GO Galaxy

6th International Biocuration Conference

8 April 2013, Cambridge, UK

Suzanna Lewis, Chris Mungall, Seth Carbon
Lawrence Berkeley National Lab
http://www.berkeleybop.org/
Amelia Ireland, University of California Berkeley
http://gmod.org/
Dave Clements, Emory University
http://galaxyproject.org/

Agenda

15:00 Galaxy Overview &
Basic Analysis with Galaxy

16:00 Ontological Analysis with GO Galaxy

17:00 Done

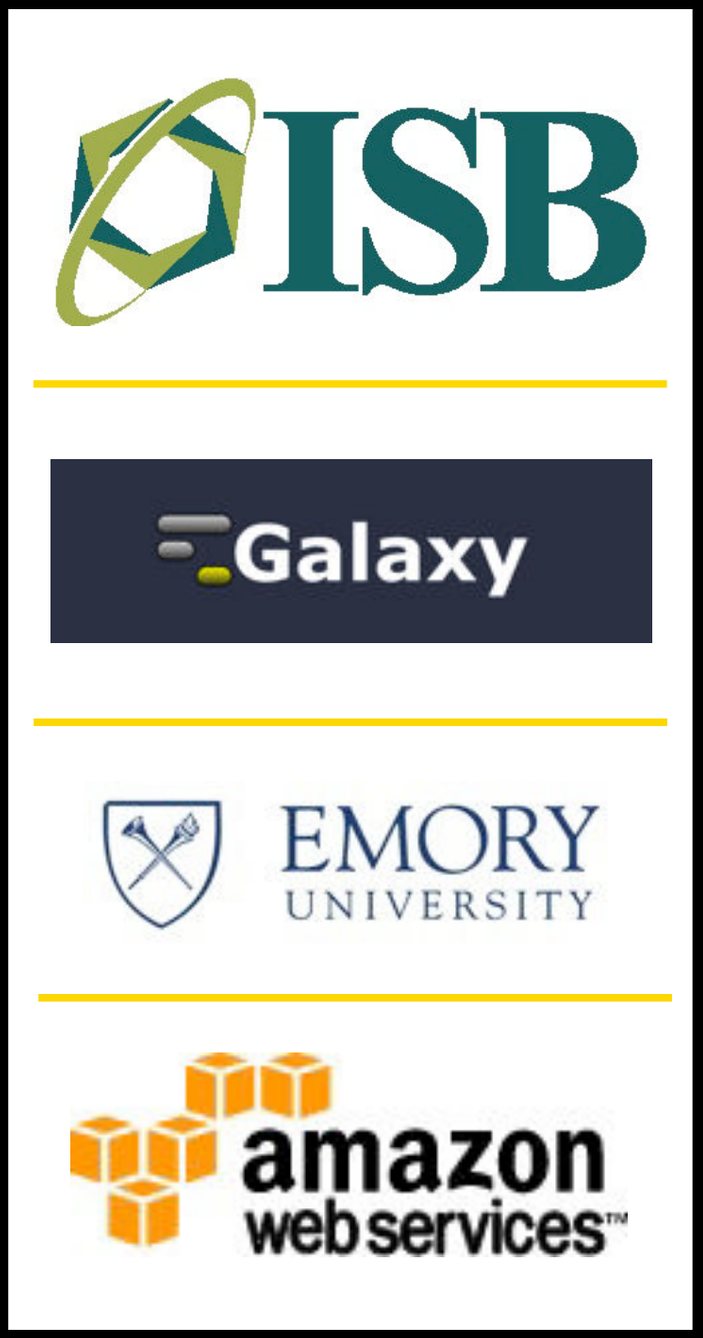
Introduction to Galaxy

6th International Biocuration Conference

8 April 2013, Cambridge, UK

Dave Clements, Emory University

http://galaxyproject.org/



Agenda

Basic Analysis with Galaxy

Basic Analysis into Reusable Workflows

Sharing: How to Play Well with Others

Galaxy Project Overview

Further reading

What is Galaxy?

What is Galaxy?

Let's answer that by demonstrating it

- Please follow along
- Please share the joy with your neighbors. Not all of you have laptops
- We will have wandering TA's to help; just flag them down.
- Given size of the audience, and the time limits the presenter won't wait for everyone to catch up.

What is Galaxy?

Let's answer that by demonstrating it

- Please follow along
- Please share the joy with your neighbors. Not all of you have laptops
- We will have wandering TA's to help; just flag them down.
- Given size of the audience, and the time limits the presenter won't wait for everyone to catch up.

"you'll definitely want to have a plan B in case the network explodes"

a Biocuration 2013 organizer

Basic Analysis

On human chromosome 22,
which coding exons have the most repeats in them?

cloud1.galaxyproject.org (I)
cloud2.galaxyproject.org (S)
cloud3.galaxyproject.org (B)

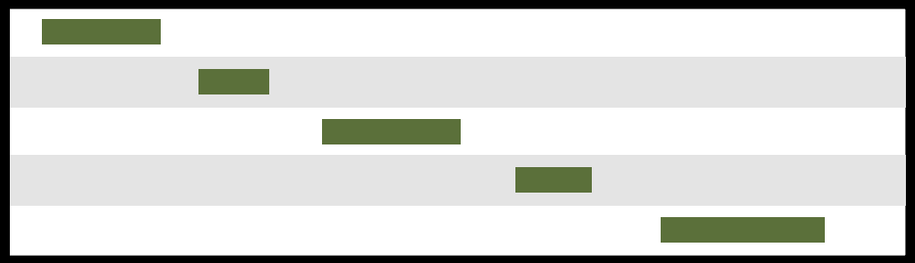
(~ http://usegalaxy.org/galaxy101)

