Galaxy Community Conference 2017: Visualization Workshop

Aysam Guerler
Johns Hopkins University

Saskia Hiltemann
ErasmusMC

Jeremy Goecks
Oregon Health & Science University
Recommended Web Browser

• Chrome will probably work best

• Updated Safari/Firefox should work well

• Internet Explorer and old versions of Safari/Firefox may have problems
Topics

• Why Visualize?
• What kind of Visualizations are available?
• Integration into a research pipeline.
Why Visualize?
Why Visualize?

Quick check: did it work?

Exploration and hypothesis generation

Sharing/publishing

*Understanding technical models and biological processes*
Anscombe’s Quartet

http://en.wikipedia.org/wiki/Anscombe's_quartet
Datasaurus

https://www.autodeskresearch.com/publications/samestats
Understanding t-SNE

http://distill.pub/2016/misread-tsne/
Why Visualize?

- Execute Tool
- Analyze Results
- Derive Conclusion
- Quality Control
- Biological Sample
Why Visualize?

More realistic:

- Biological Sample
- Quality Control
- Execute Tool
- Analyze Results
- Derive Conclusion

× Samples
× Measures
× Tools × Parameters
The challenge: Why Visualize?
Timeline of Visualizations in Galaxy

- 2005: 1st Galaxy paper published
- 2008: Display applications
- 2010: Visualization development started
- 2011: 1st visualization paper
- 2017: 40+ visualizations
Workshop Goals

Participants: learn about how to visualize your data in Galaxy
- numerical visualizations
- biological visualizations
- some analysis tools that do visualization well

Instructors: feedback from you about what you like, don’t like, and where to go next
The ‘right’ Visualization

Lots of different visualization out there, and many are very similar like e.g.:

   Genome browsers, scatterplots, and box plots

We’ll introduce you to some common and productive options, you choose what’s best for you.
Tools vs. Plugins

Tools can generate static plots and HTML pages with visualizations like e.g.:

FastQC, DeepTools and cummeRbund

Visualizations do not run as tools but can access data dynamically like e.g.:

Charts, Trackster and Circster
1st class Galaxy objects

Visualizations are first-class objects in Galaxy, just like tools

A visualization can be added to Galaxy via a configuration file that specifies:
✦ datasets that can be used
✦ location of visualization code (client-side or on server)

Galaxy handles visualization integration and data management
✦ users can focus on analyzing data
✦ developers can focus on creating visualizations
1st class Galaxy objects

Can be saved and versioned for reproducibility

Have a human-readable URL for sharing a fully interactive visualization:

http://usegalaxy.org/u/jgoecks/v/tumor-mutations

Can embed interactive visualizations in online supplementary materials via Galaxy Pages
Visual Analytics
Visual Analytics

http://www.visual-analytics.eu/faq/
Visualization Architecture

- Client-server architecture

- Lots of moving pieces
  - prepare/process data on server
  - send to client
  - render on client
Topics

- Why Visualize?
- What kind of Visualizations are available?
- Integration into a research pipeline.
Available Visualizations

- Examples of Visualization Plugins
- What is Galaxy Charts?
- Hands-on examples
Visualization Pathway
FastQC
Aims to provide a simple way to do some quality control checks on raw sequence data coming from high throughput sequencing pipelines.

Reference
www.bioinformatics.bbsrc.ac.uk
**cummeRbund**

A visualization package for Cufflinks sequencing data. It assists in the analysis of RNA-Seq differential expression analysis.

**Reference**

[http://compbio.mit.edu](http://compbio.mit.edu)
Circster
Renders genome-wide data. Position-based data is laid out in concentric circles representing chromosome positions.

Reference
Web-based visual analysis for high-throughput genomics, Goecks et al., BMC Genomics 2013, 14:397
Trackster
A genome browser with the ability to sample parameters and re-run tools on specific genome regions

Reference
NGS analyses by visualization with Trackster, Goecks et al., Nat. Biotech. 30(11), 1036-9
deepTools
A suite of user-friendly tools for the visualization, quality control and normalization of data from deep-sequencing experiments.

