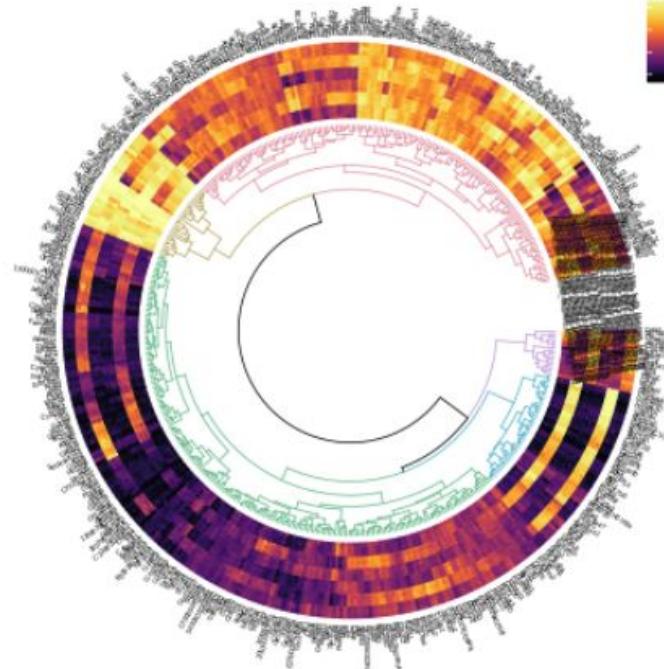
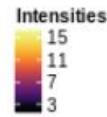
Heatmap

Heatmap plot

Heatmap



Heatmap



Circular plot

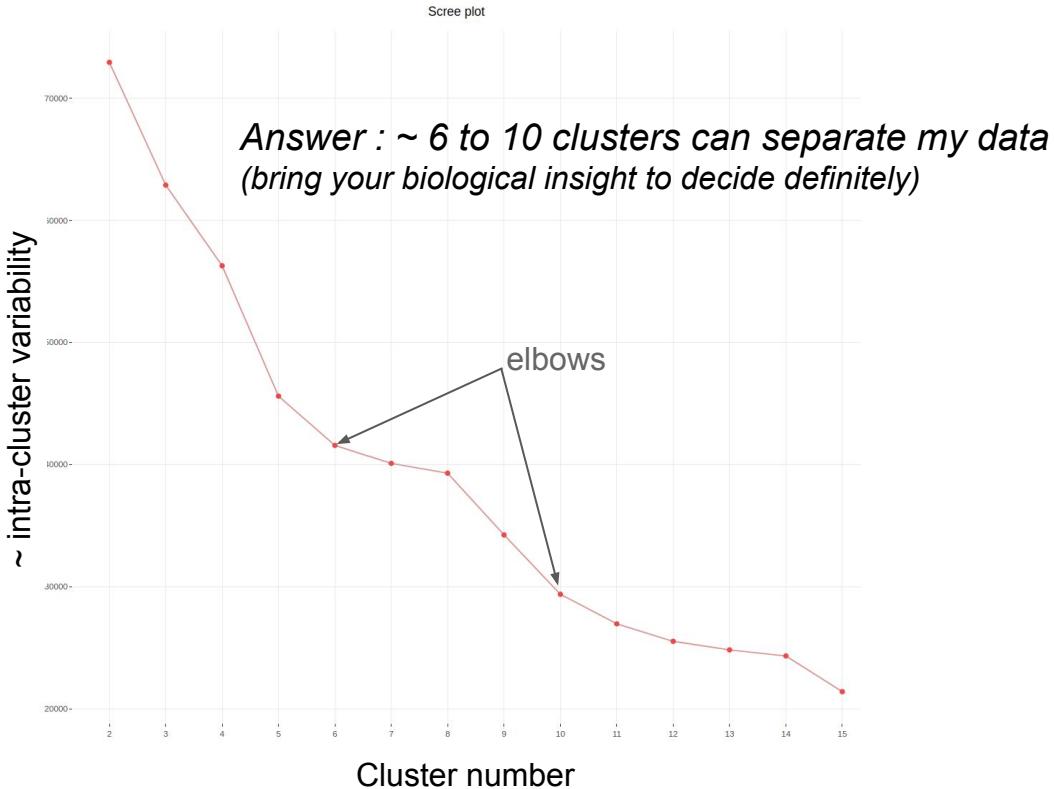
[Circular plot](#)

Heatmap

How many clusters are informative/useful ?

