Accessible, Transparent, Reproducible Research with Galaxy

The Genome Analysis Centre (TGAC)
Norwich, UK
12 April 2013

Dave Clements, Emory University
http://galaxyproject.org/

As science becomes increasingly dependent on computation:

- How best to ensure that analysis are **reproducible**?
- How can methods best be made **accessible** to scientists?
- How to facilitate **transparent** communication of analyses?

A crisis in genomics research:

reproducibility

Key Reproducibility Problems

- **Datasets**: not all available, difficult to access
- **Tools**: inaccessible, hard to record details
- **Publication**: results, data, methods separate

50 papers citing bwa

31 provide no version and no settings

8 lists versions

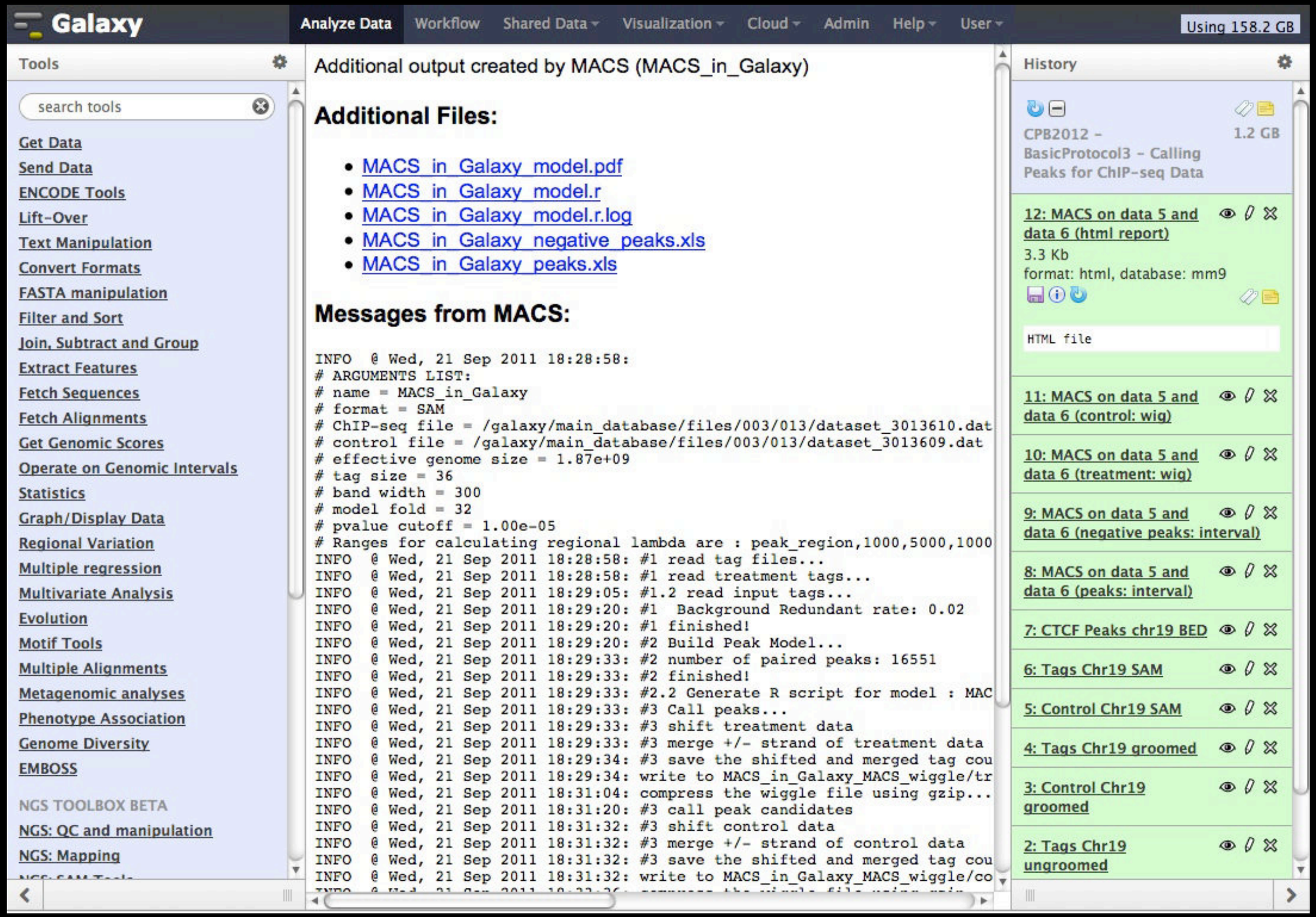
4 list settings

7 lists versions and settings

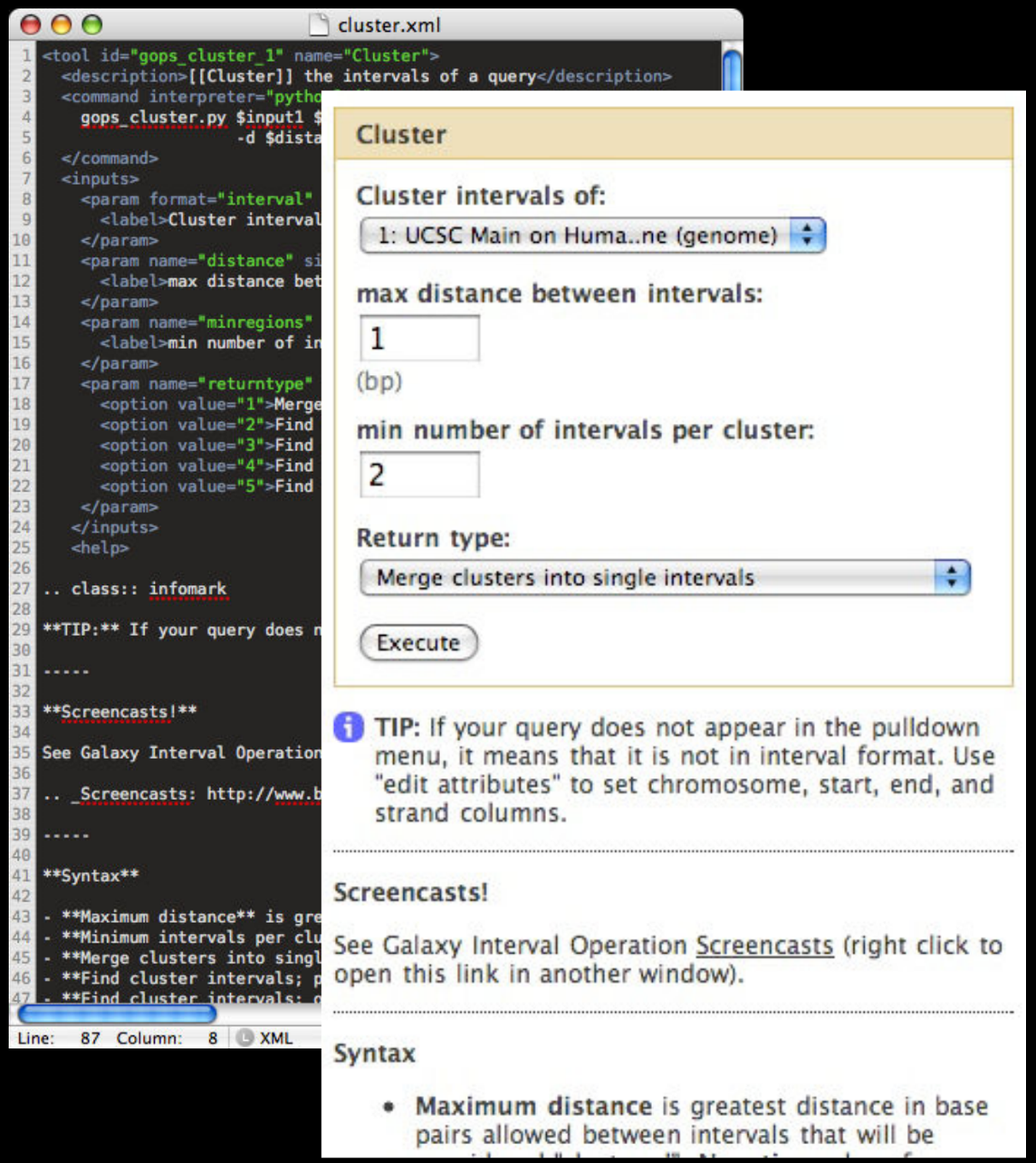
26 do not provide access to data

Nekrutenko & Taylor, "Next-generation sequencing data interpretation: enhancing reproducibility and accessibility" Nature Reviews Genetics 13, 667-672 (September 2012) doi:10.1038/nrg3305