Reference
Ramírez et al. deepTools: a flexible platform for exploring deep-sequencing data. NAR, 2014
deepTools - Heatmaps
Heatmaps of gene expression and CpG methylation of gene bodies and flanking regions (±5 kbp) with differential methylation in neonatal and adult cardiomyocytes vs. stem cells.
Types of Visualizations

1. Only available on an external website?
   - Yes: Display Application
   - No: Computationally intensive?

2. Computationally intensive?
   - Yes: Regular Tool
   - No: Written in JavaScript?

3. Written in JavaScript?
   - Yes: Charts Plugin
   - No: Generic Plugin
A shared interface for a range of visualization types i.e. bar diagrams, pie charts, scatter plots, heat maps and others.

Reference

www.nvd3.com
www.jqplot.com
Configuration Forms
An interface for a range of common visualization types i.e. bar diagrams, pie charts, scatter plots, heat maps and others.

Reference
www.nvd3.com
www.jqplot.com
Custom Plugins

If a plugin is unavailable, custom visualization types like this heat map here can be added. Charts is able to pre-process large-scale data behind the scenes.

Showing Protein-Protein Interaction data from:

www.compsysbio.org/bacteriome/
Example of a PDB-Viewer
...and many others
Community Driven

More than 28 visualizations available in Charts alone!

Data-Driven Documents
jqPlot
BioJS
Cytoscape
NVD3

Re-usable charts for d3.js
Topics

• Why Visualize?
• What kind of Visualizations are available?
• Integration into a research pipeline.
Quality Control for Raw Sequence Data
Motivating Questions

What are the characteristics of my sequencing data?

Are there any important issues to understand and/or address?

Does it meet my analysis minimum requirements?
Inputs

Import FASTQ datasets from Data Library:

Data Libraries -> Training Data -> Visualisation Workshop

Used throughout slides to show actions to take
Inputs

Import FASTQ datasets from Data Library:

Data Libraries -> Training Data -> Visualisation Workshop
Inputs
Select files and import to history

- adrenal1
- adrenal2
Run FastQC tool

1. Search for FastQC tool in the Galaxy tool menu.
2. Select the dataset(s) for FastQC analysis.
3. Review the FastQC Read Quality reports.
4. Execute the FastQC tool.
GC distribution over all sequences

GC count per read
Theoretical Distribution

Mean GC content (%)
Per base sequence content

Sequence content across all bases

Position in read (bp)
FastQC Takeaways

Great for first look at raw sequence data

Several nice plots to view data

Investigate but be skeptical of automated pass/fail
- FastQC designed for whole genome sequencing and may not be applicable for your data
- e.g., RNA-seq data often fails several checks due to unique characteristics that are different than WGS
Visualizing Primary Genomic Data
Analysis goal: what similarities and differences can be found in cancer cell lines using exome and transcriptome sequencing?
Sequencing and Analysis

Sequenced exomes and transcriptomes of 3 pancreatic cancer cell lines
- MiaPaCa2, HPAC, and PANC-1

Datasets available in data library:
- Exome subset: KRAS, STK11, ERBB2 aligned reads, removed dups, created read pileup
- transcriptome subset: KRAS, STK11, ERBB2 aligned reads
- whole transcriptome aligned reads coverage
- gene annotation
# Import Data

<table>
<thead>
<tr>
<th>name</th>
<th>description</th>
<th>data type</th>
<th>size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell line variants</td>
<td></td>
<td>vcf</td>
<td>110.1 KB</td>
</tr>
<tr>
<td>Gene fusions</td>
<td></td>
<td>chrint</td>
<td>227 bytes</td>
</tr>
<tr>
<td>HER_Rep1_ERCC-Mix2_Build37-ErccTranscripts-chr22.read1.fastq</td>
<td></td>
<td>fastq</td>
<td>2.9 MB</td>
</tr>
<tr>
<td>HER_Rep1_ERCC-Mix2_Build37-ErccTranscripts-chr22.read2.fastq</td>
<td></td>
<td>fastq</td>
<td>2.9 MB</td>
</tr>
<tr>
<td>HPAC Assembled Transcripts</td>
<td></td>
<td>gtf</td>
<td>18.9 KB</td>
</tr>
<tr>
<td>HPAC Transcriptome Coverage (Genome)</td>
<td></td>
<td>bigwig</td>
<td>121.9 MB</td>
</tr>
<tr>
<td>HPAC Transcriptome (KRAS, STK11, ERBB2)</td>
<td></td>
<td>bam</td>
<td>1.5 MB</td>
</tr>
<tr>
<td>MiPaCa2 Assembled Transcripts</td>
<td></td>
<td>gtf</td>
<td>14.9 KB</td>
</tr>
<tr>
<td>MiPaCa2 Transcriptome Coverage (Genome)</td>
<td></td>
<td>bigwig</td>
<td>148.7 MB</td>
</tr>
<tr>
<td>PANIC1 Assembled Transcripts</td>
<td></td>
<td>gtf</td>
<td>14.4 KB</td>
</tr>
<tr>
<td>PANIC1 Transcriptome Coverage (Genome)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UCSC Main on Human: refFlat (genome)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. Select datasets to import into history.
2. Click on "Import selected datasets into history."
Import Data

