

Accessible, Transparent and Reproducible Analysis with Galaxy

Application of Next Generation Sequencing Technologies for Whole Transcriptome and Genome Analysis

ABRF 2013

Saturday, March 2, 2013

Palm Springs, California, United States

Dave Clements
Emory University



This Workshop

Demonstrate **Galaxy** with a **hands-on** walk through of
an example **RNA-Seq** analysis

Introduce Galaxy and Galaxy Project as we go.

Complements talk on Monday:

Galaxy for Core Facilities

*(W6) Community Resource Solutions to Analyzing Large
Genomic Data Sets*

Slides are at

bit.ly/ABRFgxyWS1 wiki.galaxyproject.org/Events

Demonstrate **Galaxy** with a **hands-on**
walk through of an **RNA-Seq** analysis

<http://usegalaxy.org/u/jeremy/p/galaxy-rna-seq-analysis-exercise>

<http://bit.ly/gxyrnaseq>

<http://bit.ly/ABRFgxy1>

<http://bit.ly/ABRFgxy2>

<http://bit.ly/ABRFgxy3>

RNA-seq Exercise: A Plan

- Get input datasets; hg19, will mostly map to chr19
- Look at quality
- Trim as we see fit.
- Map the reads to the human reference using Tophat
- Run Cufflinks on Tophat output to assemble reads into transcripts
- Visualize it

bit.ly/gxyrnaseq

bit.ly/ABRFgxy[123]

RNA-seq Exercise: A Plan

- Get input datasets; hg19, will mostly map to chr19
 - All datasets are FASTQ and from the Body Map 2.0 project
- **Shared Data → Data Libraries**

bit.ly/gxyrnaseq

bit.ly/ABRFgxy[123]

What is FASTQ?

- Specifies sequence (FASTA) and quality scores (PHRED)
- Text format, 4 lines per entry

```
@SEQ_ID
GATTTGGGGTTCAAAGCAGTATCGATCAAATAGTAAATCCATTTGTTCAACTCACAGTTT
+
!''*((( (**+))%%%+)(%%%).1***-+*''))**55CCF>>>>>CCCCCCC65
```

- FASTQ is such a cool standard, there are 3 (or 5) of them!

```
SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS
.....IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
.....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
!"#$%&'()*+,-./0123456789:;<=>?@ABCDEFGHIJKLMN O PQRSTU VWXYZ[\]^_`abcdefghijklmnopqrstuvwxy z{|}~
|                                     |         |           |                               |                                   |
33                                59        64          73                                  104                                      126

S - Sanger      Phred+33,   93 values    (0, 93) (0 to 60 expected in raw reads)
I - Illumina 1.3 Phred+64,   62 values    (0, 62) (0 to 40 expected in raw reads)
X - Solexa      Solexa+64,   67 values    (-5, 62) (-5 to 40 expected in raw reads)
```

http://en.wikipedia.org/wiki/FASTQ_format

RNA-seq Exercise: A Plan

- Get input datasets; hg19, will mostly map to chr19
- Look at quality: Option 1
 - NGS QC and Manipulation → **Compute Quality Statistics**
 - NGS QC and Manipulation → **Draw quality score boxplot**
 - Gives you no control over how it is calculated or presented.

<http://bit.ly/gxynaseq>

RNA-seq Exercise: A Plan

- Get input datasets; hg19, will mostly map to chr19
- Look at quality: Option 2
 - NGS QC and Manipulation → **FastQ Summary Statistics**
 - Graph / Display Data → **Boxplot of quality statistics**
 - Gives you a lot of control over what the box plot looks like, but no additional information

<http://bit.ly/gxyrnaseq>

RNA-seq Exercise: A Plan

- Get input datasets; hg19, will mostly map to chr19
- Look at quality: Option 3
 - NGS QC and Manipulation → **Fastqc**
 - Gives you a lot more information but little control over how it is calculated or presented.

<http://bit.ly/gxyrnaseq>

RNA-seq Exercise: A Plan

- Get input datasets; hg19, will mostly map to chr19
- Look at quality
- Trim as we see fit: Option 1
 - **NGS QC and Manipulation** → **FASTQ Trimmer by column**
 - Trim same number of columns from every record
 - Can specify different trim for 5' and 3' ends

<http://bit.ly/gxyrnaseq>

RNA-seq Exercise: A Plan

- Get input datasets; hg19, will mostly map to chr19
- Look at quality
- ~~Trim~~ Filter as we see fit: Option 2
 - NGS QC and Manipulation → **Filter FASTQ reads by quality score and length**
 - Keep or discard whole reads at a time
 - Can have different thresholds for different regions of the reads.
 - Keeps original read length.