Exons & Repeats: A General Plan

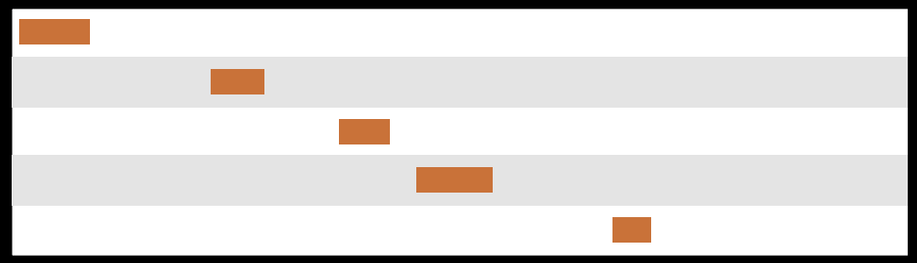
- Get some data
  - Coding exons on chromosome 22
  - Repeats on chromosome 22
- Mess with it
  - Identify which exons have repeats
  - Count repeats per exon
  - Save, download, ... exons with most repeats

cloud1.galaxyproject.org (I)
cloud2.galaxyproject.org (S)
cloud3.galaxyproject.org (B)

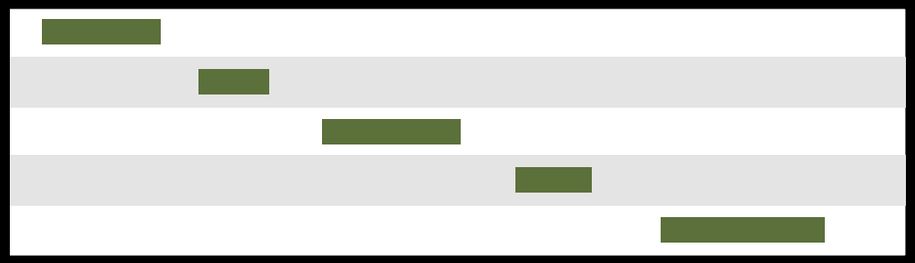
(~ http://usegalaxy.org/galaxy101)



