



University of
Zurich^{UZH}

URPP Systems Biology / Functional Genomics



Swiss Institute of
Bioinformatics

Galaxy and edgeR

Mark D. Robinson
University of Zurich

My work

- Research:
 - 50%: statistical method development for genomics
 - 35%: applying statistical methods for/with collaborators
 - 15%: tool development (Bioconductor)
- Software:
 - 85% R/Bioconductor
 - 5% browsers – UCSC/IGV
 - 5% command line
 - 5% other – perl/python/Galaxy
- Galaxy: **Interfaces to R/Bioconductor**



Method

Open Access

Bioconductor: open software development for computational biology and bioinformatics

Robert C Gentleman¹, Vincent J Carey², Douglas M Bates³, Ben Bolstad⁴, Marcel Dettling⁵, Sandrine Dudoit⁴, Byron Ellis⁶, Laurent Gautier⁷, Yongchao Ge⁸, Jeff Gentry¹, Kurt Hornik⁹, Torsten Hothorn¹⁰, Wolfgang Huber¹¹, Stefano Iacus¹², Rafael Irizarry¹³, Friedrich Leisch⁹, Cheng Li¹, Martin Maechler⁵, Anthony J Rossini¹⁴, Gunther Sawitzki¹⁵, Colin Smith¹⁶, Gordon Smyth¹⁷, Luke Tierney¹⁸, Jean YH Yang¹⁹ and Jianhua Zhang¹

[4341]

About Bioconductor

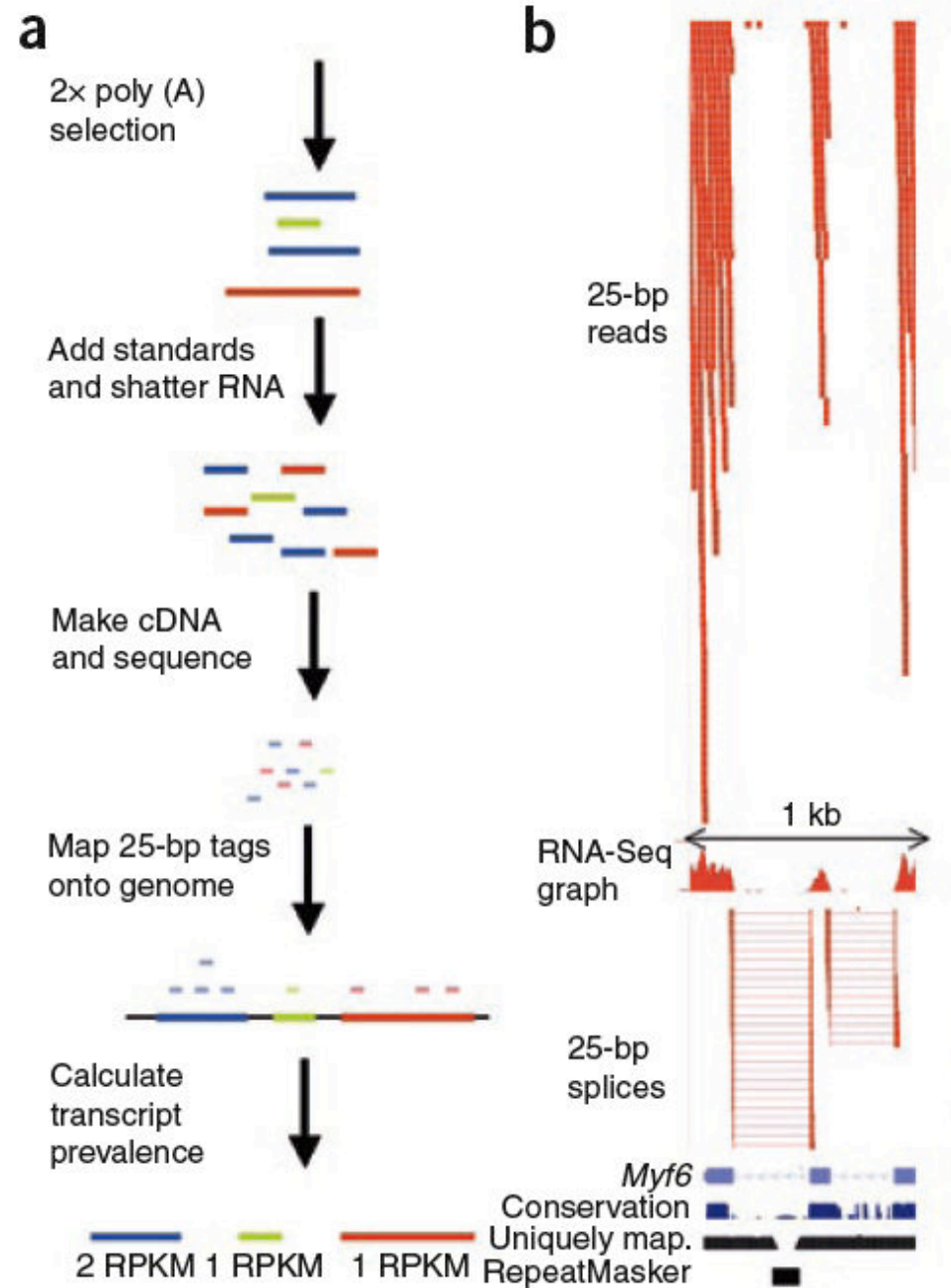
Bioconductor provides tools for the analysis and comprehension of high-throughput genomic data. Bioconductor uses the R statistical programming language, and is open source and open development. It has two releases each year, [554 software packages](#), and an active user community. Bioconductor is also available as an [Amazon Machine Image \(AMI\)](#).