<http://bit.ly/gxyrnaseq>

RNA-seq Exercise: A Plan

- Get input datasets; hg19, will mostly map to chr19
- Look at quality
- Trim as we see fit: Option 3
 - NGS QC and Manipulation → **FASTQ Quality Trimmer by sliding window**
 - Trim from both ends, using sliding windows, until you hit a high-quality section.
 - **Produces variable length reads**

<http://bit.ly/gxynaseq>

RNA-seq Exercise: A Plan

- Get input datasets; hg19, will mostly map to chr19
- Look at quality
- Trim as we see fit.
- Map the reads to the human reference using Tophat
 - *Imagine pages and pages of discussion on the intricacies and pitfalls of RNA-seq mapping here.*

<http://bit.ly/gxyrnaseq>

What is Galaxy?

- **A free (for everyone) web service** integrating a wealth of tools, compute resources, terabytes of reference data and permanent storage
- **Open source software** that makes integrating your own tools and data and customizing for your own site simple
- These options result in several **ways to use Galaxy**

<http://galaxyproject.org>

Galaxy is available ...

- As a free (for everyone) web service

<http://usegalaxy.org>

However, *a centralized solution cannot scale to meet the analysis needs of the entire world.*

Galaxy is available ...

- As a free (for everyone) web service

<http://usegalaxy.org>

- **As open source software**

<http://getgalaxy.org>

As Open Source Software: Local Galaxy Instances

- Galaxy is designed for local installation and customization
- Easily integrate new tools
- Easy to deploy and manage on nearly any (unix) system
- Run jobs on existing compute clusters
- Requires a computational resource on which to be deployed