Galaxy: an accessible analysis system

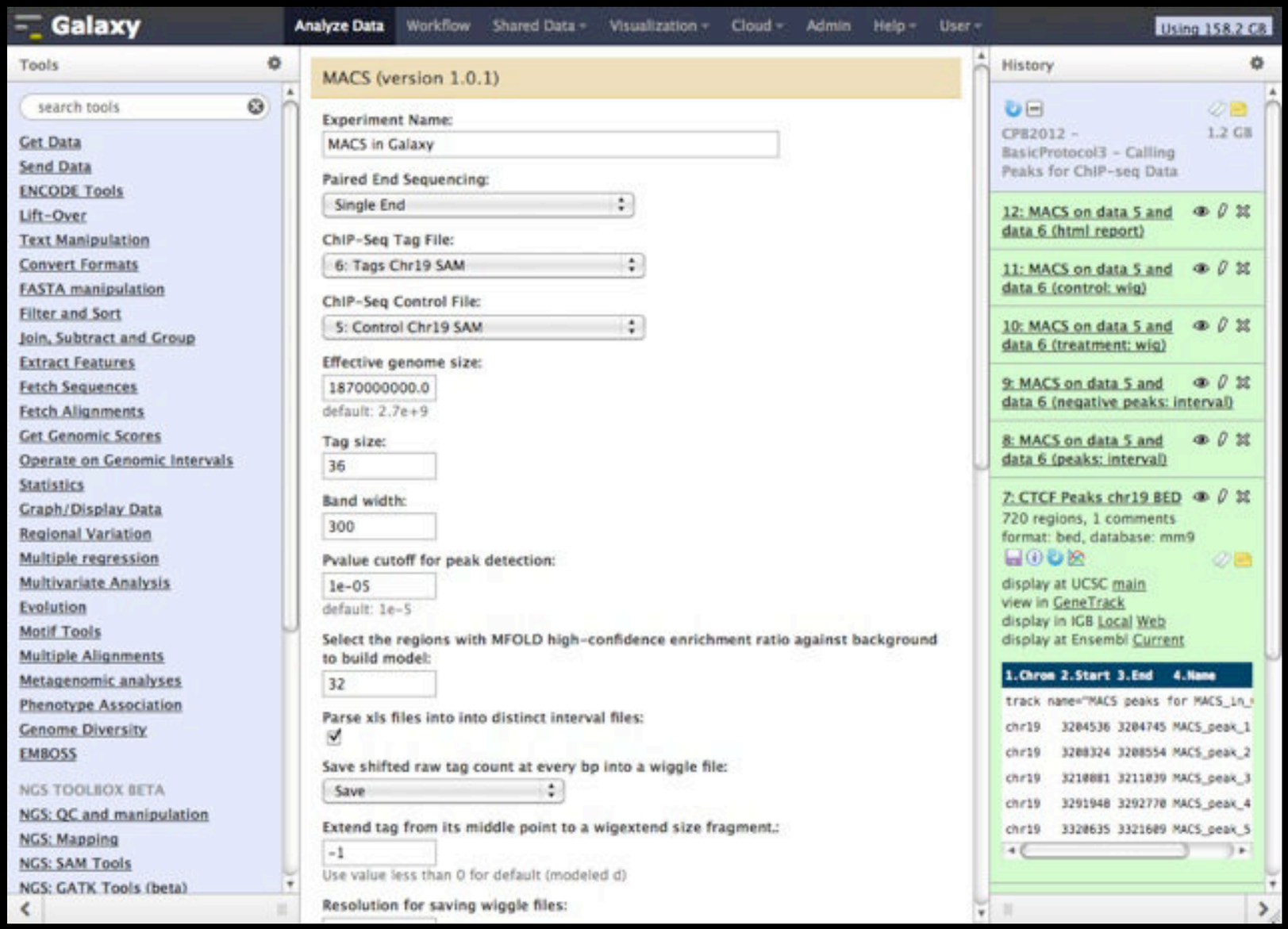


Integrating existing tools into a uniform framework



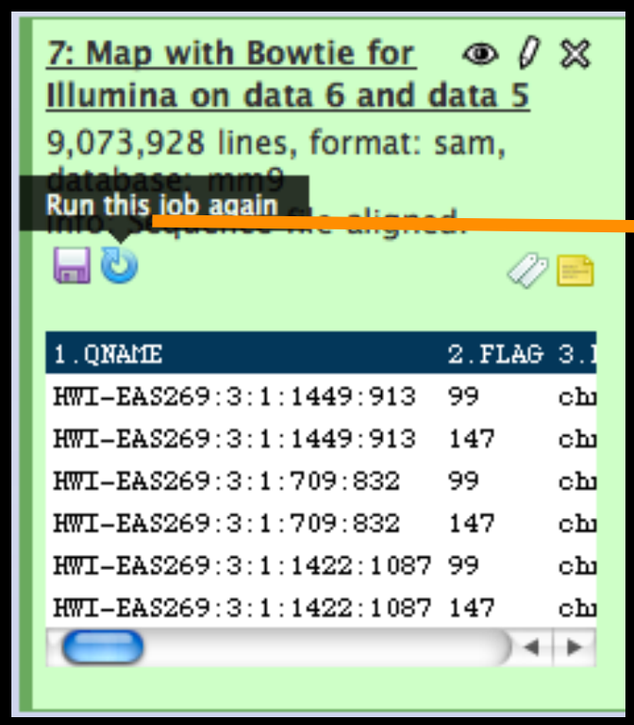
- Defined in terms of an abstract interface (inputs and outputs)
- Designed to be as easy as possible for tool authors, while still allowing rigorous reasoning