RNA-seq

differential expression
analysis

differential isoform
analysis



Current RNA-seq workflows in Galaxy

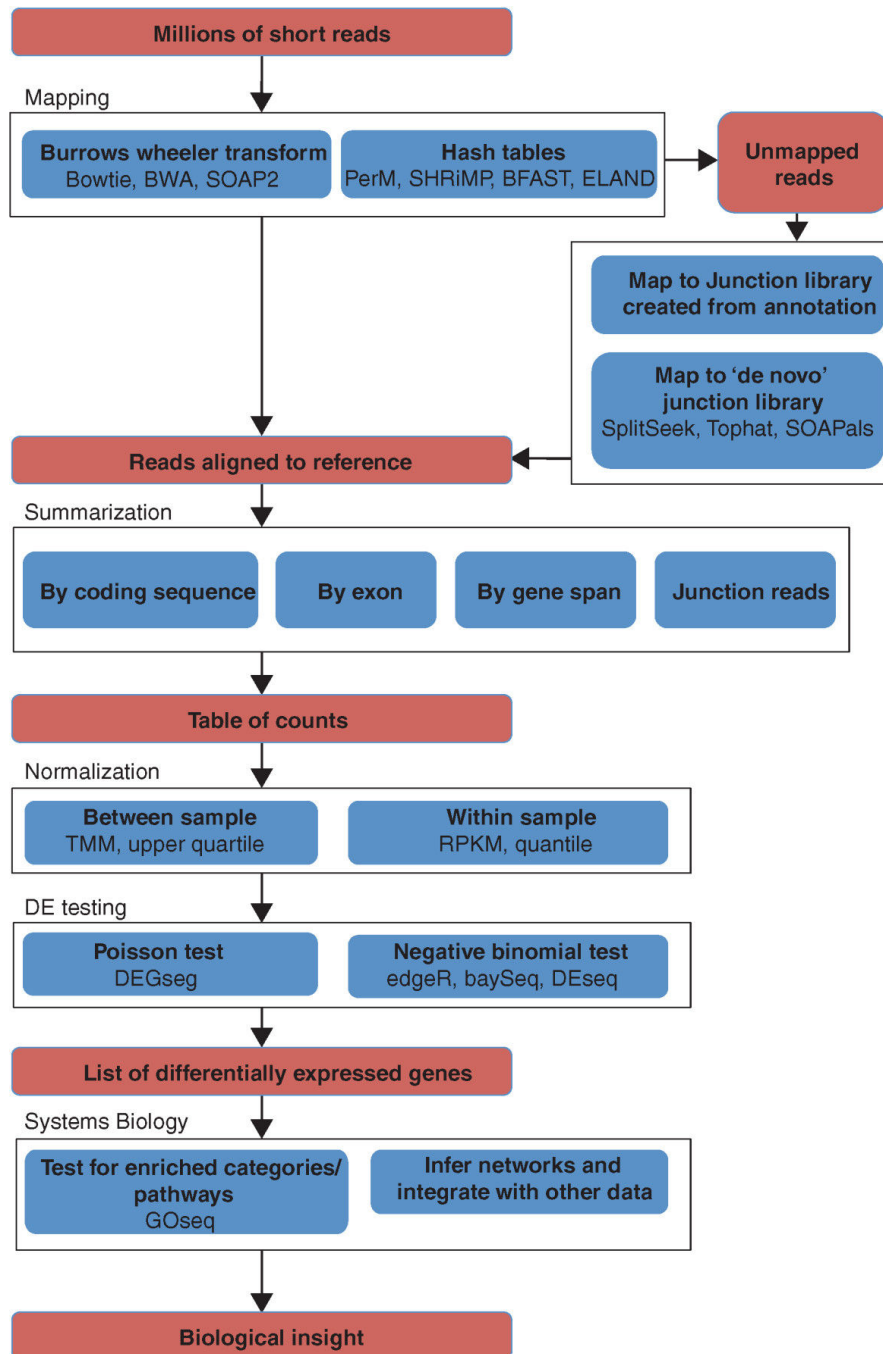
Bowtie, tophat – mapping reads

edgeR is an alternative to ‘cuffdiff’

Potential disadvantages of cuffdiff:

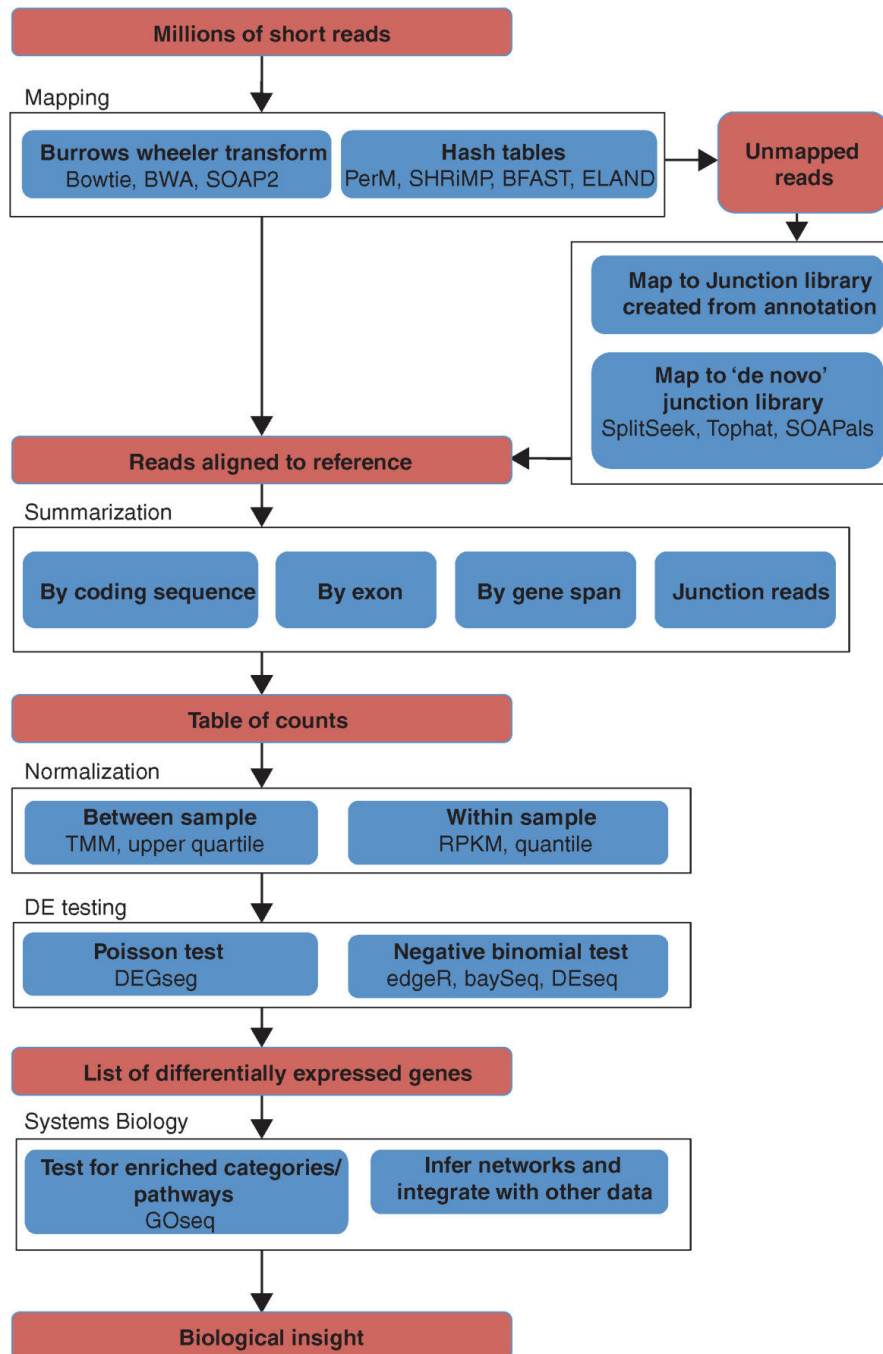
- only allows 2-sample comparisons
- Can be problematic at low depth / short read (our obs.)
- complicated