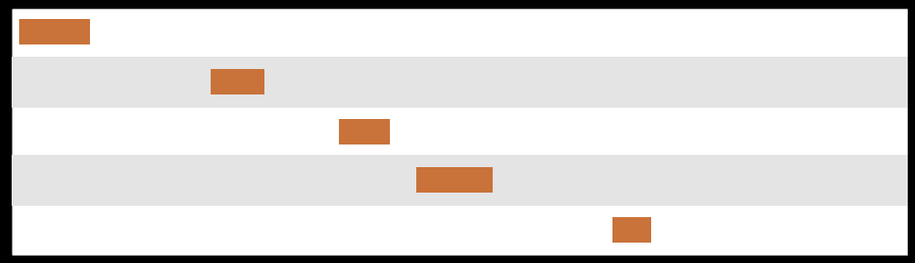
Exons, from UCSC



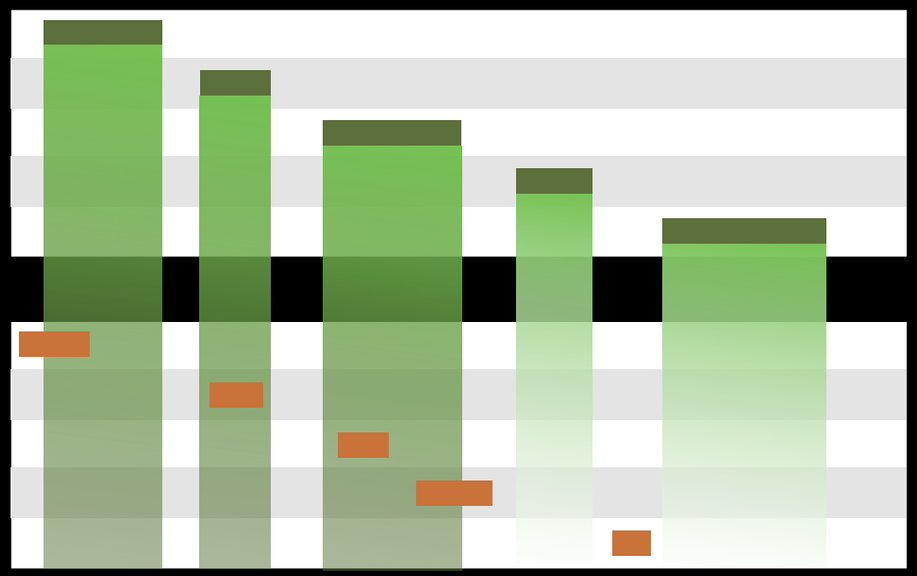
Repeats, from UCSC



Exons, from UCSC



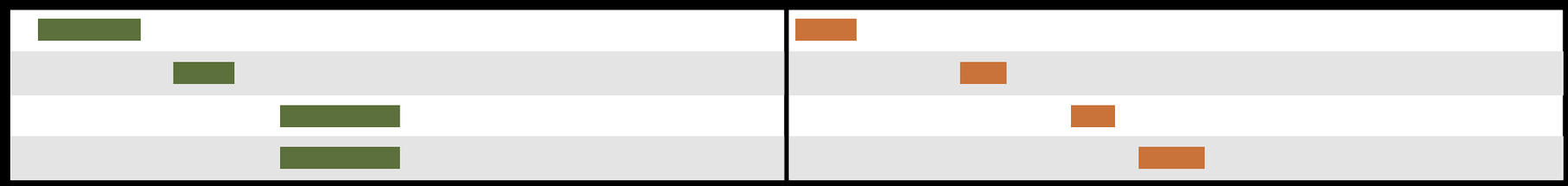
Repeats, from UCSC

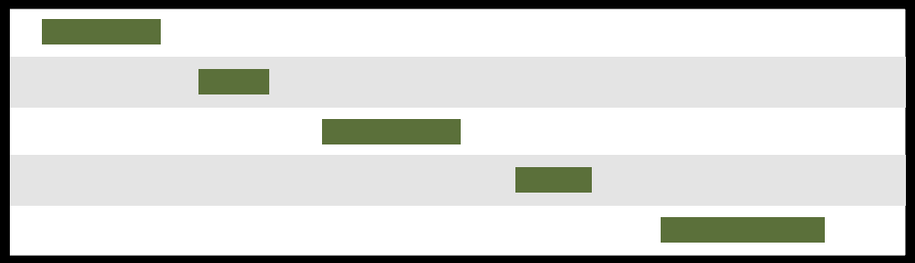


Exons, from UCSC

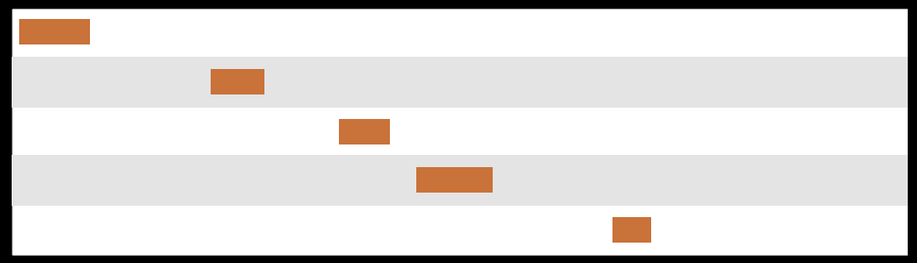
Repeats, from UCSC

Overlap pairings

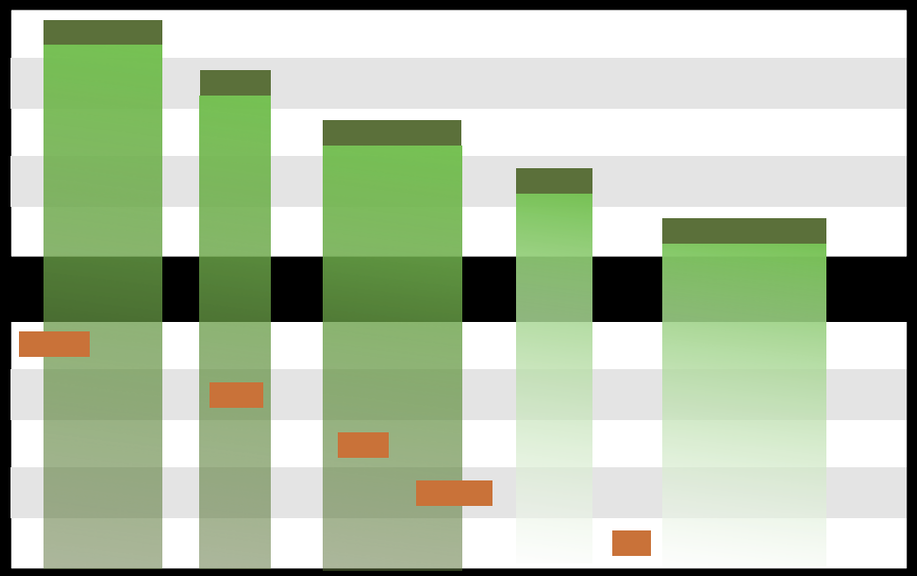




Exons, from UCSC



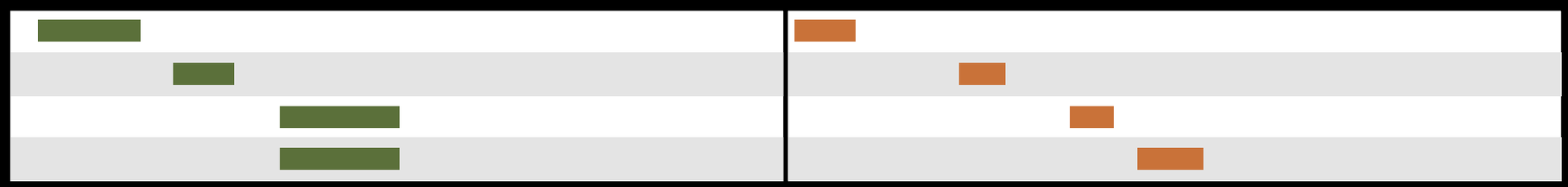
Repeats, from UCSC

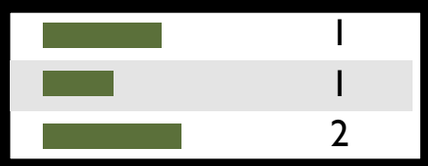


Exons, from UCSC

Repeats, from UCSC

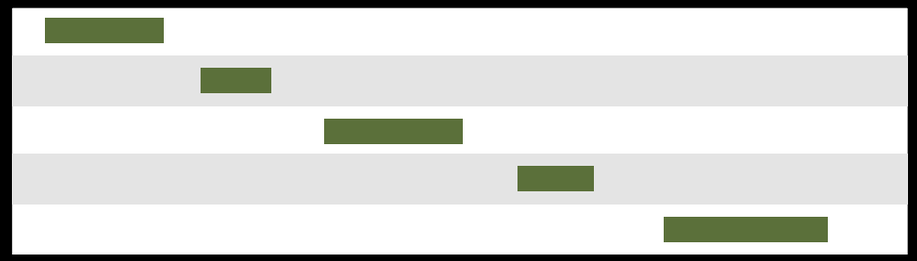
Overlap pairings





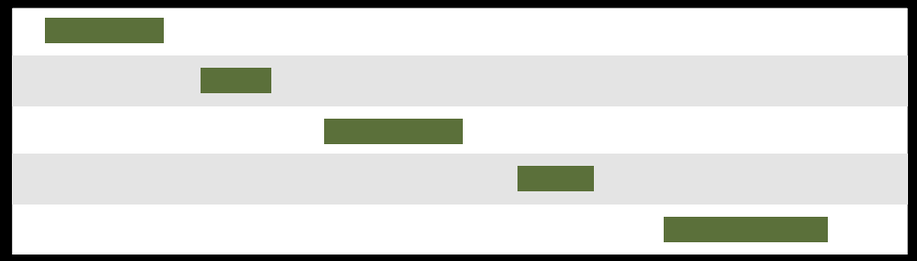
Exon overlap counts

Exon overlap counts



Exons, from UCSC

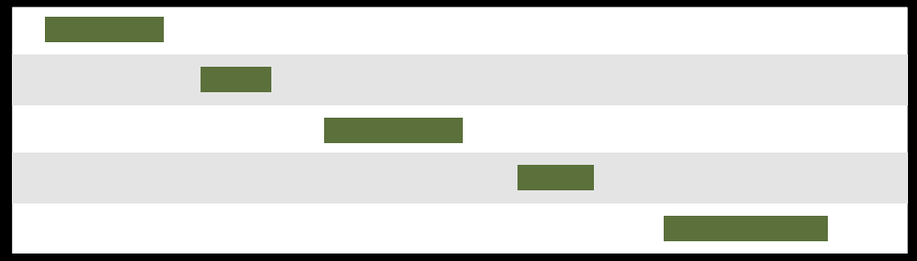
Exon overlap counts