Scree plot

[Scree plot](#)



Heatmap

Clustering tabular

Clustering results

FedVsFasted_Liver						
Gene	Info	Cluster	p-val	FDR.p-val	FC	log2(FC)
Cdkn1a	cyclin-dependent kinase inhibi	4	9.414e-15	1.404e-10	0.02529	-5.306
Serpina7	serine (or cysteine) peptidase	1	1.092e-14	1.404e-10	27.18	4.765
Gm15998	predicted gene 15998	1	1.943e-14	1.666e-10	28.31	4.823
Hsd3b5	hydroxy-delta-5-steroid dehydr	1	8.1e-13	3.472e-09	94.14	6.557
Pcsk9	proprotein convertase subtilis	1	6.793e-13	3.472e-09	10.66	3.414
Zbtb16	zinc finger and BTB domain con	2	7.096e-13	3.472e-09	0.137	-2.868
Usp2	ubiquitin specific peptidase 2	4	1.855e-11	6.816e-08	0.1662	-2.589
Lss	lanosterol synthase	1	2.245e-11	7.219e-08	6.157	2.622
Per1	period circadian clock 1	4	3.353e-11	9.583e-08	0.1789	-2.483
Srebf1	sterol regulatory element bind	4	8.962e-11	2.305e-07	11.75	3.554

Showing 1 to 10 of 465 entries

Previous 1 2 3 4 5 ... 47 Next

Heatmap

GIANT

[GIANT-QC Plots](#) Descriptive plots of .CEL collections or normalized expression data

[GIANT-Heatmap and Hierarchical clustering](#) Run hierarchical clustering and plot heatmap from expression data and/or differential expression analysis

[GIANT-Normalization with APT Summarize](#) Apply Affymetrix Power Tool summarize function to .CEL collection

[GIANT-GSEA Formating](#) Format input files for GSEA software

[GIANT-Differential Expression with LIMMA](#) Use LIMMA to detect differentially expressed genes

GIANT-Heatmap and Hierarchical clustering Run hierarchical clustering and plot heatmap from expression data and/or differential expression analysis (Galaxy Version 0.1)

Title for output
Heatmap_BasedOnLimmaResults

Data to cluster
Differential expression analysis results

Differential analysis tabular file (as given by LIMMA diff.exp. tool)
48: LIMMA_FedVsFasted_LIMMAstatistics

Select comparisons to cluster
 Select/Unselect all
 FedVsFasted_Liver FedVsFasted_AllTissues

Probes/genes filtering
Filter probes/genes only in tabular output file

Filter
Based on diff. exp. parameters (FC and p-val)

Fold change threshold for output (both 'threshold' and '1/threshold' values will be used)
2.0
Minimum value is 1 (ie. all probes/genes are kept)

FDR p-val threshold for output
0.05
When several comparisons are selected a conservative rule is applied (see details below)

Advanced parameters

Requested cluster number
3

Use scree plot to adjust number of clusters

Cluster samples
 Yes No
To apply hierarchical clustering to samples

Maximum gene number to plot
1000

Personalized colors
 Yes No

Output format
 PNG format PDF format

Html snapshot format
 PNG format SVG format

Execute

Choose a title

Choose data to cluster, here contrast FC from differential analysis results

Select optional filtering policy and corresponding options

Define clustering options :