The screenshot shows the Galaxy web interface at <https://main.g2.bx.psu.edu>. The left sidebar contains a 'Tools' menu with categories: NGS: GATK TOOLS (beta), NGS: Variant Detection, NGS: Indel Analysis, NGS: Peak Calling, and NGS: RNA Analysis. Under 'NGS: RNA Analysis', there is a sub-section 'RNA-SEQ' with a list of tools: Tophat for Illumina Find splice junctions using RNA-seq data, Cufflinks transcript assembly and FPKM (RPKM) estimates for RNA-Seq data, Cuffcompare compare assembled transcripts to a reference annotation and track Cufflinks transcripts across multiple experiments, Cuffmerge merge together several Cufflinks assemblies, and Cuffdiff find significant changes in transcript expression, splicing, and promoter use. Below this is a 'FILTERING' section with 'Filter Combined Transcripts using tracking file'. Further down are 'NGS: Picard (beta)', 'BEDTools', and 'snpEff'. The main panel displays the 'Tophat for Illumina (v2.1.1)' tool configuration. It includes fields for 'RNA-Seq FASTQ file:', 'Nucleotide-space: Must have', 'Will you select a reference?' (set to 'Use a built-in index'), 'Built-ins were indexed using', 'Select a reference genome' (set to 'Arabidopsis lyrata: Araly1'), 'If your genome of interest is', 'Is this library mate-paired?' (set to 'Single-end'), 'TopHat settings to use:' (set to 'Default settings'), and 'Use the Full parameter list to'. An 'Execute' button is at the bottom. Below the tool configuration is a 'Tophat Overview' section with text: 'TopHat is a fast splice junction to mammalian-sized genome Bowtie, and then analyzes the exons. Please cite: Trapnell, splice junctions with RNA-Seq'. At the very bottom, there is a 'Know what you are doing' section.



edgeR workflow

(similar to DESeq,
baySeq, ...)



edgeR workflow

Counting:

- htseq-count (python)
- various tools within Bioconductor

What a standard edgeR analysis might look like

```
library(edgeR)
d <- DGEList(counts=D, group=grp)
d <- calcNormFactors(d)
d <- estimateTagwiseDisp(d)
```

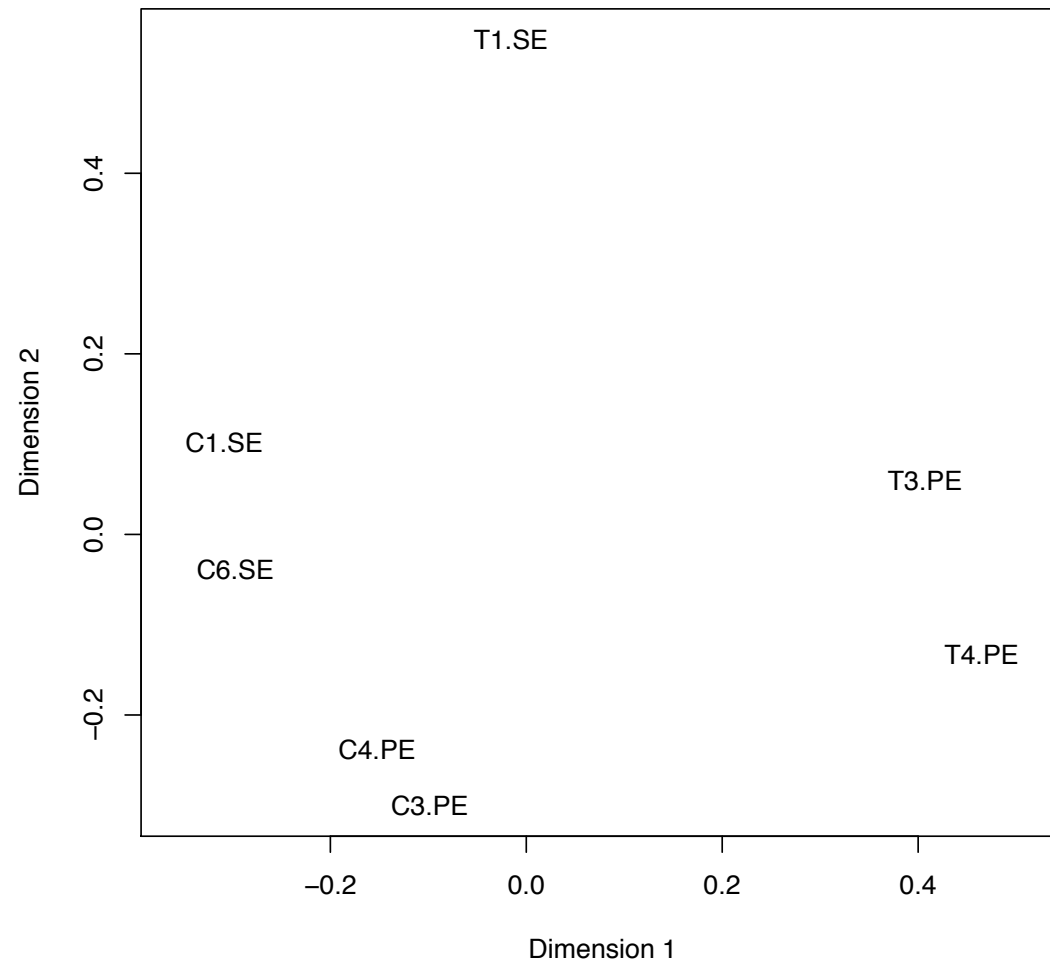
Or, start with table of metadata:

```
> samples
```

	rep	condition	libtype	shortname	countfile
1	T1	T	SE	T1.SE	S2_DRSC_CG8144_RNAi-1.count
2	T3	T	PE	T3.PE	S2_DRSC_CG8144_RNAi-3.count
3	T4	T	PE	T4.PE	S2_DRSC_CG8144_RNAi-4.count
4	C1	C	SE	C1.SE	S2_DRSC_Untreated-1.count
5	C6	C	SE	C6.SE	S2_DRSC_Untreated-6.count
6	C3	C	PE	C3.PE	S2_DRSC_Untreated-3.count
7	C4	C	PE	C4.PE	S2_DRSC_Untreated-4.count