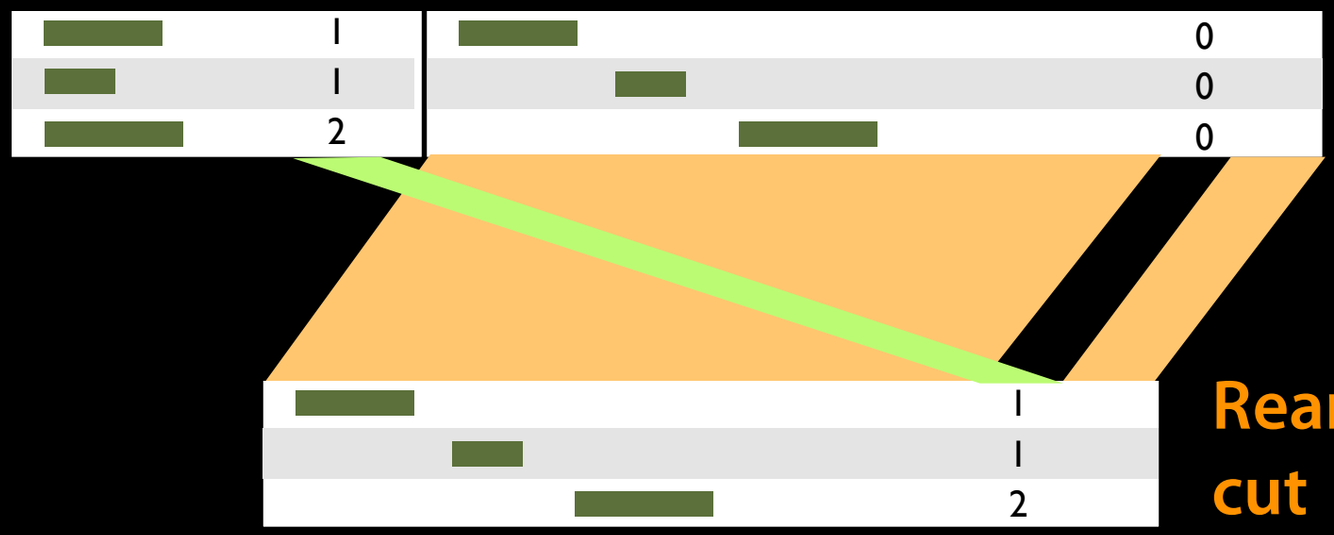
Exons, from UCSC

Join on exon name

Exon overlap counts



Exons, from UCSC



Join on exon name

Rearrange columns w/ cut

Agenda

Basic Analysis with Galaxy

Basic Analysis into Reusable Workflows

Sharing: How to Play Well with Others

Galaxy Project Overview

Further reading

Exons and Repeats *History* → Reusable *Workflow?*

- The analysis we just finished was about
  - Human chromosome 22
  - Overlap between exons and repeats
- But, ...
  - there is nothing inherently in the analysis about humans, chromosomes, exons or repeats
  - It is a series of steps that sets the score of one set of features to the number of overlaps from another set of features.

Create a generic *Overlap* Workflow

Extract Workflow from history

Create a workflow from this history.