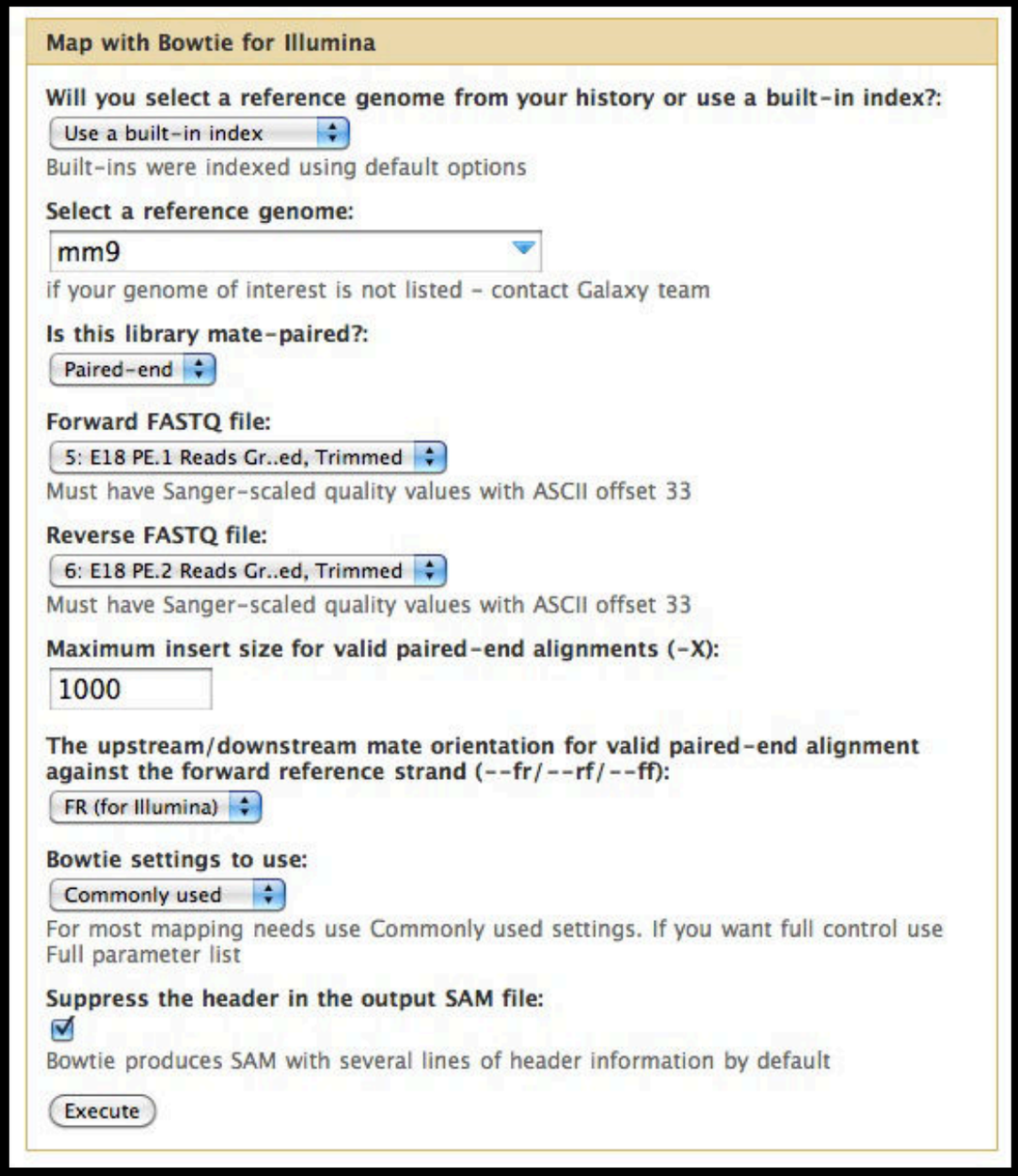
Galaxy analysis interface



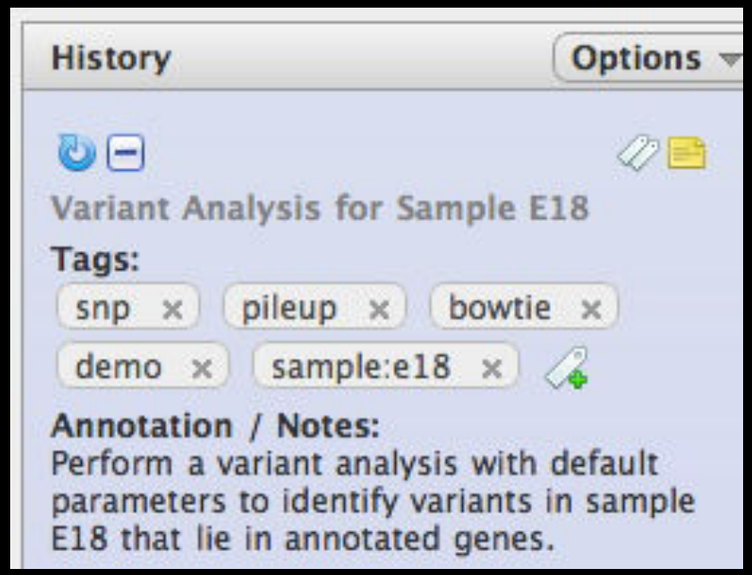
- Consistent tool user interfaces that are automatically generated
- History system facilitates and tracks multistep analyses
- Exact parameters of a step can always be inspected, and easily rerun

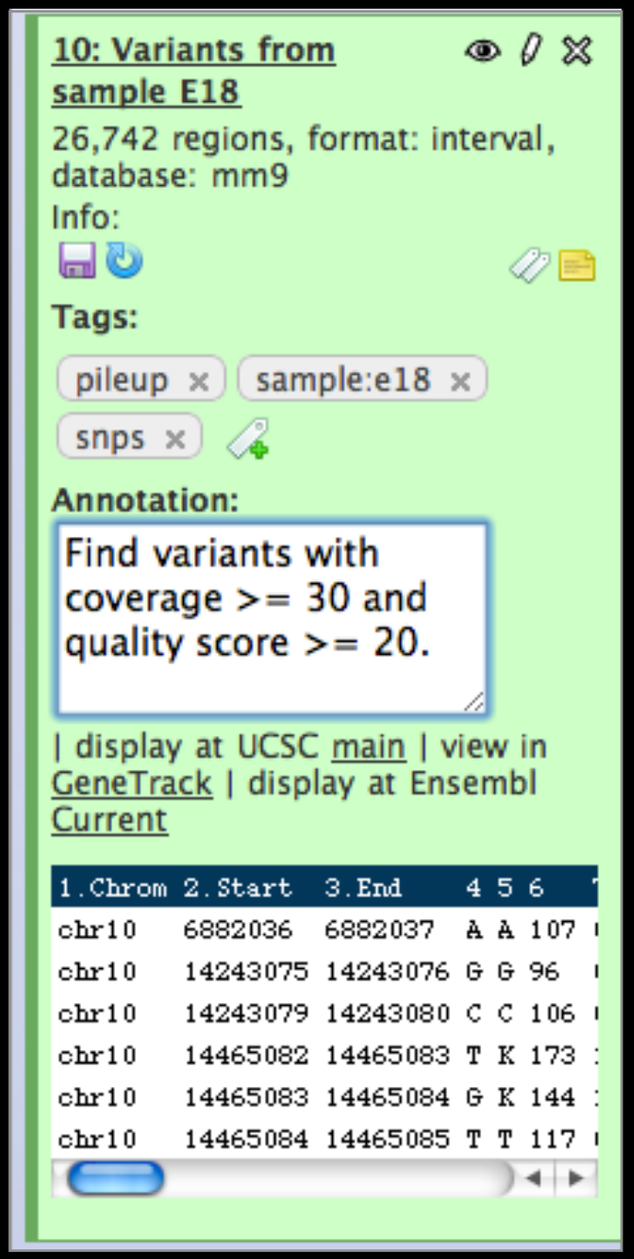
Automatically tracks every step of every analysis