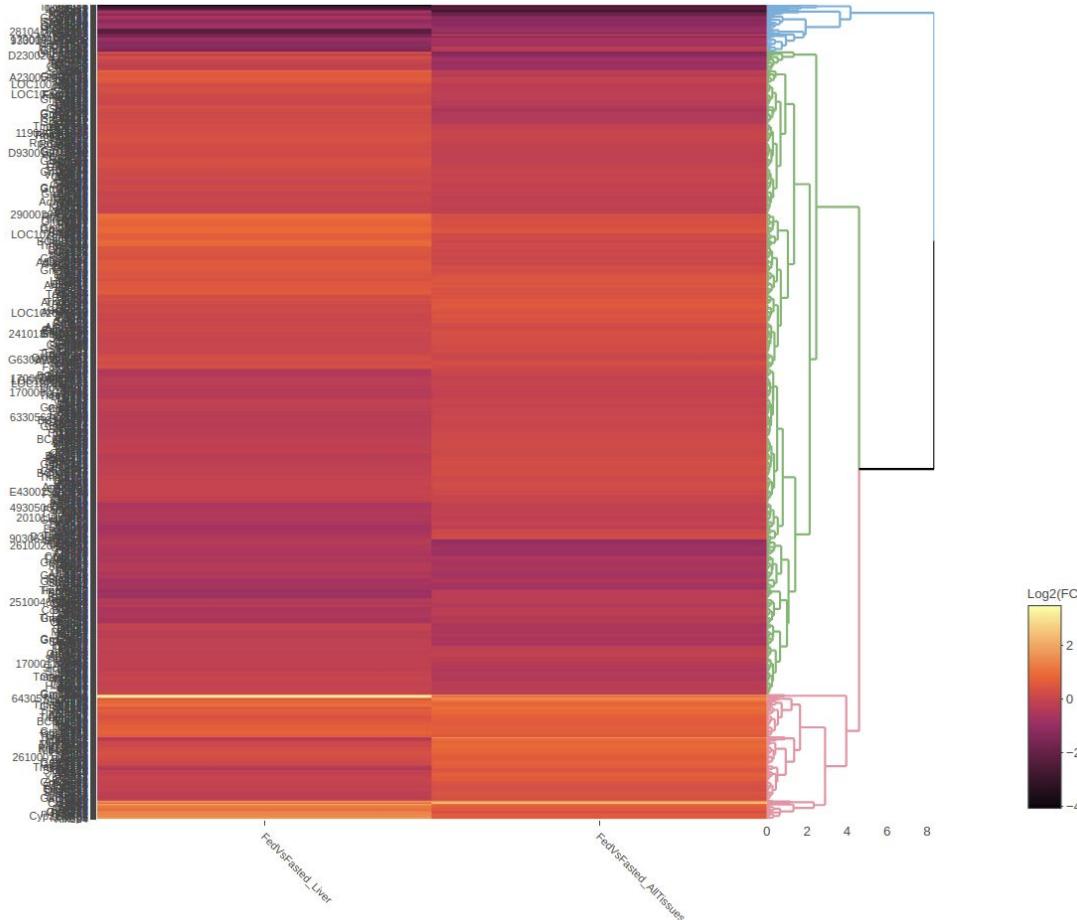
- cluster number
- cluster samples/contrasts
- color gradient

And execute !