Edit it to make some things clearer.

Run / test it

Tonight at a pub

Count # CpG islands in each exon

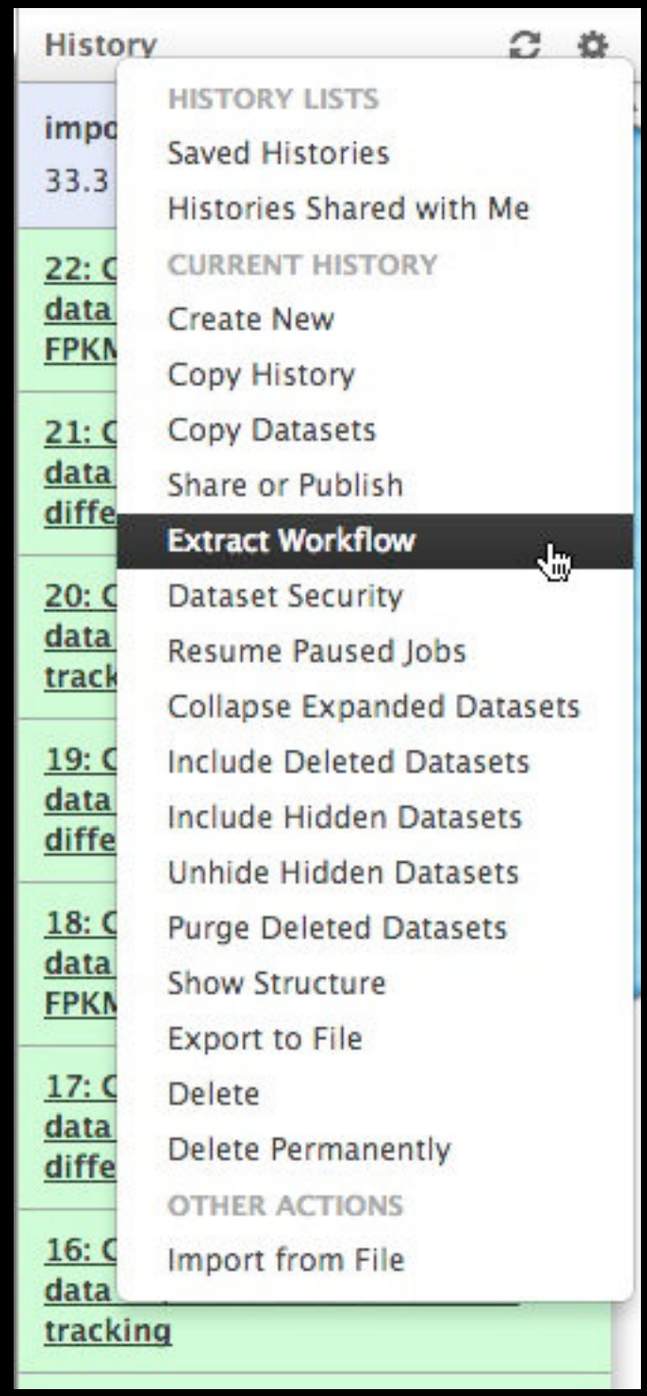
Did that work?

And

Count # of exons in each repeat

Did that work? *Why not?*

Edit workflow: doc assumptions



Agenda

Basic Analysis with Galaxy

Basic Analysis into Reusable Workflows

Sharing: How to Play Well with Others

Galaxy Project Overview

Further reading

Sharing & Publishing enables **Reproducibility**

~~Weather~~ Reproducibility: Everybody talks about it, but ...

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

All analysis and metadata 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

CSH PRESS GENOME RESEARCH

Windshield splatter analysis with the Galaxy metagenomic pipeline

Sergei Kosakovsky Pond[1,2,6,9], Samir Wadhawan[3,6,7], Francesca Chiaromonte[4], Guruprasad Ananda[1,3], Wen-Yu Chung[1,3,8], James Taylor[1,5,9], Anton Nekrutenko[1,3,9] and The Galaxy Team[1]

OPEN ACCESS ARTICLE

This Article

Published in Advance October 9, 2009, doi: 10.1101/gr.094508.109

» Abstract *Free*
» Full Text (PDF) *Free*

Current Issue
October 2010, 20 (10)

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

Sharing & Publishing enables Reproducibility

CSH PRESS GENOME RESEARCH

Windshield splatter analysis with the Galaxy metagenomic pipeline

Sergei Kosakovsky Pond[1,2,6,9], Samir Wadhawan[3,6,7],
Francesca Chiaromonte[4], 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

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

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 and sharing tools and datatypes