Go back to data library and import remaining datasets on second page
Display Applications

1: MiaPaCa2 Exome (KRA S, STK11, ERBB2)
297.5 MB
format: bam, database: hg19
uploaded bam file

display at UCSC main
display at Ensembl Current
display with IGV local Human hg19
display in IGB View

Binary bam alignments file
Display Applications

1. MiaPaCa2 Exome (KRA S, STK11, ERBB2)
297.5 MB
format: bam, database: hg19

uploaded bam file

display at UCSC main
display at Ensembl Current
display with IGV local Human hg19
display in IGB View

Binary bam alignments file
Display Applications
Display Applications

Advantages
- use familiar tools
- Often easy to view your data alongside public datasets

Disadvantages
- cannot easily share/version visualization
- many more visualizations than display applications in Galaxy
Trackster—Galaxy’s Genome Browser
Trackster—Galaxy’s Genome Browser

Genome browsers are a foundational genome visualization tool

Trackster is for the high-throughput sequencing era
- no need to download/move/copy datasets
- maximum flexibility for customization (e.g. rainbow tracks)
- 2-3 indices per dataset for fast visualization

Datasets supported
- intervals: BED, BigBed, GFF/GTF, interval
- reads and variants: SAM/BAM, VCF
- continuous/numerical: Wiggle, BigWig, BedGraph
Let’s visualize our data in Trackster

1. Create visualization
2. Add gene annotation (RefSeq)
3. Save visualization
4. Exit
5. Reopen visualization
Let’s visualize our data in Trackster

1. Create visualization
Let’s visualize our data in Trackster

1. Create visualization
Let’s visualize our data in Trackster

1. Create visualization
2. Add gene annotation (RefSeq)
Let’s visualize our data in Trackster

1. Create visualization
2. Add gene annotation (RefSeq)
Let’s visualize our data in Trackster

1. Create visualization
2. Add gene annotation (RefSeq)
3. Save visualization
Let’s visualize our data in Trackster

4. Exit
Let’s visualize our data in Trackster

5. Reopen visualization
Behind the Scenes

Galaxy is indexing datasets for

✦ viewing large genomic regions (coverage plots)
✦ viewing small genomic regions (getting individual data points)
✦ feature names and locations

Indexes is the primary way that big datasets are visualized quickly
Display Modes

Tracks can be displayed differently
- coverage to individual features
- similar language to UCSC

Let’s try different modes
- this is fast because data is sent from Galaxy server and rendered in your Web browser
Searching

Can search for named features such as gene annotations
✦ BED, GFF/GTF

Let’s try searching for a gene: ERBB2
Let’s add data to Trackster

Add exome data for all cell lines and called variants...

1. Select datasets for new tracks
2. Add selected datasets
3. Confirm selection and add data
Let’s add data to Trackster

Add exome data for all cell lines and called variants...

but all these reads make it hard to see the data below

so change exome mode to coverage
Variant tracks
Trackster: Rainbow Track for Coverage

1. Navigate to ERBB2 gene
2. Create group
3. Add transcriptome coverage tracks to group
4. Create composite track
5. Adjust max
6. what do we see?
Trackster: Rainbow Track for Coverage

1. Remove gene fusions track
Trackster: Rainbow Track for Coverage

2. Navigate to ERBB2 gene
Trackster: Rainbow Track for Coverage

3. Create group
Trackster: Rainbow Track for Coverage

4. Add transcriptome coverage tracks to group
Trackster: Rainbow Track for Coverage

5. Create composite track
Trackster: Rainbow Track for Coverage

6. Adjust display; Adjust max and name; change height
Add More Data

Add RNA-seq mapped reads and assembled transcripts

Look at ERBB2
  ✦ bookmark

Look at STK11
  ✦ bookmark

Look at KRAS
  ✦ bookmark
Interested in genes with mutations that are also expressed