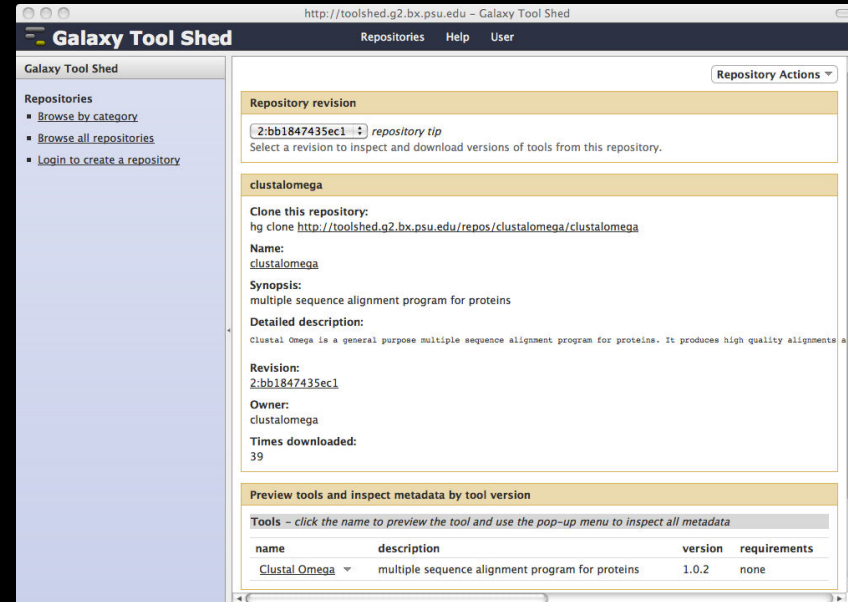
<http://getgalaxy.org>

Encourage **Local** Galaxy Instances

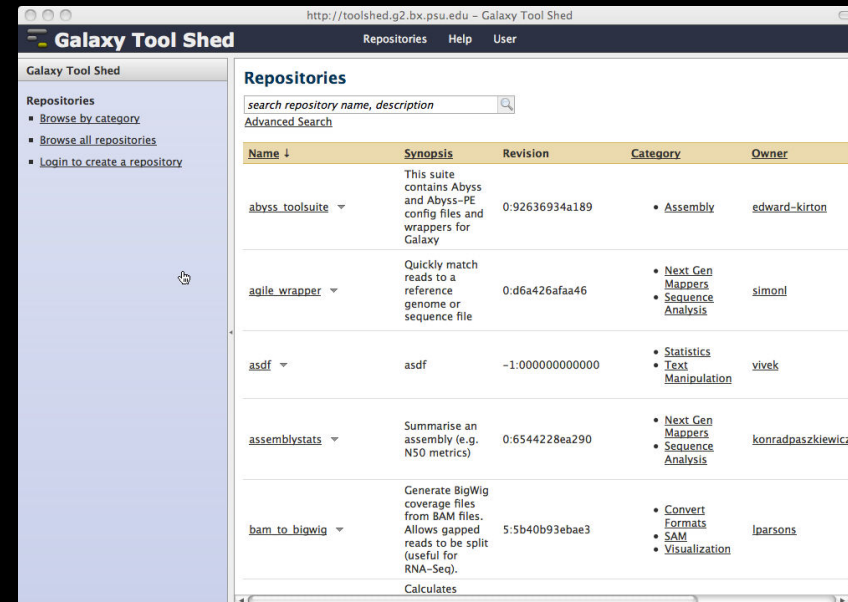
- Encourage and support Local Galaxy Instances
- Support **increasingly decentralized model** and improve access to existing resources
- Focus on building **infrastructure to enable the community to integrate and share tools, workflows, and best practices**

Galaxy Tool Shed

<http://toolshed.g2.bx.psu.edu>



The screenshot shows the Galaxy Tool Shed interface for the 'clustalomega' repository. The left sidebar contains links to 'Browse by category', 'Browse all repositories', and 'Login to create a repository'. The main content area displays the 'Repository revision' as '2:bb1847435ec1' with a 'repository tip' and a note to 'Select a revision to inspect and download versions of tools from this repository.' Below this, the repository name 'clustalomega' is shown, followed by a 'Clone this repository:' link, the 'Name: clustalomega', a 'Synopsis: multiple sequence alignment program for proteins', and a 'Detailed description: Clustal Omega is a general purpose multiple sequence alignment program for proteins. It produces high quality alignments'. The 'Revision: 2:bb1847435ec1', 'Owner: clustalomega', and 'Times downloaded: 39' are also listed. At the bottom, a table titled 'Tools - click the name to preview the tool and use the pop-up menu to inspect all metadata' shows one tool: 'Clustal Omega' with a description 'multiple sequence alignment program for proteins', version '1.0.2', and 'none' requirements.



The screenshot shows the Galaxy Tool Shed interface with a list of repositories. The left sidebar is the same as the previous screenshot. The main content area has a search bar and an 'Advanced Search' link. Below is a table with columns: 'Name', 'Synopsis', 'Revision', 'Category', and 'Owner'. The table lists several repositories: 'abyss_toolsuite' (This suite contains Abyss and Abyss-PE config files and wrappers for Galaxy), 'agile_wrapper' (Quickly match reads to a reference genome or sequence file), 'asdf' (asdf), 'assemblystats' (Summarise an assembly (e.g. N50 metrics)), and 'bam to bigwig' (Generate BigWig coverage files from BAM files. Allows gapped reads to be split (useful for RNA-Seq). Calculates). The 'Category' column lists categories like 'Assembly', 'Next Gen Mappers', 'Sequence Analysis', 'Statistics', 'Text Manipulation', 'Convert Formats', 'SAM', and 'Visualization'. The 'Owner' column lists names like 'edward-kirton', 'simonl', 'vivek', 'konradpaskiewicz', and 'Inarsons'.

Encourage **Public** Galaxy Instances

<http://wiki.galaxyproject.org/PublicGalaxyServers>

Interested in:

ChIP-chip and ChIP-seq?

✓ Cistrome

Statistical Analysis?

✓ Genomic Hyperbrowser

Protein synthesis?

✓ GWIPS-viz

de novo assembly?

✓ CBIIT Galaxy

Reasoning with ontologies?

✓ OPPL Galaxy

Repeats

✓ RepeatExplorer

Everything?

✓ Andromeda

As Open Source Software: Local Galaxy Instances

- Galaxy is designed for local installation and customization
- Easily integrate new tools
- Easy to deploy and manage on nearly any (unix) system
- Run jobs on existing compute clusters
- Requires a **computational resource** on which to be deployed

<http://getgalaxy.org>

Got your own cluster?

- Control **where** tool execution happens
- Galaxy **works with any DRMAA** compliant cluster job scheduler (which is most of them).
- Galaxy is **just another client** to your scheduler.



Galaxy is available ...