What a standard edgeR analysis might look like

```
plotMDS(d, col=c("blue","orange")[factor(grp)])
```



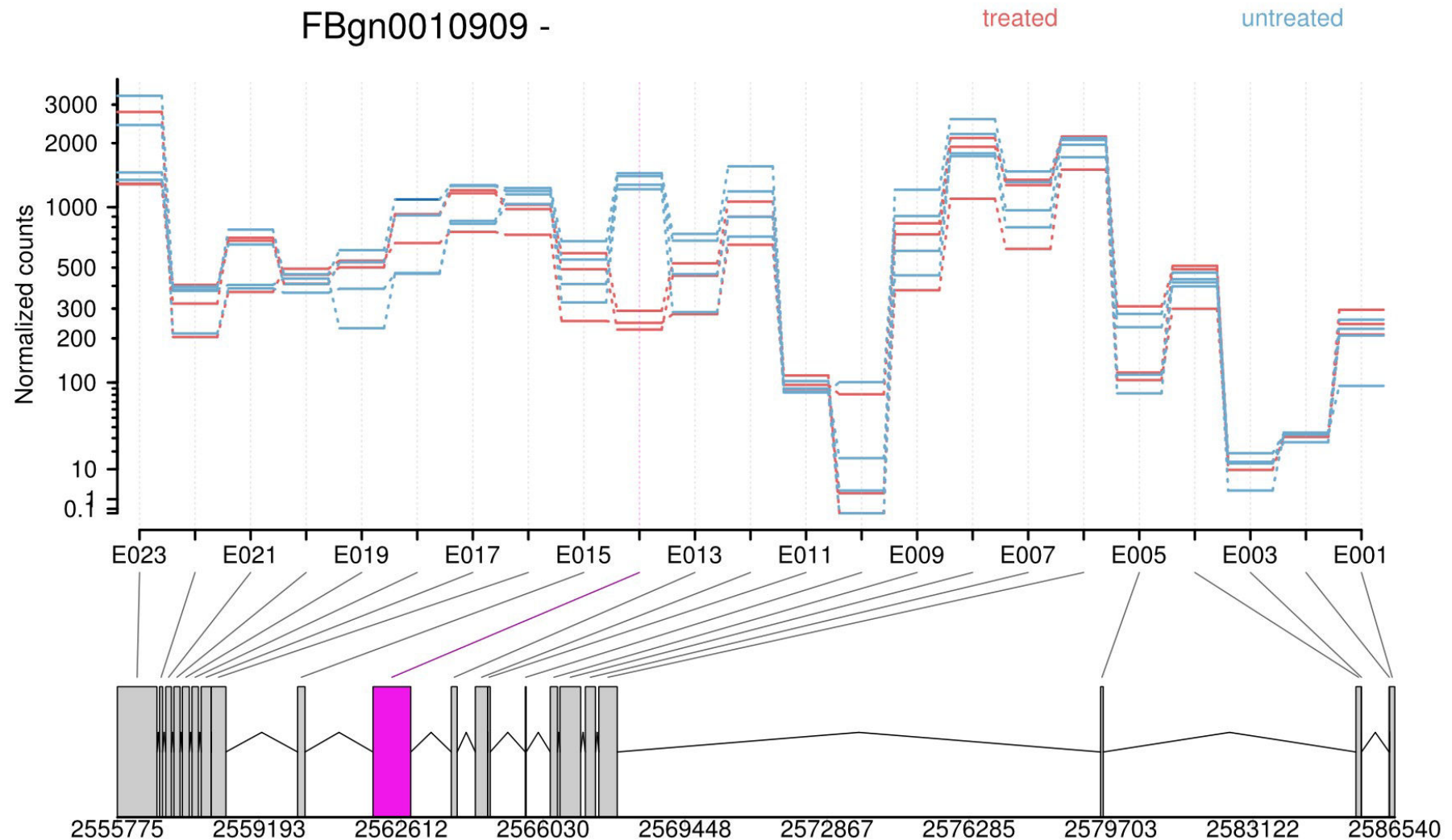
What a standard edgeR analysis might look like

```
de <- exactTest(d)
tt <- topTags(de, n=nrow(d))
rn.de <- rownames(tt)[tt$table$FDR < .05][1:500]
head(tt)
```

```
> head(tt$table)
```

	logFC	logCPM	PValue	FDR
ENSG00000151224	7.714164	11.404623	4.599520e-167	7.907494e-163
ENSG00000134339	7.428391	13.098576	6.366207e-164	5.472392e-160
ENSG00000173432	7.335286	13.061117	3.179161e-163	1.821871e-159
ENSG00000138115	7.431459	11.339722	2.968370e-159	1.275806e-155
ENSG00000141485	8.123902	9.329354	3.578253e-158	1.230347e-154
ENSG00000160868	7.239692	11.218554	3.996155e-154	1.128469e-150

DEXSeq – exon counts – differential isoform usage (edgeR has a similar mode)



edgeR integrated into Savant

Savant Genome Browser 2: visualization and analysis for population-scale genomics

Marc Fiume¹, Eric J. M. Smith¹, Andrew Brook¹, Dario Strbenac², Brian Turner³, Aziz M. Mezlini⁴, Mark D. Robinson⁵, Shoshana J. Wodak^{2,6} and Michael Brudno^{1,4,*}

¹Department of Computer Science, University of Toronto, Ontario, Canada M5S 2E4, ²Epigenetics Laboratory Cancer Research Program, Garvan Institute of Medical Research, Sydney, New South Wales 2011, Australia, ³Molecular Structure and Function Program, ⁴Centre for Computational Medicine, Hospital for Sick Children, Toronto, Ontario, Canada M5G 1L7, ⁵Institute of Molecular Life Sciences, University of Zurich, 8057 Zürich, Switzerland and ⁶Department of Biochemistry, University of Toronto, Ontario, Canada M5S 1A8

Received March 7, 2012; Revised April 23, 2012; Accepted April 24, 2012

edgeR plugin

The analysis of quantitative HTS data (e.g. from RNA-seq or ChIP-seq) relies on statistical procedures that highlight differential regions. For example, the density of mapped reads in a particular genomic region may represent enrichment level of a protein–DNA interaction (ChIP-seq), or gene expression level (RNA-seq). The edgeR plugin is a wrapper for software written in the R statistical programming language for the detection of significantly differentially enriched regions or expressed genes, relative to observed biological variation, directly within Savant (12). The plugin computes on multiple BAM tracks, some designated as Case and others as Control, and provides a table of ranked results, including the region locations, log-fold-changes, *P*-values and estimated false discovery rates of the change between conditions.