Let’s look at KRAS

Exon 2
Share and Publish

Saved Visualizations

<table>
<thead>
<tr>
<th>Title</th>
<th>Type</th>
<th>Dbkey</th>
<th>Tags</th>
<th>Sharing</th>
<th>Created</th>
<th>Last Updated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Workshop</td>
<td>Trackster</td>
<td>hg19</td>
<td>0 Tags</td>
<td>~6 hours ago</td>
<td>~2 seconds ago</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Trackster</td>
<td>hg19</td>
<td>0 Tags</td>
<td>Jun 18, 2015</td>
<td>Jun 18, 2015</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Trackster</td>
<td>hg19</td>
<td>0 Tags</td>
<td>Mar 31, 2015</td>
<td>Mar 31, 2015</td>
<td></td>
</tr>
</tbody>
</table>

Visualizations shared with you by others

No visualizations have been shared with you.
Share or Publish Visualization 'Workshop'

Make Visualization Accessible via Link and Publish It
This visualization is currently accessible via link.
Anyone can view and import this visualization by visiting the following URL:

http://localhost:8080/u/jgoecks/v/workshop

You can:
- Disable Access to Visualization Link
  Disables visualization's link so that it is not accessible.
- Publish Visualization
  Publishes the visualization to Galaxy's Published Visualizations section, where it is publicly listed and searchable.

Share Visualization with Individual Users
You have not shared this visualization with any users.
- Share with a user

Back to Visualizations List
Share and Publish
DeepTools for Understanding Numerical Genomic Data

Nucleic Acids Research

deepTools: a flexible platform for exploring deep-sequencing data

Fidel Ramírez1, Friederike Dündar1,2, Sarah Diehl1, Björn A. Grüning2 and Thomas Manke1

*Author Affiliations

To whom correspondence should be addressed. Tel: +49 0 761 5108 738; Fax: +49 0 761 5108 80738; Email: manke@em.uni-freiburg.de

Received February 4, 2014. Revision received April 5, 2014. Accepted April 15, 2014.

http://nar.oxfordjournals.org/content/42/W1/W187
Motivating Questions

• What are the characteristics of my aligned sequence reads?

• Coverage especially important
  ✦ average across all regions
  ✦ coverage per gene

• Are there correlations between samples?
plotCoverage for exome sequencing

plotCoverage assesses the sequencing depth of BAM files (Galaxy Version 2.2.3.0)

**Bam File**

- 6: HPAC Transcriptome (K8AS, STK11, ERBB2)
- 5: PAM1 Transcriptome (K8AS, STK11, ERBB2)
- 4: MiaPaCa2 Transcriptome (K8AS, STK11, ERBB2)
- 3: HPAC Exome (K8AS, STK11, ERBB2)
- 2: PAM1 Exome (K8AS, STK11, ERBB2)

Show advanced options

- Yes

**Number of samples**

- 100000

Number of samples taken from the genome to compute the scaling factors. (---numberOfSamples)

**Region of the genome to limit the operation to**

This is useful when testing parameters to reduce the time required. The format is chr:start:end, for example "chr10" or "chr10:456700:891000" (---region)

**Extend reads to the given average fragment size.**

- No extension. The default value and most typically appropriate.

(1) Single-end reads and singletons are extended to match this length. (2) Paired-end reads are extended to match the fragment size, regardless of what is set here. By default, "each" read mate is extended. This can be modified using the SAM flags (see: ---samFlagInclude and ---samFlagExclude options) to keep only the first or the second mate. Unmapped reads, mate reads that map to different chromosomes or too far apart are extended to the given value. Reads are only extended if --extendReads is set to a value greater than the read length. *NOTE*: For spliced-read data, this option is not recommended as it will extend reads over skipped regions, e.g., introns in RNA-seq data.