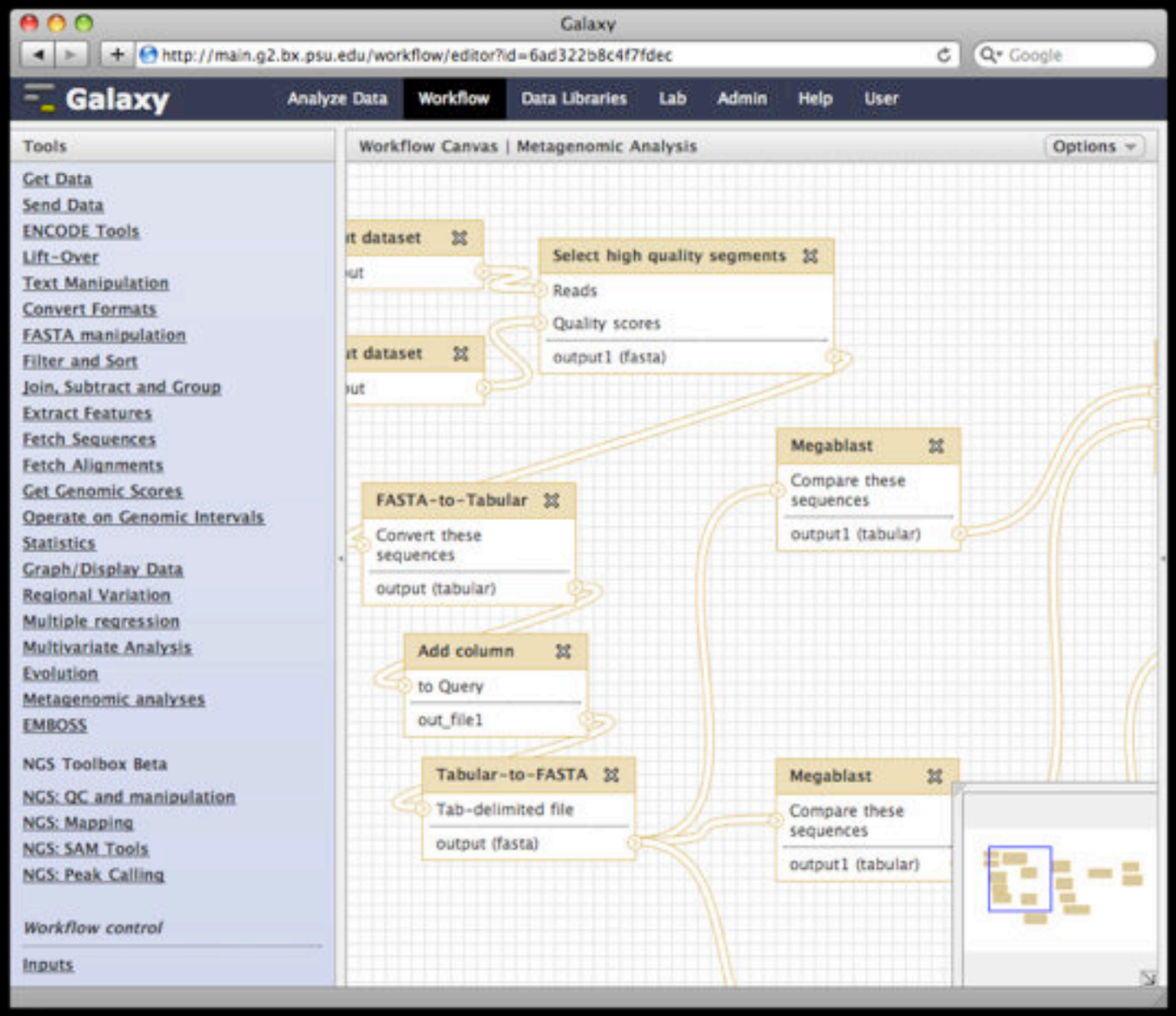


As well as user-generated metadata and annotation...