General, can be applied to other types of data (e.g. ChIP-seq)

Bioconductor and Galaxy, generally

- RGalaxy package

The following example illustrates these best practices (this function is in the *RGalaxy* package under the name `functionToGalaxify`):

```
function (inputfile1 = GalaxyInputFile(), inputfile2 = GalaxyInputFile(),
  plotTitle = character(), plotSubTitle = "My subtitle", outputfile1 = GalaxyOutput("mydata",
    "csv"), outputfile2 = GalaxyOutput("myplot", "pdf"))
{
  data1 <- tryCatch({
    as.matrix(read.delim(inputfile1, row.names = 1))
  }, error = function(err) {
    stop("failed to read first data file: ", conditionMessage(err))
  })
  data2 <- tryCatch({
    as.matrix(read.delim(inputfile2, row.names = 1))
  }, error = function(err) {
    stop("failed to read second data file: ", conditionMessage(err))
  })
  data3 <- data1 + data2
  write.csv(data3, file = outputfile1)
  pdf(outputfile2)
  if (missing(plotTitle))
    plotTitle <- ""
  plot(data3, main = plotTitle, sub = plotSubTitle)
  dev.off()
}
<environment: namespace:RGalaxy>
```

Bioconductor and Galaxy, generally

- RGalaxy package

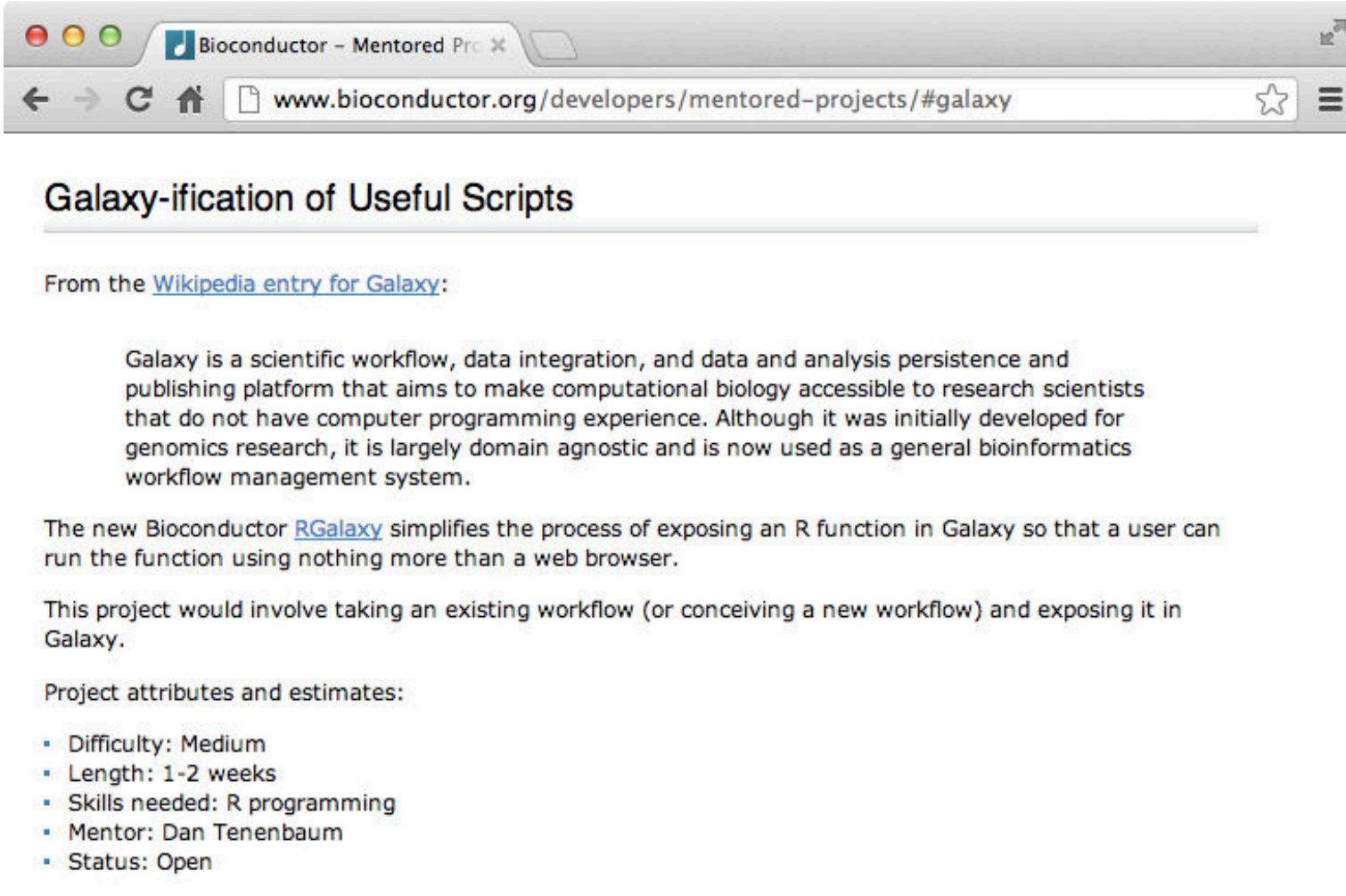
6 Adding a function to Galaxy

Now that you have written a function that follows the best practices described above, you can make it available to Galaxy as follows:

```
> galaxy(functionToGalaxify,  
+   manpage="functionToGalaxify",  
+   version="0.99.0",  
+   name="Add",  
+   package="RGalaxy",  
+   galaxyConfig=GalaxyConfig(getwd(), "mytool", "Test Section",  
+   "testSectionId")  
+   )
```

Bioconductor and Galaxy, generally

- Mentored projects at BioC



The screenshot shows a web browser window with the address bar displaying www.bioconductor.org/developers/mentored-projects/#galaxy. The page title is "Galaxy-ification of Useful Scripts". The content includes a quote from the Wikipedia entry for Galaxy, a description of the new Bioconductor RGalaxy project, and a list of project attributes and estimates.

Galaxy-ification of Useful Scripts

From the [Wikipedia entry for Galaxy](#):

Galaxy is a scientific workflow, data integration, and data and analysis persistence and publishing platform that aims to make computational biology accessible to research scientists that do not have computer programming experience. Although it was initially developed for genomics research, it is largely domain agnostic and is now used as a general bioinformatics workflow management system.