**Ignore duplicates**

- Yes

If set, reads that have the same orientation and start position will be considered only once. If reads are paired, the mate position also has to coincide to ignore a read. (---ignoreDuplicates)

**Center regions with respect to the fragment length**

- Yes

For paired-end data the fragment is defined by the bounds of the reads. For single-end data the bounds are defined by the read and the user-definable fragment/extension length. This option is useful to get a sharper signal around enriched regions. (---centerReads)

**Minimum mapping quality**

- 1

If set, only reads with a mapping quality score higher than this value are considered. (---minMappingQuality)

**Include reads based on the SAM flag**

For example, to get only reads that are the first mate use a flag of 64. This is useful to count properly paired reads only once, otherwise the second mate will be also considered for the coverage. (---samFlagInclude)

**Exclude reads based on the SAM flag**

For example, to get only reads that map to the forward strand, use ---samFlagExclude 16, where 16 is the SAM flag for reads that map to the reverse strand. (---samFlagExclude)

**Skip zeros**

- Yes

If set, then zero counts that happen for all BAM files given are ignored. This may result in fewer considered regions. (---skipZeros)

**Title of the plot**

Title of the plot, to be printed on top of the generated image. (---plotTitle)
plotCoverage for exome sequencing
Data Preparation for Transcription

Datasets: Compute Matrix

computeMatrix prepares data for plotting a heatmap or a profile of given regions (Galaxy Version 2.2.3.0)

Select regions

1: Select regions

Regions to plot

- File, in BED format, containing the regions to plot.

Score file

- 13: UCSC Main on Human: refFlat (genome)

computeMatrix has two main output options

scale-regions

In the scale-regions mode, all regions in the BED file are stretched or shrunk to the same length (in bases) that is indicated by the user. Reference-point refers to a position within the BED regions (start or end of each region). In the reference-point mode only those genomic positions before (upstream) and/or after (downstream) the reference point will be considered.

Distance in bases to which all regions are going to be fit

- 500

Set distance up- and downstream of the given regions

- no

Show advanced output settings

- no

Show advanced options

- no

Job Resource Parameters

- Use default job resource parameters

Execute
plotProfile

46: plotProfile image

40.5 KB
format: png, database: hg19
Image in png format

MiaPaCa2

PANC1

HPAC

genes

TSS  TSS  TSS
TES  TES  TES

10  15  20  25  30  35  40
plotHeatmap

<table>
<thead>
<tr>
<th>Matrix file from the computeMatrix tool</th>
</tr>
</thead>
<tbody>
<tr>
<td>(matrixFile)</td>
</tr>
<tr>
<td>Show advanced output settings</td>
</tr>
<tr>
<td>yes</td>
</tr>
<tr>
<td>Image file format</td>
</tr>
</tbody>
</table>

The x-axis label:
- distance from TSS (top)
- (+-xAxisLabel)

The y-axis label for the top panel:
- genes
- (-yAxisLabel)

Heatmap width in cm:
- 5
- The minimum value is 1 and the maximum is 100. (--heatmapWidth)

Heatmap height in cm:
- 20
- The minimum value is 1 and the maximum is 100. (--heatmapHeight)

What to show:
- summary plot, heatmap and colorbar
- The default is to include a summary or profile plot on top of the heatmap and a heatmap colorbar. (--whatToShow)

Label for the region start:
- TSS
- Only for scale-regions mode. Label shown in the plot for the start of the region. Default is TSS (transcription start site), but could be changed to anything, e.g. "peak start". (%startLabel)

Label for the region end:
- TES
- Only for scale-regions mode. Label shown in the plot for the region end. Default is TES (transcription end site). (%endLabel)

Reference point label:
- TSS
- Label shown in the plot for the reference-point. Default is the same as the reference point selected (e.g. TSS), but could be anything, e.g. "peak start" etc. (%referencePointLabel)

Labels for the samples (each bigwig) plotted:
- MiaPaCa2 PANC1 HPAC
- The default is to use the file name of the sample. The sample labels should be separated by spaces and quoted if a label itself contains a space e.g. label=1 "label 2" (%samplesLabel)
Data Preparation: multiBigwigSummary

multiBigwigSummary calculates average scores for a list of two or more bigwig files (Galaxy Version 2.2.3.0)

Choose computation mode

Bins
In the bins mode, the correlation is computed using equally sized bins. In the BED file mode, a list of genomic regions in BED format has to be given. For each region in the BED file, the number of overlapping reads is counted in each of the BigWig files. Then the correlation is computed.