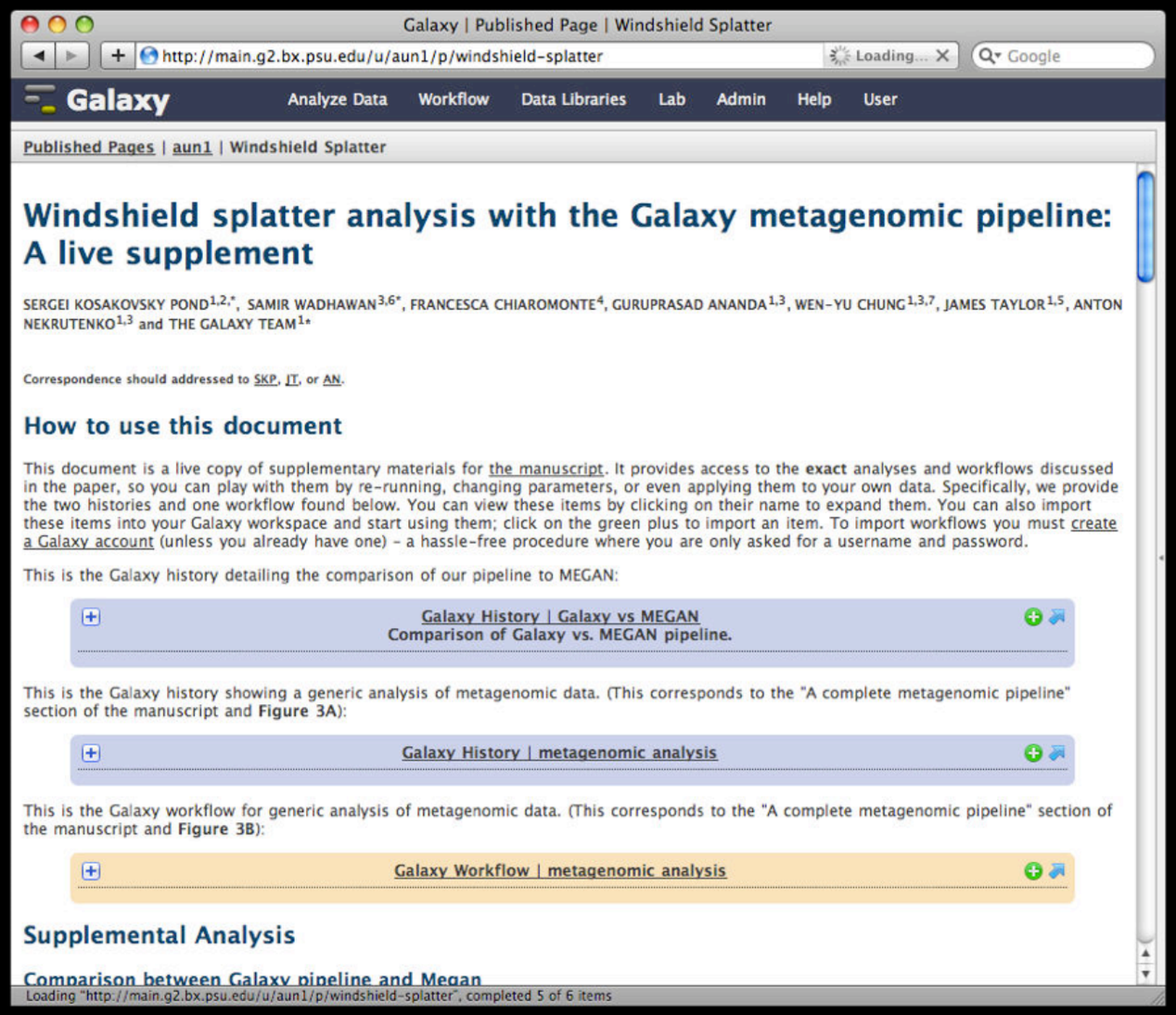


Galaxy workflow system



- Workflows can be constructed from scratch *or* extracted from existing analysis histories
- Facilitate reuse, as well as providing precise reproducibility of a complex analysis

Tranparency: Sharing and publishing



- All analysis components (datasets, histories, workflows) can be *shared* among Galaxy users and *published*
- Annotation and Galaxy Pages allow analyses to be augmented with textual content and provided in the form of an integrated document

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]

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

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.]

Windshield splatter analysis with the Galaxy metagenomic pipeline: A live supplement

SERGEI KOSAKOVSKY POND[1,2,*], SAMIR WADHAWAN[3,6*], FRANCESCA CHIAROMONTE[4], GURUPRASAD ANANDA[1,3], WEN-YU CHUNG[1,3,7], JAMES TAYLOR[1,5], ANTON NEKRUTENKO[1,3] and THE GALAXY TEAM[1*]

Correspondence should addressed to SKP, JT, or AN.

How to use this document