The new Bioconductor [RGalaxy](#) simplifies the process of exposing an R function in Galaxy so that a user can run the function using nothing more than a web browser.

This project would involve taking an existing workflow (or conceiving a new workflow) and exposing it in Galaxy.

Project attributes and estimates:

- Difficulty: Medium
- Length: 1-2 weeks
- Skills needed: R programming
- Mentor: Dan Tenenbaum
- Status: Open

Ross Lazarus (BakerIDI, Melbourne)

The screenshot shows a web browser window with a YouTube video player at the top. The video title is "Galaxy ToolFactory Medieval music background". Below the video player, the Galaxy ToolFactory web interface is visible. The interface has a dark header with the "Galaxy / ToolFactory" logo and navigation links: "Analyze Data", "Workflow", "Shared Data", "Admin", "Help", and "User". A status bar on the right indicates "Using 2.9 Mb".

The left sidebar contains a "Tools" section with a list of tool categories: "Make new tools or run scripts", "Get Data", "Send Data", "ENCODE Tools", "Lift-Over", "Text Manipulation", "Filter and Sort", "Join, Subtract and Group", "Convert Formats", "Extract Features", "Fetch Sequences", "Fetch Alignments", "Get Genomic Scores", "Operate on Genomic Intervals", "Statistics", "Wavelet Analysis", "Graph/Display Data", "Regional Variation", "Multiple regression", "Multivariate Analysis", "Evolution", "Motif Tools", "Multiple Alignments", "Metagenomic analyses", "FASTA manipulation", "NCBI BLAST+", and "NGS: QC and manipulation".

The main content area is titled "Tool Factory (version 0.06)". It contains several sections for configuring a new tool:

- Select an input file from your history:** A dropdown menu showing "Selection is Optional".
- Title for job outputs and new tool ID:** A text input field containing "My dynamic script".
- Create a tar.gz file ready for local toolshed entry:** A dropdown menu showing "No. Just run the script please".
- Create an HTML report with links to all output files and thumbnail links to PDF images:** A dropdown menu showing "Yes, arrange all outputs in an HTML output".
- Create a new (default tabular) history output:** A dropdown menu showing "My script writes to a new history output".
- Galaxy datatype for your tool's output file:** A dropdown menu showing "Tabular".
- Select the interpreter for your code. This must be available on the path of the execution host:** A dropdown menu showing "Rscript".
- Cut and paste the script to be executed here:** A large text area for pasting the script.

The right sidebar contains a "History" section with a list of tool runs. The first entry is "1: p value test" with 9 lines of output. The output is displayed in a table:

rawP
0.002
0.03
0.2
0.19
0.00001

Ross Lazarus (BakerIDI, Melbourne)

www.youtube.com/watch?v=Nzzc9zHZjJE&feature=plcp

Bioconductor - Install | aroma-proje | Import to Mendeley | Import to Mendeley | 10 Day Weather Fore | Genes and Immunity | graphics - Fonts in F | UZH/ETH - FGZ - S

Galaxy / ToolFactory

Analyze Data | Workflow | Shared Data | Admin | Help | User | Using 2.9 Mb

Tools

Make new tools or run scripts

- Tool Factory Makes scripts into tools

Get Data

Send Data

ENCODE Tools

Lift-Over

Text Manipulation

Filter and Sort

Join, Subtract and Group

Convert Formats

Extract Features

Fetch Sequences

Fetch Alignments

Get Genomic Scores

Operate on Genomic Intervals

Statistics

Wavelet Analysis

Graph/Display Data

Regional Variation

Multiple regression

Multivariate Analysis

Evolution

Motif Tools

Multiple Alignments

Metagenomic analyses

FASTA manipulation

NCBI BLAST+

NGS: QC and manipulation

Create a new (default tabular) history output:

My script writes to a new history output

This is useful if your script creates a single new tabular file you want to appear in the history after the tool executes

Galaxy datatype for your tool's output file:

Tabular

You may need to edit the xml to extend this list

Select the interpreter for your code. This must be available on the path of the execution host:

Rscript

Cut and paste the script to be executed here:

```
ourargs = commandArgs(TRUE)
inf = ourargs[1]
outf = ourargs[2]
inp = read.table(inf,head=T,row.names=NULL,sep='\t')
p = inp[,column]
q = p.adjust(p,method='BH')
outp = cbind(inp,'BH Adjusted p-value'=q)
write.table(outp,outf, quote=FALSE, sep='\t',row.names=F,col.names=T)
```

Script must deal with two command line parameters: Path to input tabular file path (or 'None' if none selected) and path to output tabular history file (or 'None').

Execute

Local Admins ONLY Only users whose IDs found in the local admin_user configuration setting in universe_wsgi.ini can run this tool.

What it does This tool enables a user to paste and submit an arbitrary R/python/perl script to Galaxy.

Input options This version is limited to simple transformation or reporting requiring only a single input file selected from the history.

Output options Optional script outputs include one single new history tabular file, or for scripts that create multiple outputs, a new HTML report linking all the files and images created by the script can be automatically generated.

Tool Generation option Once the script is working with test data, this tool will optionally generate a new Galaxy tool in a gzip file ready to upload to your local toolshed for sharing and installation. Provide a small sample input when you run generate the tool because it will become the input for the generated functional test.

History

ToolFactory June 2012 47 bytes

1: p value test

9 lines
format: tabular, database: 2
Info: None

1
rawP
0.002
0.03
0.2
0.19
0.00001

Ross Lazarus (BakerIDI, Melbourne)

www.youtube.com/watch?v=Nzzc9zHZJJE&feature=plcp

nductor - Install | Install | aroma-proje | Import to Mendeley | Import to Mendeley | TWC 10 Day Weather Fore | Genes and Immunity | graphics - Fonts in P | UZH/ETH - FGCZ - S | >> Other Book

Galaxy / ToolFactory

Analyze Data | Workflow | Shared Data | Admin | Help | User | Using 3.0 Mb

Tools

Make new tools or run scripts

- Tool Factory Makes scripts into tools

Get Data

Send Data

ENCODE Tools

Lift-Over

Text Manipulation

Filter and Sort

Join, Subtract and Group

Convert Formats

Extract Features

Fetch Sequences

Fetch Alignments

Get Genomic Scores

Operate on Genomic Intervals

Statistics

Wavelet Analysis

Graph/Display Data

Regional Variation

Multiple regression

Multivariate Analysis

Evolution

Motif Tools

Multiple Alignments

Metagenomic analyses

FASTA manipulation

NCBI BLAST+

NGS: QC and manipulation

rawP	BH Adjusted p-value
0.002	0.008
0.03	0.06
0.2	0.228571428571429
0.19	0.228571428571429
1e-05	8e-05
0.029	0.06
0.1	0.16
0.3	0.3