- Bin size in bp
  - 10000
  - Length in bases for a window used to sample the genome. (--binSize)

- Distance between bins
  - 0
  - By default, multiBamSummary considers consecutive bins of the specified 'Bin size'. However, to reduce the computation time, a larger distance between bins can be given. Larger distances result in fewer bins being considered. (--distanceBetweenBins)

Region of the genome to limit the operation to

- This is useful when testing parameters to reduce the time required. The format is chr:start:end, for example "chr10" or "chr10:456700-891000". (--region)

Save raw counts (scores) to file

- Yes
- No
- (--outRawCounts)

Job Resource Parameters

- Use default job resource parameters

Execute
plotCorrelation

**plotCorrelation** Create a heatmap or scatterplot of correlation scores between different samples (Galaxy Version 2.2.3.0)

Matrix file from the multiBamSummary tool

- 49: multibamwigwsummary on data 9, data 8, and data 7: correlation matrix

Correlation method

- Spearman

Plotting type

- Heatmap

Minimum value for the heatmap intensities

If not specified the value is set automatically. (--min)

Maximum value for the heatmap intensities

If not specified the value is set automatically. (--max)

Color map to use for the heatmap

- RdYBu

Available color map names can be found here: http://matplotlib.org/examples/color/colormaps_reference.html

Title of the plot

Title of the plot, to be printed on top of the generated image. (--plotTitle)

Plot the correlation value

- Yes

If set, then the correlation number is plotted on top of the heatmap. (--plotNumbers)

Skip zeros

- Yes

If set, then zero counts that happen for *all* BAM files given are ignored. This may result in fewer considered regions. (--skipZeros)

Image file format

- png

(--outFileFormat)

Remove regions with very large counts

- Yes

If set, bins with very large counts are removed. Bins with abnormally high reads counts artificially increase Pearson correlation; that's why, by default, plotCorrelation tries to remove outliers using the median absolute deviation (MAD) method applying a threshold of 200 to only consider extremely large deviations from the median. ENCODE blacklist page (https://sites.google.com/site/arsenalunddeu/projects/blacklists) contains useful information about regions with unusually high counts. (--removeOutliers)

Save the matrix of values underlying the heatmap

- Yes

Execute
*Warning*. 512385 NaN values were found. They will be removed along with the corresponding bins in other samples for the computation and plotting.
RNA-seq

Sequence RNA molecules to characterize:
- transcribed genes
- individual transcripts
- differential expression
Motivating Question

Tutorial: https://galaxyproject.org/tutorials/rb_rnaseq/

- Tophat $\rightarrow$ Htseq-count $\rightarrow$ DESeq2

How does using HISAT instead of Tophat impact the results?
DESeq2 determines differentially expressed features from count tables (Galaxy Version 2.1.8.3)

**Factor**

1: Factor

**Specify a factor name**

Conditions

Only letters, numbers and underscores will be retained in this field

**Factor level**

2: Factor level

**Specify a factor level**

Condition 1

Only letters, numbers and underscores will be retained in this field

Counts file(s)

- [ ] 84: htseq-count on collection 37

2: Factor level

**Specify a factor level**

Condition 2

Only letters, numbers and underscores will be retained in this field

Counts file(s)

- [ ] 92: htseq-count on collection 57

**Visualising the analysis results**

- [ ] Yes
- [ ] No

Output all levels vs all levels of primary factor (use when you have > 2 levels for primary factor)

- [ ] Yes
- [ ] No

DESeq2 performs independent filtering by default using the mean of normalized counts as a filter statistic.

**Fit type**

- [ ] parametric

**Turn off outliers replacement (only affects with >6 replicates)**

- [ ] Yes
- [ ] No

When there are more than 6 replicates for a given sample, the DESeq2 will automatically replace counts with large Cook’s distance with the trimmed mean over all samples, scaled up by the size factor or normalization factor for that sample.

**Turn off outliers filtering (only affects with >2 replicates)**

- [ ] Yes
- [ ] No

When there are more than 2 replicates for a given sample, the DESeq2 will automatically filter genes which contain a Cook’s distance above a cutoff.

**Turn off independent filtering**

- [ ] Yes
- [ ] No

DESeq2 performs independent filtering by default using the mean of normalized counts as a filter statistic.

[ ] Execute
MA-plot for Conditions: ConditionX1 vs ConditionX2

MA-plot for Conditions: Condition1 vs Condition2
Histogram of p-values for Conditions: ConditionX1 vs ConditionX2

Histogram of p-values for Condition: c1 vs c2
Workshop Materials

Will be available on training day page

For this workshop:
- Galaxy page on usegalaxy.org with:
  - published history
  - published visualization

https://usegalaxy.org/u/jeremy/p/visualization-workshop
(not updated yet to for this year’s materials)
Thanks!