Agenda

Basic Analysis with Galaxy

Basic Analysis into Reusable Workflows

Sharing: How to Play Well with Others

Galaxy Project Overview

Further reading

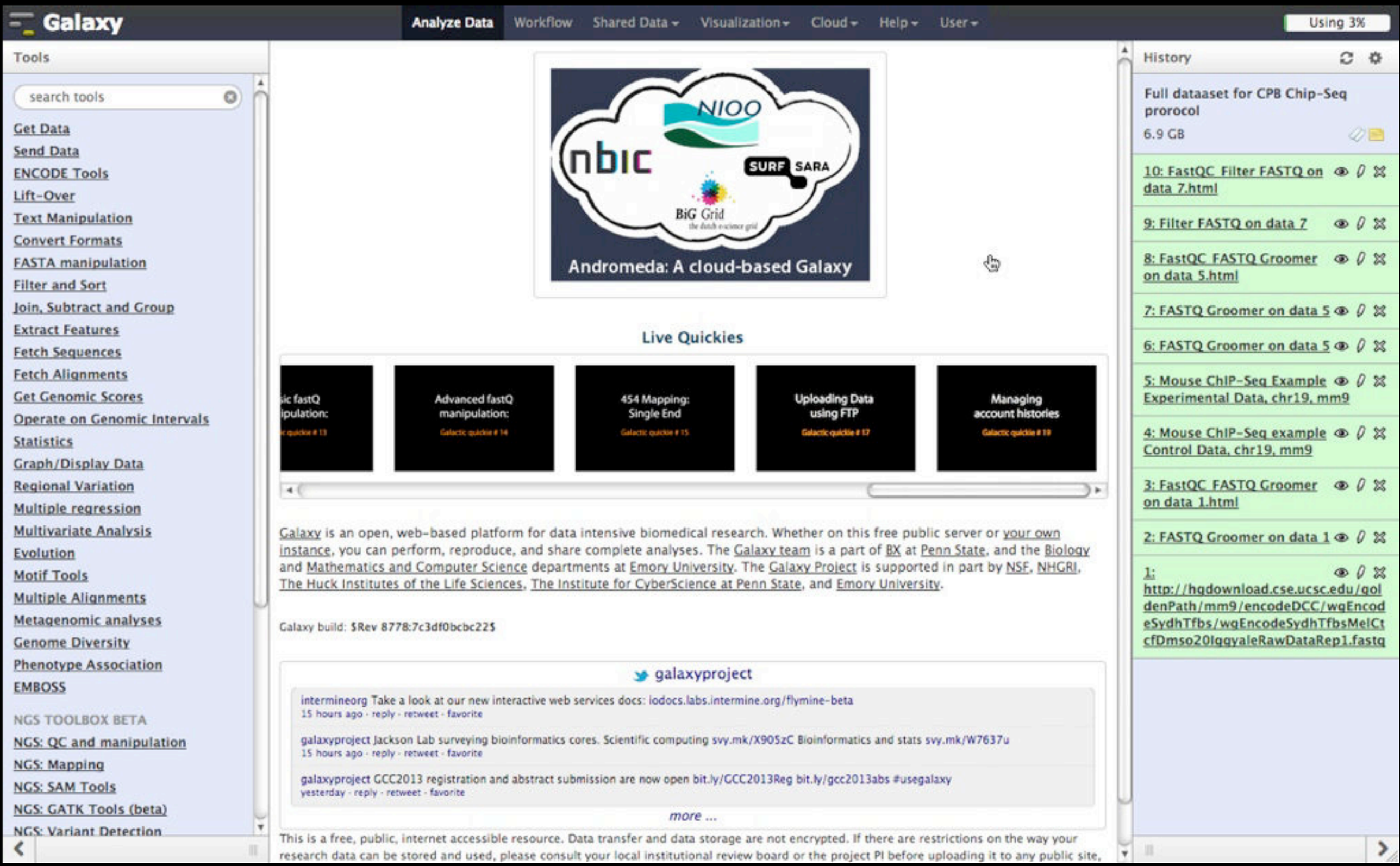
What is Galaxy, and how can I get some?

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

- Galaxy is available as:
  - **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
  - **Free cloud images** that can be deployed by informatics novices

http://galaxyproject.org

A free for everyone web-based service: usegalaxy.org



(See also: http://bit.ly/gxyServers)

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

Open Source Software: getgalaxy.org

- Galaxy is designed for local installation and customization
  - Easily integrate new tools
  - Requires a computational resource on which to be deployed

http://getgalaxy.org

Galaxy is available *on the cloud*

- Using this today
- Start with a **fully configured and populated** (tools and data) Galaxy instance.
- Allows you to scale up and down your compute assets as needed.
- Someone else manages the data center
- We are using this today

http://usegalaxy.org/cloud
http://aws.amazon.com/education

Galaxy Resources and Community: 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