This document is a live copy of supplementary materials for the manuscript. It provides access to the **exact** analyses and workflows discussed in the paper, so you can play with them by re-running, changing parameters, or even applying them to your own data. Specifically, we provide the two histories and one workflow found below. You can view these items by clicking on their name to expand them. You can also import these items into your Galaxy workspace and start using them; click on the green plus to import an item. To import workflows you must create a Galaxy account (unless you already have one) – a hassle-free procedure where you are only asked for a username and password.

This is the Galaxy history detailing the comparison of our pipeline to MEGAN:

Galaxy History | Galaxy vs MEGAN
Comparison of Galaxy vs. MEGAN pipeline.

This is the Galaxy history showing a generic analysis of metagenomic data. (This corresponds to the "A complete metagenomic pipeline" section of the manuscript and **Figure 3A**):

Galaxy History | metagenomic analysis

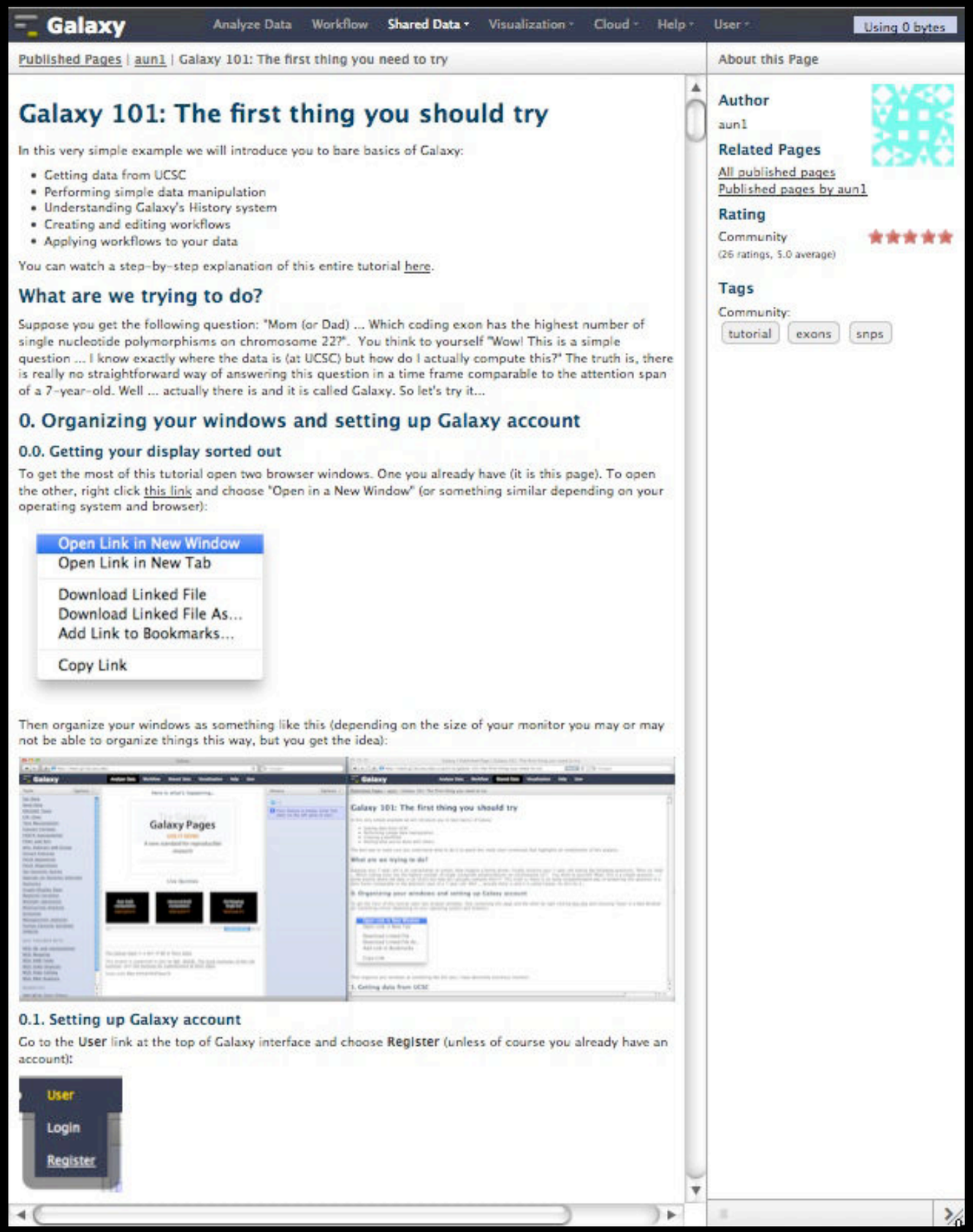
This is the Galaxy workflow for generic analysis of metagenomic data. (This corresponds to the "A complete metagenomic pipeline" section of the manuscript and **Figure 3B**):