History

ToolFactory June 2012 7.2 Kb

3: [pvalueadjuster.toolshed.gz](#) 7.0 Kb
format: toolshed.gz, database: 2
Info: wrote pvalueadjuster/pvalueadjuster.py

2: [pvalueadjuster.toolshed.gz](#) 7.0 Kb
format: toolshed.gz, database: 2
Info: wrote pvalueadjuster/pvalueadjuster.py

1: [pvalueadjuster.toolshed.gz](#) 7.0 Kb
format: toolshed.gz, database: 2
Info: wrote pvalueadjuster/pvalueadjuster.py

9-times

format: tabular, database: 2
Info: None

1

rawP

0.002

0.03

0.2

0.19

0.00001

Ross Lazarus (BakerIDI, Melbourne)

→ ↻ www.youtube.com/watch?v=Nzzc9zHZjJE&feature=plcp

bioconductor - Install | aroma-proje | Import to Mendeley | Import to Mendeley | TWC 10 Day Weather Fore | Genes and Immunity | graphics - Fonts in P | UZH/ETH - FGZ - S

Galaxy Tool Shed

Repositories Admin Help User

1 valid tools on Jun 24, 2012

Search

- Search for valid tools
- Search for workflows

Repositories

- Browse by category
- Browse my repositories
- Browse my invalid tools
- Create new repository

Repository Actions

Upload a single file or a tarball

⚠ Uploading may take a while, depending upon the size of the file. Wait until a message is displayed in your browser after clicking the **Upload** button below.

File:

No file chosen

Url:

Enter a URL to upload your files via http.

Uncompress files?

Supported compression types are gz and bz2. If **Yes** is selected, the uploaded file will be uncompressed. However, if the uploaded file is an archive that contains compressed files, the contained files will not be uncompressed. For example, if the uploaded compressed file is some_file.tar.gz, some_file.tar will be uncompressed and extracted, but if some_file.tar contains some_contained_file.gz, the contained file will not be uncompressed.

Change set commit message:

This is the commit message for the mercurial change set that will be created by this upload.

3:40 / 7:35

Ross Lazarus (BakerIDI, Melbourne)

www.youtube.com/watch?v=Nzzc9zHZJjE&feature=plcp

Bioconductor - Install | aroma-project | Import to Mendeley | Import to Mendeley | TWC 10 Day Weather Forecast | Genes and Immunity | graphics - Fonts in R | UZH/ETH - FCCZ - S | » Other E

Galaxy / ToolFactory

Analyze Data Workflow Shared Data Admin Help User Using 3.0 Mb

Administration

Security

- Manage users
- Manage groups
- Manage roles

Data

- Manage quotas
- Manage data libraries

Server

- Tool versions
- Reload a tool's configuration
- Profile memory usage
- Manage jobs
- Manage installed tool shed repositories

Tool sheds

- Search and browse tool sheds

Form Definitions

- Manage form definitions

Sample Tracking

- Manage sequencers and external services
- Manage request types
- Sequencing requests
- Find samples

Valid repositories

Name ↓	Synopsis	Revision	Owner
pvalueadjuster	Adjust p values	0:aa484a793fb2	rosstheboss
transpose	Transpose a tabular file	0:b6f89499c538	rosstheboss

4:25 / 7:35

Ross Lazarus (BakerIDI, Melbourne)

The screenshot displays the Galaxy ToolFactory web interface. The top navigation bar includes 'Galaxy / ToolFactory', 'Analyze Data', 'Workflow', 'Shared Data', 'Admin', 'Help', and 'User'. A status bar on the right indicates 'Using 3.0 Mb'.

Tools Panel (Left): A list of tool categories is shown, including Graph/Display Data, Regional Variation, Multiple regression, Multivariate Analysis, Evolution, Motif Tools, Multiple Alignments, Metagenomic analyses, FASTA manipulation, NCBI BLAST+, NGS: QC and manipulation, NGS: Picard (beta), NGS: Mapping, NGS: Indel Analysis, NGS: RNA Analysis, NGS: SAM Tools, NGS: GATK Tools (beta), NGS: Variant Detection, NGS: Peak Calling, NGS: Simulation, SNP/WGA: Data; Filters, SNP/WGA: QC; LD; Plots, SNP/WGA: Statistical Models, Phenotype Association, VCF Tools, and ToolFactory Generated tools. The 'pvalueadjuster tool_desc' is selected under the last category.

Tool Configuration (Center):

- Tool:** pvalueadjuster (version 0.01)
- Select a suitable input file from your history:** 2: pvalueadjuster.tabular
- Supply a name for the outputs to remind you what they contain:** pvalueadju
- Execute** button

What it Does: Does a BH p value adjustment for FDR control

Script: Pressing execute will run the following code over your input file and generate some outputs in your history:

```
# use p.adjust - assumes a HEADER row and column 1 - please fix for any real use
column = 1 # adjust if necessary for some other kind of input
ourargs = commandArgs(TRUE)
inf = ourargs[1]
outf = ourargs[2]
inp = read.table(inf, head=T, row.names=NULL, sep='\t')
p = inp[,column]
q = p.adjust(p, method='BH')
outp = cbind(inp, 'BH Adjusted p-value'=q)
write.table(outp, outf, quote=FALSE, sep='\t', row.names=F, col.names=T)
```

Attribution: This Galaxy tool was created by ross.lazarus@gmail.com at 24/06/2012 15:58:34 using the Galaxy ToolFactory. See <https://bitbucket.org/fubar/galaxytoolfactory> for details of that project

History Panel (Right):

- ToolFactory June 2012** 7.2 Kb
- 3: pvalueadjuster.toolshed.gz** 7.0 Kb
format: toolshed.gz, database: 2
Info: wrote pvalueadjuster/pvalueadjuster.py
- 2: pvalueadjuster.tabular**
- 1: p value test** 9 lines
format: tabular, database: 2
Info: None

The '1: p value test' entry is expanded, showing a table with a single column 'rawP' and five rows of p-values:

rawP
0.002
0.03
0.2
0.19
0.00001

The bottom of the image shows a video player interface with a progress bar at 5:11 / 7:35.