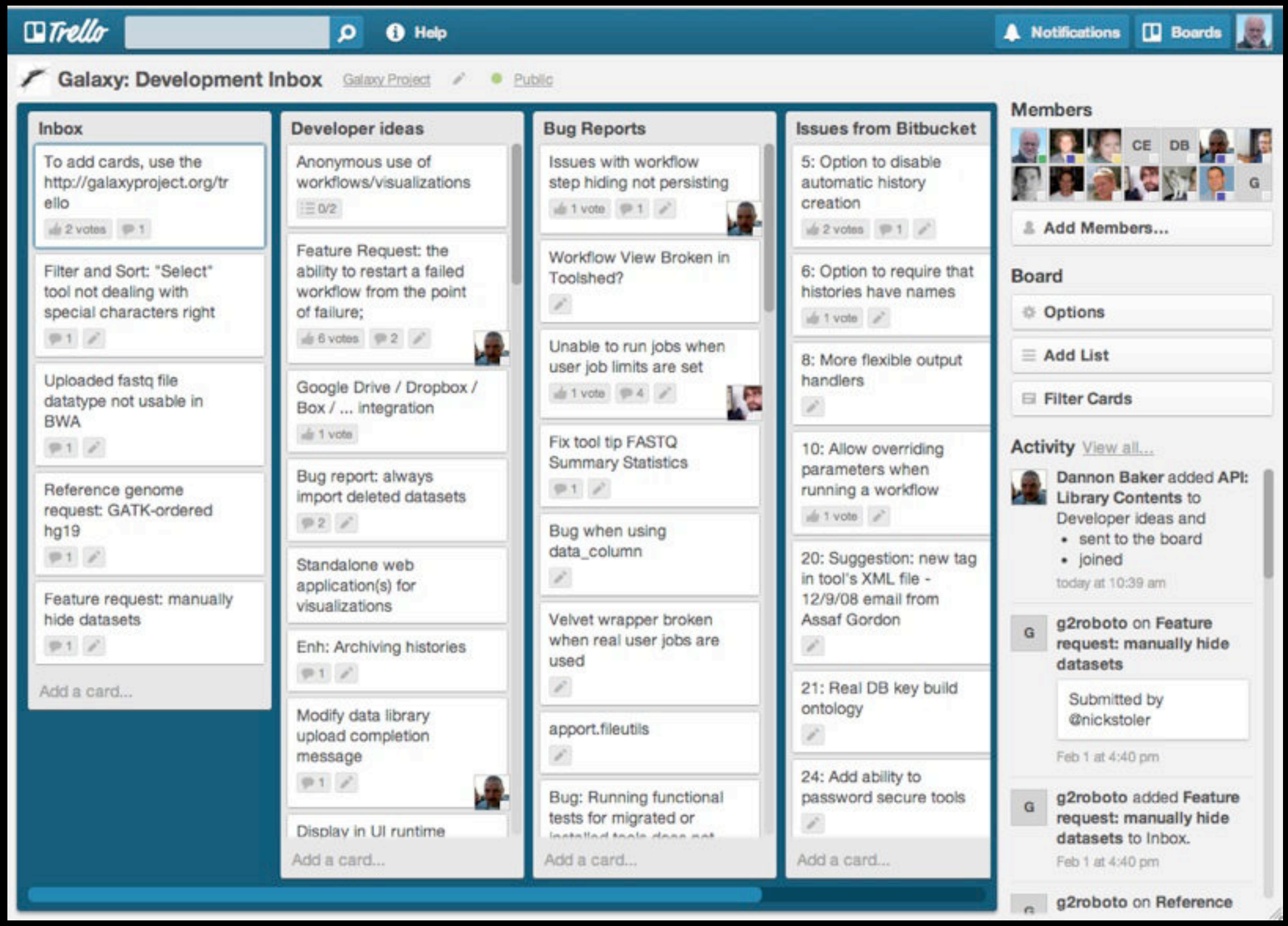
High volume (2900 posts, 2700 members in 2012)

Galaxy-Dev

Questions about developing for and deploying Galaxy

High volume (4500 posts, 850 members in 2012)

Community can create, vote and comment on issues



http://bit.ly/gxytrello

http://wiki.galaxyproject.org

Galaxy

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.

usegalaxy.org

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

getgalaxy.org

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!**

Galaxy Community Conference 30 June - 2 July 2013 OSLO

Talk abstracts due 12 April
Be published in *GigaScience*

Use Galaxy

Use Main *(about)*
Use Others! • 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 Wiki

DaveClements Settings Logout | Search:

Events

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 .

Cont
1.
2.
3.

Upcoming Events

ISB TGAC The Genome Analysis Centre™ Galaxy Community Conference 30 June – 2 July 2013 OSLO

Date	Topic/Event	Venue/Location
April 7-10	*GO Galaxy Workshop*	Biocuration 2013, Cambridge, United Kingdom
April 7-8	BOSC/Broad Interoperability Hackathon	Cambridge, Massachusetts, United States
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
April 10	*Introduction to Galaxy Boot Camp* Registration is full	UC Davis Bioinformatics Core Davis, California, United States
April 11	*Introduction to RNASeq Boot Camp* Registration is full	
April 11	*Introduction to Galaxy Workshop*	The Genome Analysis Centre, Norwich, United Kingdom
April 12	*Next generation sequencing data interpretation: enhancing reproducibility and accessibility*	Reed College, Portland, Oregon, 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 16-17	*Galaxy Workflows for Bioinformatics Analysis*, and *Workshop 1A – Galaxy Workflows for Bioinformatics Analysis*	Workshop in Next-Generation Sequence Analysis and Metabolomics (WiNGS), UNC-Charlotte, North Carolina, United States
May 21 May 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	Analyse de données issues de séquenceurs nouvelle génération sous Galaxy	

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
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

News Items

February 2013 Galaxy Update

The February 2013 Galaxy Update is now available.

Galaxy UPDATE

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.*

GCC2013 Training Day

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.