Galaxy Workflow | metagenomic analysis

Supplemental Analysis

Comparison between Galaxy pipeline and Megan

Give it a spin: usegalaxy.org/galaxy101



Galaxy 101, is a hands-on exercise that demonstrates many Galaxy basics.

Galaxy 101 includes histories, datasets, and workflows, and is itself a *Galaxy Page*.

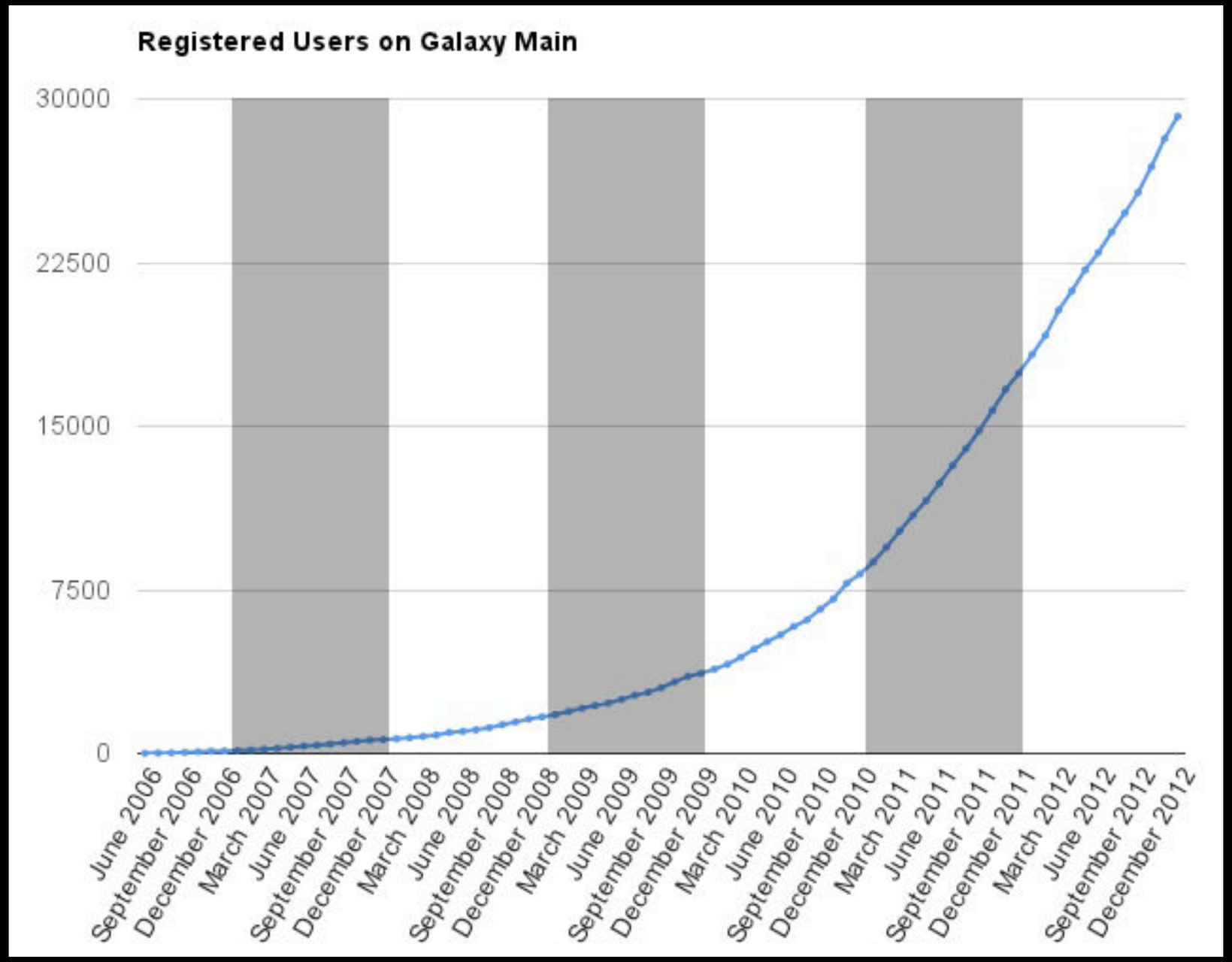
Galaxy is available ...