- As a free (for everyone) web service

<http://usegalaxy.org>

- As open source software

<http://getgalaxy.org>



- *On the Cloud*

<http://usegalaxy.org/cloud>

We are using this right now

<http://aws.amazon.com/education>

Galaxy Resources and Community

Mailing Lists (very active)

Unified Search

Issues Board

Events Calendar, News Feed

Community Wiki

GalaxyAdmins

Screencasts

Tool Shed

Public Installs

CiteULike group, Mendeley mirror

Annual Community Meeting

<http://wiki.galaxyproject.org>

Mailing Lists

<http://wiki.galaxyproject.org/MailingLists>

Galaxy-Announce

Project announcements, low volume, moderated

Low volume (42 posts, 1600 members in 2012)

Galaxy-User

Questions about using Galaxy and usegalaxy.org


High volume (2900 posts, 2700 members in 2012)

Galaxy-Dev

Questions about developing for and deploying Galaxy

High volume (4500 posts, 850 members in 2012)

Unified Search: <http://galaxyproject.org/search>

 Galaxy Web Search

Google™ Custom Search

Search ✕

Search the entire set of Galaxy web sites and mailing lists using Google.

[Run this search at Google.com \(useful for bookmarking\)](#)

Want a [different search](#)?

[Project home](#)

Find

Everything on ...

Tools for ...

Email about ...

Source code for ...

Published Histories, Pages, Workflows, about ...

Documentation on ...

Papers using Galaxy for ...

Related feature requests

Community can create, vote and comment on issues

The screenshot displays a Trello board for the 'Galaxy Project' under the 'Development Inbox' list. The board is organized into four main columns: 'Inbox', 'Developer ideas', 'Bug Reports', and 'Issues from Bitbucket'. Each column contains several cards representing different development tasks or issues. The 'Inbox' column has five cards, including one about adding cards and another about filtering and sorting. The 'Developer ideas' column has five cards, such as 'Anonymous use of workflows/visualizations' and 'Feature Request: the ability to restart a failed workflow'. The 'Bug Reports' column has five cards, including 'Issues with workflow step hiding not persisting' and 'Workflow View Broken in Toolshed?'. The 'Issues from Bitbucket' column has five cards, such as '5: Option to disable automatic history creation' and '6: Option to require that histories have names'. On the right side of the board, there is a 'Members' section showing a grid of member avatars, an 'Add Members...' button, and a 'Board' section with 'Options', 'Add List', and 'Filter Cards' buttons. Below these is an 'Activity' section showing recent actions, such as 'Dannon Baker added API: Library Contents to Developer ideas and' and 'g2roboto on Feature request: manually hide datasets'. Each card in the columns includes a title, a description, and a 'votes' count with a thumbs-up icon. Some cards also have a 'comments' count with a speech bubble icon. The bottom of the board has a blue bar with an 'Add a card...' button.

Trello Board: Galaxy: Development Inbox

Columns:

- Inbox**
 - To add cards, use the <http://galaxyproject.org/training>
 - Filter and Sort: "Select" tool not dealing with special characters right
 - Uploaded fastq file datatype not usable in BWA
 - Reference genome request: GATK-ordered hg19
 - Feature request: manually hide datasets
- Developer ideas**
 - Anonymous use of workflows/visualizations
 - Feature Request: the ability to restart a failed workflow from the point of failure;
 - Google Drive / Dropbox / Box / ... integration
 - Bug report: always import deleted datasets
 - Standalone web application(s) for visualizations
 - Enh: Archiving histories
 - Modify data library upload completion message
 - Display in UI runtime
- Bug Reports**
 - Issues with workflow step hiding not persisting
 - Workflow View Broken in Toolshed?
 - Unable to run jobs when user job limits are set
 - Fix tool tip FASTQ Summary Statistics
 - Bug when using data_column
 - Velvet wrapper broken when real user jobs are used
 - apport.fileutils
 - Bug: Running functional tests for migrated or installed tools does not
- Issues from Bitbucket**
 - 5: Option to disable automatic history creation
 - 6: Option to require that histories have names
 - 8: More flexible output handlers
 - 10: Allow overriding parameters when running a workflow
 - 20: Suggestion: new tag in tool's XML file - 12/9/08 email from Assaf Gordon
 - 21: Real DB key build ontology
 - 24: Add ability to password secure tools

Members: Add Members...

Board: Options, Add List, Filter Cards

Activity: View all...