galaxyproject.org/GCC2013

Galaxy Community Conference

STARTING @ €95

(GIGA)n SCIENCE

GCC2013 Training Day

galaxyproject.org/GCC2013

Galaxy Community Conference

STARTING @ €95

Talk abstracts due **12 April**

(GIGA)n SCIENCE

GCC2013 Training Day

Agenda

Basic Analysis with Galaxy

Basic Analysis into Reusable Workflows

Sharing: How to Play Well with Others

Galaxy Project Overview

Further reading

Further Reading

Learn Galaxy Hub Page

http://wiki.galaxyproject.org/Learn

Worked RNA-Seq Example

http://bit.ly/GxyRNASeqEx

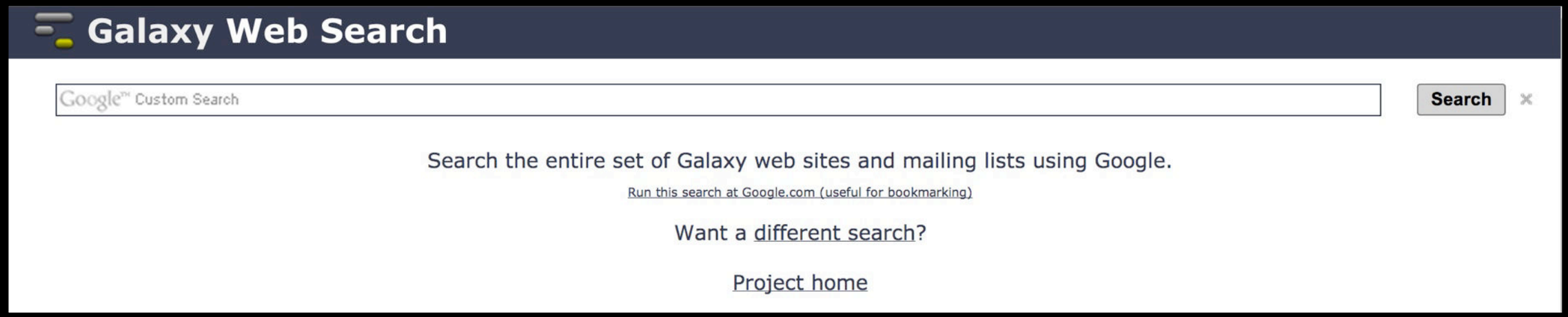
Public Galaxy Server List

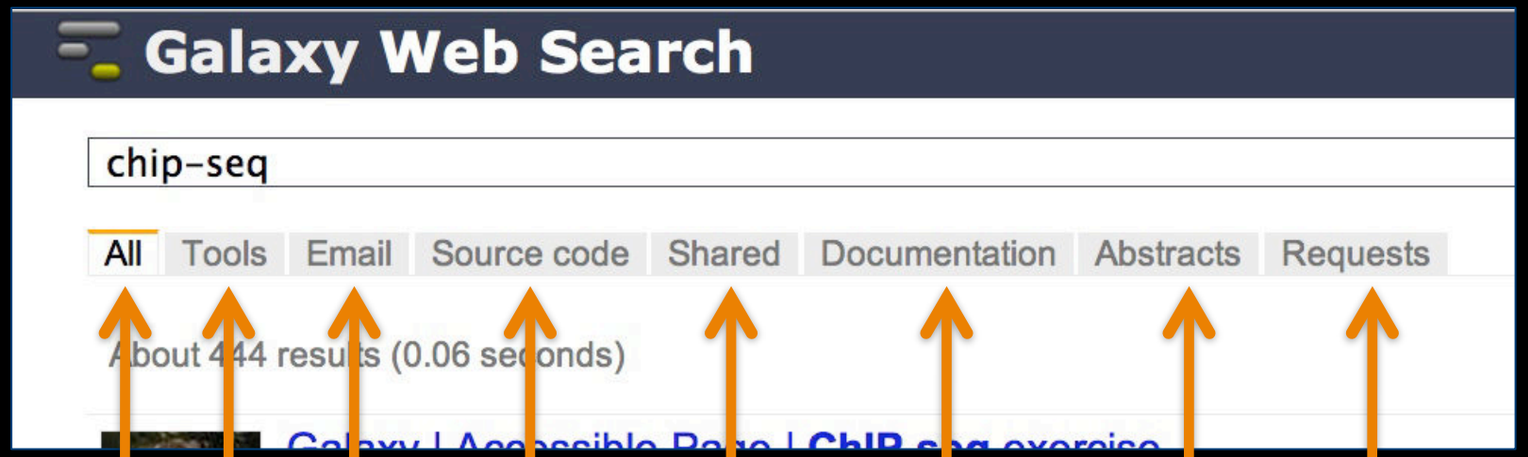
http://bit.ly/gxyServers

Community Wiki

http://wiki.galaxyproject.org

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





Find

Everything on …

Tools for …

Email about …

Source code for …

Published Histories, Pages, Workflows, about …

Documentation on …

Papers using Galaxy for …

Related feature requests

Acknowledgements

Francis Ouellette

Alex Bateman

Claire O'Donovan

Amy Cottage

Jennifer Harrow

Biocuration Organisers

Suzanna Lewis

Chris Mungall

Seth Carbon

Amelia Ireland

AWS Education Grant

NIH NSF Huck Institute

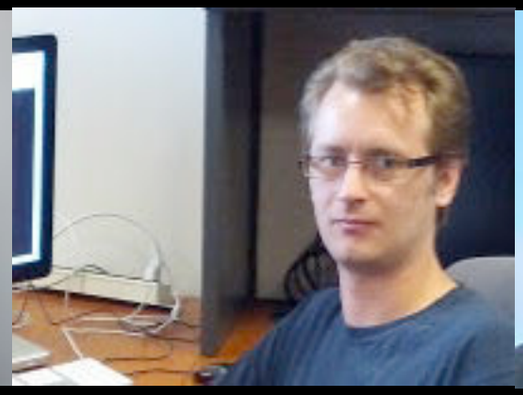
Penn State University Emory University

The Galaxy Team

Enis Afgan

Dannon Baker

Dan Blankenberg



Dave Bouvier

Dave Clements

Nate Coraor

Carl Eberhard

Dorine Francheteau

Jeremy Goecks

Sam Guerler

Jen Jackson

Greg von Kuster

Ross Lazarus

Anton Nekrutenko

James Taylor

http://wiki.galaxyproject.org/GalaxyTeam

Feedback Please!

http://bit.ly/GOISB2013

Thanks

Dave Clements

Galaxy Project
Emory University

clements@galaxyproject.org

http://bit.ly/GOISB2013