Heatmap

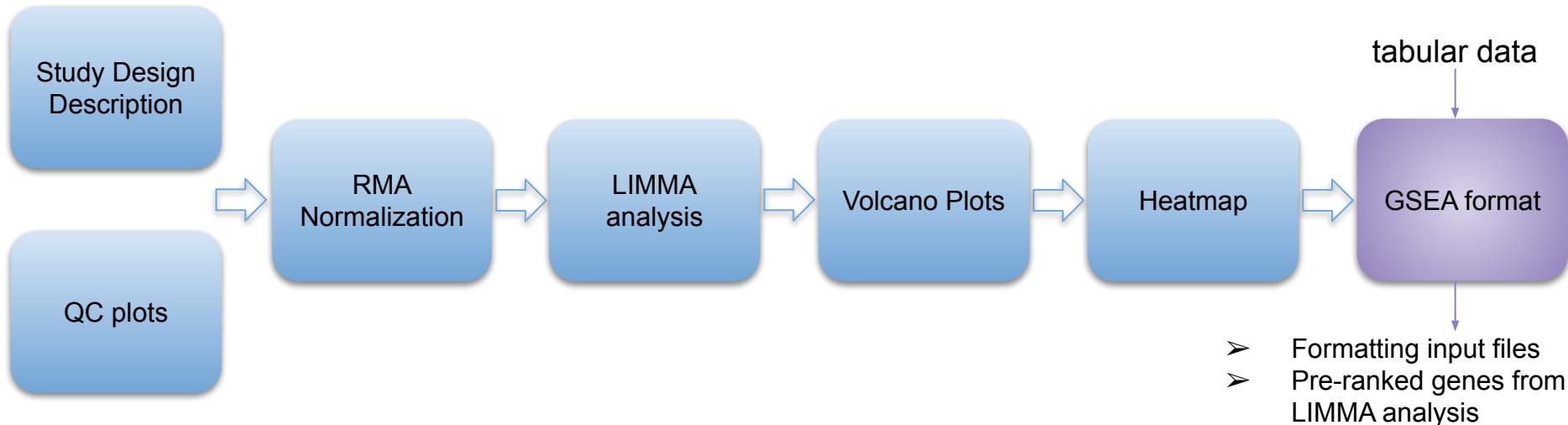
Heatmap plot

Heatmap



GIANT - Tools Suite

Microarray analysis pipeline: “ from .CEL to GSEA “



GSEA format

GIANT

[GIANT-QC Plots](#) Descriptive plots of .CEL collections or normalized expression data

[GIANT-Heatmap and Hierarchical clustering](#) Run hierarchical clustering and plot heatmap from expression data and/or differential expression analysis

[GIANT-Normalization with APT](#) Summarize Apply Affymetrix Power Tool summarize function to .CEL collection

[GIANT-GSEA Formating Format input files for GSEA software](#)

[GIANT-Differential Expression with LIMMA](#) Use LIMMA to detect differentially expressed genes

GIANT-GSEA Formating Format input files for GSEA software (Galaxy Version 0.1)

Title for output (without space)
GSEAformat_fromExpression

GSEA configuration
GSEA analysis

Normalized expression tabular file
34: APT_GcSstRNA_NormalizedData

Factor information tabular file
38: conditions.txt

Reference factor
Diet

Execute

Choose a title

For GSEA analysis, select expression and condition files with phenotype factor

And execute !

GSEA format

3 results in history :

- log file
- phenotype file
- expression file

The screenshot shows the TestGalaxy interface with a sidebar for searching data and a main area displaying historical results. The results are listed in a table with columns for name, description, and file type.

NAME	DESCRIPTION	FILE TYPE
TestGalaxy_GSE46495	50 shown, 12 deleted	
50.97 MB		
62: GSEAformat fromExpression_Log		
61: GSEAformat fromExpression Phenotypes		
60: GSEAformat fromExpression Expressions		
59: Heatmap BasedOnLimaResults_Log		

61: GSEAformat_fromExpression_Phenotypes.cls

```
30 2 1
# fed fasted
fed fasted fasted
```