- Dannon Baker added API: Library Contents to Developer ideas and
 - sent to the board
 - joinedtoday at 10:39 am
- g2roboto on Feature request: manually hide datasets
 - Submitted by @nickstolerFeb 1 at 4:40 pm
- g2roboto added Feature request: manually hide datasets to Inbox. Feb 1 at 4:40 pm
- g2roboto on Reference

<http://bit.ly/gxytrello>

http://wiki.galaxyproject.org


Galaxy Wiki

DaveClements Settings Logout | Search:

Titles Text

FrontPage

Edit History Actions




Galaxy is an open, web-based platform for *accessible, reproducible, and transparent* computational biomedical research.

- **Accessible:** Users without programming experience can easily specify parameters and run tools and workflows.
- **Reproducible:** Galaxy captures information so that any user can repeat and understand a complete computational analysis.
- **Transparent:** Users share and publish analyses via the web and create Pages, interactive, web-based documents that describe a complete analysis.

This is the Galaxy Community Wiki. It describes all things Galaxy.

Use Galaxy


Galaxy's [public service web site](#) makes analysis tools, genomic data, tutorial demonstrations, persistent workspaces, and publication services available to any scientist. Extensive [user documentation](#) (applicable to any [public](#) or local Galaxy instance) is available on [this wiki](#) and [elsewhere](#).



Deploy Galaxy

Galaxy is open source for all organizations. Local Galaxy servers can be set up by [downloading and customizing](#) the Galaxy application.

- [Admin](#)
- [Cloud](#)




Community & Project

Galaxy has a large and active user community and many ways to [Get Involved](#).


- [Community](#)
- [News](#)
- [Events](#)
- [Support](#)
- [Galaxy Project](#)

Contribute

- **Users:** [Share](#) your histories, workflows, visualizations, data libraries, and [Galaxy Pages](#), enabling others to use and learn from them.
- **Deployers and Developers:** Contribute tool definitions to the Galaxy [Tool Shed](#) (making it easy for others to use those tools on their installations), and code to the core release.
- **Everyone:** [Get Involved!](#)




Topic voting now open!



Use Galaxy

[Project Server](#) ([Use it!](#))
[Other Servers](#) • [Learn](#)
[Share](#) • [Search](#)

Communication

[Support](#) • [News](#) 
[Events](#) • [Twitter](#)
[Mailing Lists](#) ([search](#))

Deploy Galaxy

[Get Galaxy](#) • [Cloud](#)
[Admin](#) • [Tool Config](#)
[Tool Shed](#) • [Search](#)

Contribute

[Tool Shed](#) • [Share](#)
[Issues & Requests](#)
[Support](#)

Galaxy Project

[Home](#) • [About](#)
[Community](#)
[Big Picture](#)

Events

News

Galaxy Event Horizon

Events with Galaxy-related content are listed here.

Also see the [Galaxy Events Google Calendar](#) for a listing of events and deadlines that are relevant to the Galaxy Community. This is also available as an [RSS feed](#).

If you know of any event that should be added to this page and/or to the Galaxy Event Calendar, please add it here or send it to outreach@galaxyproject.org.

Upcoming Events



Date	Topic/Event	Venue/Location
February 4	Introduction to Galaxy Boot Camp	UC Davis Bioinformatics Core Davis, California, United States
March 2-5	Accessible, Transparent and Reproducible Analysis With Galaxy, part of SW1: Application of NGS Platforms for Whole Transcriptome and Genome Analysis Galaxy for Core Facilities, part of "W6: Community Resource Solutions to Analyzing Large Genomic Data Sets"	ABRF 2013 Palm Springs, California, United States
March 26-28	RNA Technologies and Analysis Workshop	DOE JGI User Meeting
April 5-6	2013 GMod Meeting	Cambridge, United Kingdom, immediately prior to Biocuration 2013
April 7-10	GO Galaxy Workshop	Biocuration 2013, Cambridge, United Kingdom
April 9-11	Workshop: Integrated Research Data Management for Next Gen Sequencing Analysis Using Galaxy and Globus Online Software-as-a-Service Talk: Integrated Research Data management and Analysis in NGS using Globus Online, Galaxy and Amazon Web Services	BioIT World, Boston, Massachusetts, United States
May 14-16	Tutorial: Exploring and Enabling Biomedical Data Analysis with Galaxy	Great Lakes Bioinformatics Conference (GLBIO) 2013, Pittsburgh, Pennsylvania, United States
May 21-29	Initiation à l'utilisation de Galaxy Les deux ateliers sont maintenant complets	Cycle "Bioinformatique par la pratique" 2013, INRA Jouy-en-Josas, France
May 22-30	Analyse de données issues de séquenceurs nouvelle génération sous Galaxy Les deux ateliers sont maintenant complets	
June 6-7	Informatics on High Throughput Sequencing Data Workshop	Toronto, Ontario, Canada

News

Announcements of interest to the Galaxy Community. These can include items from the Galaxy Team or the Galaxy community and can address anything that is of wide interest to the community.