Ross Lazarus (BakerIDI, Melbourne)

→ ↻ www.youtube.com/watch?v=Nzzc9zHZJjE&feature=plcp

Bioconductor - Install | aroma-proj | Import to Mendeley | Import to Mendeley | TWC 10 Day Weather Fore | Genes and Immunity | graphics - Fonts in P | UZH/ETH - FGCZ - S

Galaxy / ToolFactory

Analyze Data Workflow Shared Data Admin Help User Using 3.0 Mb

Tools

- Graph/Display Data
- Regional Variation
- Multiple regression
- Multivariate Analysis
- Evolution
- Motif Tools
- Multiple Alignments
- Metagenomic analyses
- FASTA manipulation
- NCBI BLAST+
- NGS: QC and manipulation
- NGS: Picard (beta)
- NGS: Mapping
- NGS: Indel Analysis
- NGS: RNA Analysis
- NGS: SAM Tools
- NGS: GATK Tools (beta)
- NGS: Variant Detection
- NGS: Peak Calling
- NGS: Simulation
- SNP/WGA: Data; Filters
- SNP/WGA: QC; LD; Plots
- SNP/WGA: Statistical Models
- Phenotype Association
- VCF Tools
- ToolFactory Generated tools
 - pvalueadjuster tool_desc

Workflows

rawP	BH Adjusted p-value
0.002	0.008
0.03	0.06
0.2	0.228571428571429
0.19	0.228571428571429
1e-05	8e-05
0.029	0.06
0.1	0.16
0.3	0.3

History

ToolFactory June 2012 7.3 Kb

4: test 9 lines
format: tabular, database: ?

1	2
rawP	BH Adjusted p-value
0.002	0.008
0.03	0.06
0.2	0.228571428571429
0.19	0.228571428571429
1e-05	8e-05

3: pvalueadjuster.toolshed.qz 7.0 Kb
format: toolshed.qz, database: ?
Info: wrote pvalueadjuster/pvalueadjuster.py
binary data

2: pvalueadjuster.tabular

1: p value test

5:57 / 7:35

Some challenges in making R/ Bioconductor tools available (to Galaxy)

- How to allow facilitate general design matrices?

```
> design <- model.matrix( ~ libtype + condition, samples)
> design
```

	(Intercept)	libtypeSE	conditionT
S2_DRSC_CG8144_RNAi-1.bam	1	1	1
S2_DRSC_CG8144_RNAi-3.bam	1	0	1
S2_DRSC_CG8144_RNAi-4.bam	1	0	1
S2_DRSC_Untreated-1.bam	1	1	0
S2_DRSC_Untreated-6.bam	1	1	0
S2_DRSC_Untreated-3.bam	1	0	0
S2_DRSC_Untreated-4.bam	1	0	0

Some challenges in making R/ Bioconductor tools available (to Galaxy)

- How to specify metadata?
- How to adequately give all the plotting options?
- Versioning
- Making reproducible analyses available
- Only polished analyses made available?
- Citations
- Best practices – multiple best practices?

Copy-number-aware differential analysis of quantitative DNA sequencing data

Mark D. Robinson [1,2,3,*], Dario Strbenac [3], Clare Stirzaker [3,6], Aaron L. Statham [3], Jenny Song [3], Terence P. Speed [4,5], Susan J. Clark [3,6]

[1] Institute of Molecular Life Sciences, University of Zurich, Winterthurerstrasse 190 CH-8057 Zurich, Switzerland

[2] Swiss Institute of Bioinformatics, University of Zurich, Zurich, Switzerland

[3] Epigenetics Laboratory, Cancer Research Program, Garvan Institute of Medical Research, Sydney 2010, New South Wales, Australia

[4] Bioinformatics Division, Walter and Eliza Hall Institute of Medical Research, Parkville, Melbourne 3052, Victoria, Australia

[5] Department of Medical Biology, University of Melbourne, 3050, Victoria, Australia

[6] St Vincent's Clinical School, University of NSW, Sydney, NSW, Australia

[*] Corresponding author: mark.robinson@imls.uzh.ch.

Summary

Developments in microarray and high throughput sequencing (HTS) technologies have resulted in a rapid expansion of research into epigenomic changes that occur in normal development and in the progression of disease, such as cancer. Not surprisingly, copy number variation (CNV) has a direct effect on HTS read densities and can therefore bias differential detection results. We have developed a flexible approach called ABCD-DNA (Affinity Based Copy-number-aware Differential quantitative DNA sequencing analyses) that integrates CNV and other systematic factors directly into the differential enrichment engine.

Supplementary Data (semi-processed), R Code for all Figures and analyses:

- Archive with all data/code [[ABCD-DNA_Supplement_RCode_Data.tar.gz](#), 649 MB]
- Sweaved PDF document with annotated ABCD-DNA analysis [[ABCD-DNA.pdf](#)]
- [Track] file (use with UCSC Genome Browser - hg18) for ABCD-DNA (and other algorithm) calls, Illumina 450k array data

Bioconductor meeting in Zurich – December 13-14

UZH/ETH – FGZ – Overview

www.fgcz.ch/Bioconductor2012/Overview

Bioconductor European Developers' Workshop2012
[Overview](#)
[Invited Speakers](#)
[Programme outline](#)
[Location](#)
[Registration](#)
[Hotels](#)
[Contact](#)

Overview

This workshop is aimed at Bioconductor contributors and bioinformaticians who wish to contribute packages to the Bioconductor project. The aim of the meeting is to foster the exchange of technical expertise, to keep contributors up to speed with the latest developments and to coordinate related efforts.


Topics this year will include:


- New developments in Bioconductor core packages and infrastructure
- Infrastructure and methods for understanding gene regulation
- Next Generation Data handling and analysis
- Reproducible research and software engineering
- Tools for the management and analysis of proteomics datasets


Time and location


December 13-14th 2012
University of Zürich, Irchel Campus, room Y55 L06/08
Invited speakers


The workshop is supported by

**University of Zurich**

**SIB**
Swiss Institute of Bioinformatics

**FNSNF**
FONDS NATIONAL SUISSE
SCHWEIZERISCHER NATIONALFONDS
FONDO NAZIONALE SVIZZERO
SWISS NATIONAL SCIENCE FOUNDATION

**SystemsX.ch**
The Swiss Initiative in Systems Biology


**Agilent Technologies**

News & Events

[NGS Education Facility - Courses](#)
[1st CECE Bioinformatics and Functional Genomics Workshop](#)
[FGZ Next Generation Sequencing Newsletter](#)
[PRIME-XS Proteomics Transnational Access Program](#)
[Bioconductor European Developers' Workshop 2012](#)

Quicklinks

[Intranet](#): Info and Docu
[Project](#): Request Projects
[Order](#): Order Services



TOP