60: GSEAformat_fromExpression_Phenotypes.gct

```
#1.2
25723 30
NAME DESCRIPTION GSM1131278_3502_19461_fedF1_MoGene1_1ST.CEL GSM1131279_3502_19462_fedF2_MoGene1_1ST.CEL GSM1131280_3502_19463_fedF3_MoGene1_1ST.CEL
Cep72 na 5.79442 6.13562 5.84982 5.98774 5.92123 5.33125 5.58054 5.55378 5.46471 5.23132 4.93863 4.99035 5.00348 4.70104 5.28328 5.8962
Unc93b1 na 10.344 10.3944 10.2048 10.4642 10.4454 9.47285 9.77977 9.77173 9.96121 9.7751 7.11397 7.33832 6.99454 7.05715 7.10572 10.4969
n-R5s217 na 4.32256 4.14826 4.35658 4.24367 4.56956 4.53336 4.57694 4.39047 4.36536 4.54664 4.44205 4.48468 4.49446 4.71065 4.57785
Gm28388 na 4.16515 4.16639 4.18478 3.98049 4.40679 4.33571 3.99504 4.39457 4.32379 4.24779 4.31598 4.78819 4.47686 4.65536 4.34895 3.96071
C920025E04Rik na 4.71314 5.56514 5.28329 5.36201 5.81648 7.3407 6.88435 7.23555 7.97076 7.0976 4.14262 4.55554 4.70368 4.45674 4.42379
Fdxr na 5.41385 5.48893 5.36192 5.31349 5.25819 7.04597 6.97918 6.95787 7.15202 7.05526 5.31033 5.05699 5.46236 5.04118 5.21375 5.22125
Cep76 na 6.91781 6.92847 6.72883 6.63956 6.92951 7.17886 7.13444 7.31411 6.83268 7.31083 8.05704 8.0441 8.10768 8.30292 7.8898 6.98429
Uba1 na 13.7473 13.8369 13.596 13.6206 13.723 13.6431 13.7014 13.7335 13.5518 13.5609 13.4516 13.3693 13.6276 13.3863 13.5045 13.3821
Cep78 na 8.03659 8.1785 8.2553 8.24717 7.8063 6.88295 6.66618 6.84899 6.24082 6.97469 7.97384 7.99195 7.81602 8.0196 8.16006 7.89462
Uba2 na 10.9947 10.9617 11.0736 10.6946 11.0001 9.95672 10.1506 10.2292 9.92522 10.1729 11.2999 11.2354 11.2891 11.2002 11.3825 10.3875
Spo11 na 3.7536 3.76609 3.76075 4.15384 3.8664 3.99895 4.07094 3.89099 4.39034 4.04493 3.86349 4.10587 4.16552 4.06177 4.21527 3.67195
Uba3 na 11.775 11.9832 11.9057 11.6576 11.8011 12.0993 11.9783 12.1653 11.337 12.1241 12.7865 12.5603 12.6863 12.545 12.81 11.1022
Dmpk na 12.4793 12.4088 12.3493 12.5491 12.4996 9.40599 8.93488 9.3373 9.48732 9.33554 14.362 14.3146 14.4238 14.2209 14.2243 9.76844
1700084C06Rik na 4.02575 4.19404 4.34464 4.08788 4.22004 3.94578 4.30068 4.2995 4.42673 4.75624 4.31217 4.17765 4.05961 4.46004 4.56209
Uba5 na 9.94948 9.95413 9.57235 9.80935 9.75292 9.94691 9.88339 10.2151 9.60875 10.1696 9.318 8.93312 9.21813 8.84039 9.1798 9.01907
Fubp1 na 11.7124 11.8233 11.9381 11.8558 11.7228 11.2408 11.2254 11.3883 10.1782 11.2232 10.9961 10.5842 10.8039 10.7685 10.7047 11.9859
```

GSEA format

GIANT

[GIANT-QC Plots](#) Descriptive plots of .CEL collections or normalized expression data

[GIANT-Heatmap and Hierarchical clustering](#) Run hierarchical clustering and plot heatmap from expression data and/or differential expression analysis

[GIANT-Normalization with APT](#)
[Summarize](#) Apply Affymetrix Power Tool summarize function to .CEL collection

[GIANT-GSEA Formating](#) Format input files for GSEA software

[GIANT-Differential Expression with LIMMA](#) Use LIMMA to detect differentially expressed genes

GIANT-GSEA Formating Format input files for GSEA software (Galaxy Version 0.2.0)

Versions Options

Title for output (without space)
GSEAformat_toPersonalize

GSEA configuration
Pre-ranked GSEA analysis

Differential analysis tabular file (as given by LIMMA diff.exp. tool)
39: Volcano_toPersonalize_LIMMAstatistics
This file should contain only annotated gene names or only probe identifiers as rows but no both kinds.

Reference contrast
Fasted versus Fed

Reference statistic
 Relative value of Log2(Fold Change)
 Absolute value of Log2(Fold Change)
 Relative value of moderated t-statistic
 Absolute value of moderated t-statistic

FDR p-val threshold
0.05

Execute

Recommended

Choose a title

For pre-ranked analysis, select diff. analysis results with contrast and sorting parameter to use

And execute !

This new option and recommendation is based on litterature & discussions with Biostatisticians, but can be adapted. Be AWARE of what you want to do with GSEA before selecting such option and adapt according your question and your expertise.