The Galaxy News is also available as an [RSS feed](#).

See [Add a News Item](#) below for how to get an item on this page, and the RSS feed. Older news items are available in the [Galaxy News Archive](#).

See also

- [Distribution News Briefs](#)
- [Galaxy Updates](#)
- [Galaxy on Twitter](#)
- [Events](#)
- [Learn](#)
- [Support](#)
- [About the Galaxy Project](#)

News Items

February 2013 Galaxy Update

The February 2013 Galaxy Update is now available.

Highlights:

- [Three new public Galaxy servers](#)
- [New papers](#)
- [Open Positions](#) at five different institutions
- [GCC2013 Training Day Topic voting, Registration, and Sponsorships](#)
- [January GalaxyAdmins Web Meetup](#) slides and screencast
- [Other Upcoming Events and Deadlines](#)
- [Galaxy Distributions](#)
- [Tool Shed Contributions](#)
- [Other News](#)

If you have anything you would like to see in the March [Galaxy Update](#), please let us know.

[Dave Clements](#) and the [Galaxy Team](#)

Posted to the [Galaxy News](#) on 2013-02-01

GCC2013 Training Day Topics: Vote!

A list of possible topics for the GCC2013 Training Day is now available. Please take a few minutes to [review these possibilities](#) and then [vote](#) for your favorite three topics.*

Your votes will determine not only the topics that are offered, but also which topics should be offered more than once, assigned to which rooms, and which ones should not be scheduled at the same time. Your vote matters.

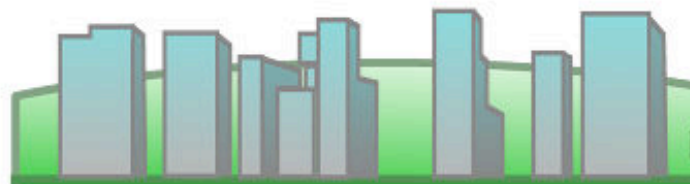
News Items

[February 2013 Galaxy Update](#)
[GCC2013 Training Day Topics: Vote!](#)
[Galaxy Project Openings](#)
[Jan 11, 2013 Distribution & News Brief](#)
[January 2013 GalaxyAdmins](#)
[January 2013 Galaxy Update](#)
[Dec 20, 2012 Distribution & News Brief](#)
[Galaxy Internships @ EMBL](#)
[Nominate GCC2013 Training Topics](#)
[Dec 3, 2012 Distribution & News Brief](#)
[December 2012 Galaxy Update](#)
[Nov 14, 2012 Distribution & News Brief](#)
[NGS Analysis by Viz. with Trackster](#)
[November 2012 GalaxyAdmins](#)

[News Archive](#)



Galaxy Community Conference



OSLO



30 June
- 2 July

2013

UiO : University of Oslo

Registration & abstract
submission are now open
<http://galaxyproject.org/GCC2013>

GCC2013
Training
Day





Enis Afgan



Dannon Baker



Dan Blankenberg



Dave Bouvier



Dave Clements



Nate Coraor



Carl Eberhard



Jeremy Goecks



Sam Guerler



Jen Jackson



Greg von Kuster



Ross Lazarus



Anton Nekrutenko



James Taylor



You

The Galaxy Team

<http://wiki.galaxyproject.org/GalaxyTeam>

RNA-seq Exercise: A Plan

- ...
- Trim as we see fit.
- Map the reads to the human reference using Tophat
- Run Cufflinks on Tophat output to assemble reads into transcripts
- *Imagine pages and pages of discussion on the intricacies and pitfalls of RNA-seq transcript prediction here.*

<http://bit.ly/gxynaseq>

RNA-seq Exercise: A Plan

- ...
- Map the reads to the human reference using Tophat
- Run Cufflinks on Tophat output to assemble reads into transcripts
 - *Imagine pages and pages of discussion on the intricacies and pitfalls of RNA-seq transcript prediction here.*
- Visualize it

<http://bit.ly/gxyrnaseq>

Visualizing Genomics

Supported external browsers

- UCSC
- Ensembl
- GBrowse
- IGB
- IGV

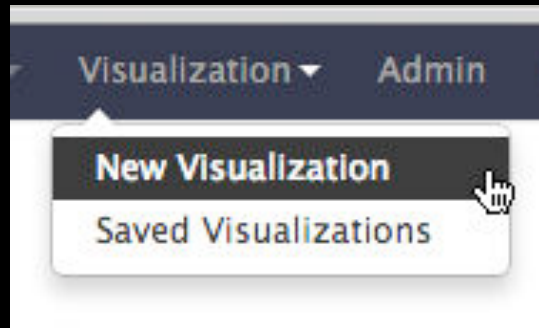
Traditional browser strengths:

- Showing what is nearby
- what else is happening here
- highlighting correlations
- integrating many datasets

But, *wouldn't it be nice to*

- Use visualization to **evaluate and refine analyses?**
- **Expose** some **basic analyses in visualization** to make it more informative?
- Make that **analyze-visualize-refine loop seamless and fast?** That is, integrate the two?
- Use visualization to **learn tools and explore their parameter space?**
- Not be tied to a **predefined reference genome?**

Create a visualization in Galaxy



or

A screenshot of the Galaxy web interface showing a visualization of Cufflinks transcripts. The visualization is titled '28: Brain: assembled transcripts from Cufflinks' and contains 211 lines of data. The format is gtf, and the database is hg19. The visualization shows a table of transcripts and exons, with a 'Visualize' button highlighted by a mouse cursor.

28: Brain: assembled transcripts from Cufflinks
211 lines
format: gtf, database: hg19
Info: cufflinks v2.0.2
cufflinks -q --no-update-check -l 300000 -F 0.100000 -j 0.150000 -p 4
display at IPSC main
display at Ensembl Current

1. Seqname	2. Source	3. Feature	4. Start
chr19	Cufflinks	transcript	33480
chr19	Cufflinks	exon	33480
chr19	Cufflinks	transcript	33490
chr19	Cufflinks	exon	33490
chr19	Cufflinks	transcript	33510
chr19	Cufflinks	exon	33510

Isn't it nice to

- To do all those things we talked about?
 - Use visualization to evaluate and refine analyses?
 - Expose some basic analyses in visualization to make it more informative?
 - Make that analyze-visualize-refine loop seamless and fast? That is, integrate the two?
 - Use visualization to learn tools and explore their parameter space?
 - Not be tied to a predefined reference genome?

Acknowledgements

Nalini Raghavachari
David Needleman
Jim Vincent

The Galaxy Team
especially
Dannon Baker

ABRF

AWS Education Grant

NIH NSF Huck Institute
Penn State University Emory University

Thanks

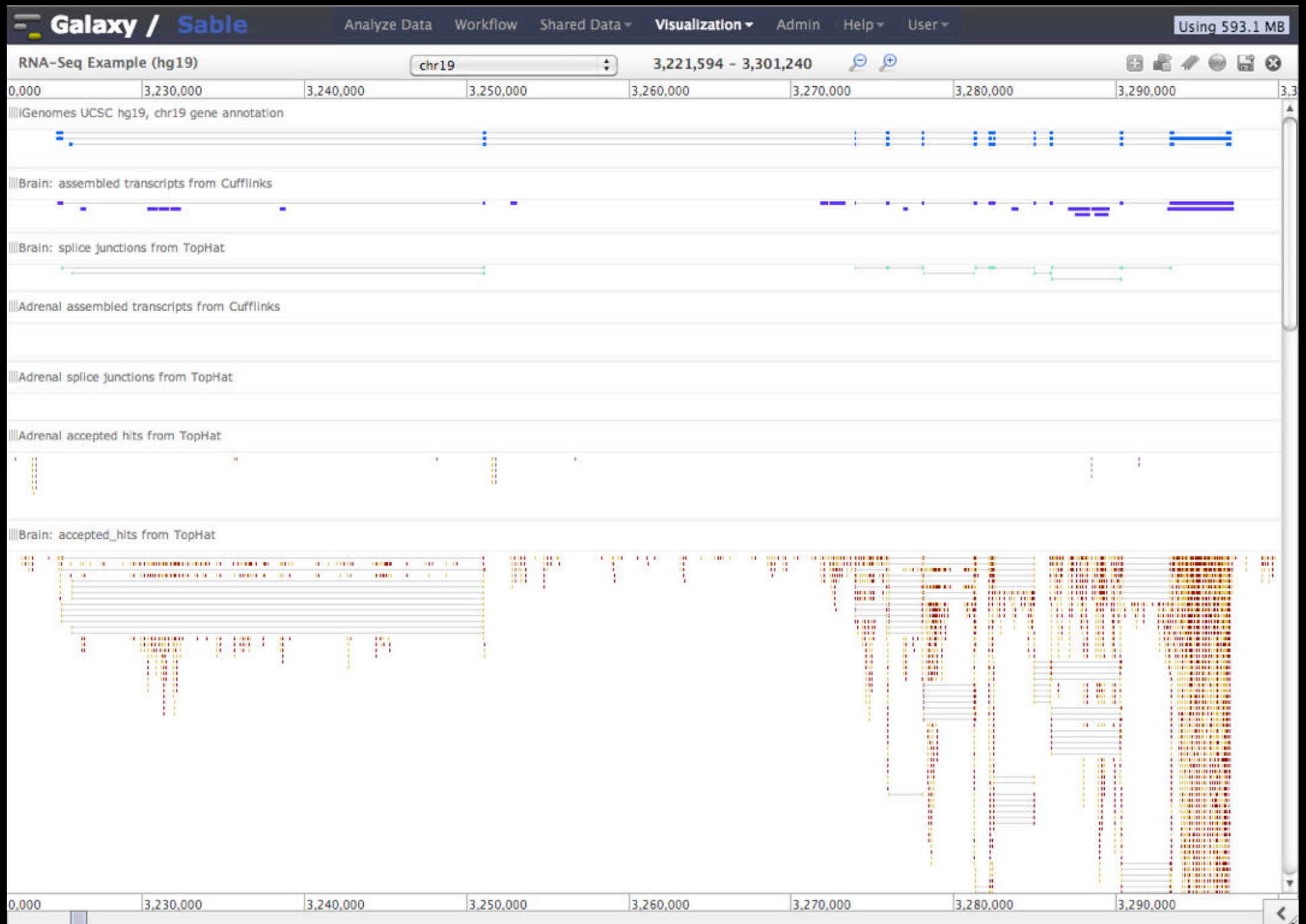


Dave Clements

Galaxy Project
Emory University

clements@galaxyproject.org

Trackster: Galaxy's embedded track browser



Some Galaxy Terminology

Dataset:

Any input, output or intermediate set of data + metadata

History:

A series of inputs, analysis steps, intermediate datasets, and outputs

Workflow:

A series of analysis steps

Can be repeated with different data

More Galaxy Terminology

Share:

Make something available to someone else

Publish:

Make something available to everyone

Galaxy Page:

Analysis documentation within Galaxy; easy to embed any Galaxy object

Let's all share...

Sharing for Galaxy Administrators Too

Data Libraries

Make data easy to find

Genome Builds

Care about a particular subset of life?

Galaxy Tool Shed

Wrapping tools and datatypes

Sharing & Publishing enables **Reproducibility**



Reproducibility: Everybody talks about it, but ...



Galaxy aims to push the goal of reproducibility from the bench to the bioinformatics realm

All analysis in Galaxy is recorded without any extra effort from the user.

Histories, workflows, visualizations and *pages* can be shared with others or published to the world.

Sharing & Publishing enables **Reproducibility**





Apply today for the Cancer GWAS Grant.

HOME | ABOUT | ARCHIVE | SUBMIT | SUBSCRIBE | ADVERTISE | AUTHOR INFO | CONTACT | HELP

Institution: PENN STATE UNIV Sign In via User Name/Password

Search for Keyword:
Advanced Search

Windshield splatter analysis with the Galaxy metagenomic pipeline

Sergei Kosakovsky Pond^{1,2,6,9}, Samir Wadhawan^{3,6,7},
Francesca Chiaromonte⁴, Guruprasad Ananda^{1,3}, Wen-Yu Chung^{1,3,8},
James T


OPEN ACCESS ARTICLE

This Article

Published in Advance October 9, 2009, doi: 10.1101/gr.094508.109
Copyright © 2009 by Cold Spring Harbor Laboratory Press

Current Issue

October 2010, 20 (10)



Footnotes

[Supplemental material is available online at <http://www.genome.org>. All data and tools described in this manuscript can be downloaded or used directly at <http://galaxyproject.org>. Exact analyses and workflows used in this paper are available at <http://usegalaxy.org/u/aun1/p/windshield-splatter>.]

<http://usegalaxy.org/u/aun1/p/windshield-splatter>