- **As a free (for everyone) web service** integrating a wealth of tools, compute resources, terabytes of reference data and permanent storage

http://usegalaxy.org

http://usegalaxy.org
(a.k.a Main)

- Free public web site
- Anybody can use it
- Persistent
- 150,000+ jobs / month
- Hundreds of tools



http://bit.ly/gxystats

usegalaxy.org: a wealth of tools

NGS: QC and manipulation

ILLUMINA DATA

- FASTQ Groomer convert between various FASTQ qual formats
- FASTQ splitter on joined pair end reads
- FASTQ joiner on paired end reads
- FASTQ Summary Statistics by column

ROCHE-454 DATA

- Build base quality distributio
- Select high quality segments
- Combine FASTA and QUAL ir FASTQ

AB-SOLID DATA

- Convert SOLiD output to fast
- Compute quality statistics fo SOLiD data
- Draw quality score boxplot f SOLiD data

GENERIC FASTQ MANIPULATION

- Filter FASTQ reads by quality score and length
- FASTQ Trimmer by column
- FASTQ Quality Trimmer by sliding window
- FASTQ Masker by quality sco
- Manipulate FASTQ reads on various attributes
- FASTQ to FASTA converter
- FASTQ to Tabular converter
- Tabular to FASTQ converter

FASTX-TOOLKIT FOR FASTQ DATA

- Quality format converter (ASCII Numeric)
- Compute quality statistics
- Draw quality score boxplot
- Draw nucleotides distribution chart
- FASTQ to FASTA converter
- Filter by quality
- Remove sequencing artifacts
- Barcode Splitter
- Clip adapter sequences
- Collapse sequences
- Rename sequences
- Reverse-Complement
- Trim sequences

FASTQ QC

- FastQC:Read QC reports using FastQC

NGS: Mapping

ILLUMINA

- Map with Bowtie for Illumina
- Map with BWA for Illumina

ROCHE-454

- Lastz map short reads against reference sequence
- Megablast compare short reads against htgs, nt, and wgs databases
- Parse blast XML output

AB-SOLID

- Map with Bowtie for SOLiD
- Map with BWA for SOLiD

NGS: SAM Tools

- Filter SAM on bitwise flag values
- Convert SAM to interval
- SAM-to-BAM converts SAM format to BAM format
- BAM-to-SAM converts BAM format to SAM format
- Merge BAM Files merges BAM files together
- Generate pileup from BAM dataset
- Filter pileup on coverage and SNPs
- Pileup-to-Interval condenses pileup format into ranges of bases
- flagstat provides simple stats on BAM files
- rmdup remove PCR duplicates
- MPileup SNP and indel caller
- Slice BAM by provided regions

NGS: GATK Tools (beta)

ALIGNMENT UTILITIES

- Depth of Coverage on BAM files
- Print Reads from BAM files

REALIGNMENT

- Realigner Target Creator for use in local realignment
- Indel Realigner - perform local realignment

BASE RECALIBRATION

- Count Covariates on BAM files
- Table Recalibration on BAM files
- Analyze Covariates - draw plots

GENOTYPING

- Unified Genotyper SNP and inde caller

ANNOTATION

- Variant Annotator

FILTRATION

- Variant Filtration on VCF files
- Select Variants from VCF files

VARIANT QUALITY SCORE RECALIBRATION

- Variant Recalibrator
- Apply Variant Recalibration

VARIANT UTILITIES

- Validate Variants
- Eval Variants
- Combine Variants

NGS: Indel Analysis

- Filter Indels for SAM
- Extract indels from SAM
- Indel Analysis

NGS: Peak Calling

- MACS Model-based Analysis of ChIP-Seq
- SICER Statistical approach for the Identification of ChIP-Enriched Regions
- GeneTrack indexer on a BED file
- Peak predictor on GeneTrack index

NGS: RNA Analysis

RNA-SEQ

- 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
- Cuffdiff find significant changes

For example, the first 5 pages of NGS tools

But, it's a big world

Main has lots of tools, storage, processor, users, ...

- But **not all tools** - there are thousands and adding new tools is not taken lightly
- But **not infinite storage and processors** - Main now has job limits and storage quotas

A centralized solution cannot scale to meet data analysis demands of the whole world

Galaxy is available ...

- **As a free (for everyone) web service** integrating a wealth of tools, compute resources, terabytes of reference data and permanent storage
- **As open source software** that makes integrating your own tools and data and customizing for your own site simple

http://getgalaxy.org

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 an existing computational resource on which to be deployed

http://getgalaxy.org

Encourage 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

- Allow sites to share “suites” containing tools, datatypes, workflows, sample data, and automated installation scripts for tool dependencies
- Integration with Galaxy instances to automate tool installation and updates

toolshed.g2.bx.psu.edu

Public Galaxy Servers

http://galaxyproject.org/wiki/PublicGalaxyServers

Interested in:

ChIP-chip and ChIP-seq?

✓ Cistrome

Statistical Analysis?

✓ Genomic Hyperbrowser

Sequence and tiling arrays?

✓ Oqtans

Text Mining?

✓ DBCLS Galaxy

Reasoning with ontologies?

✓ GO Galaxy

Internally symmetric protein structures?

✓ SymD

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 an **existing computational resource** on which to be deployed

http://getgalaxy.org

Got your own cluster?

- Galaxy works with DRMAA compliant cluster job schedulers (which is most of them).
- Galaxy is just another client to your scheduler.

CLUSTER RESOURCES
TORQUE

GRID ENGINE

Platform Computing

DRMAA
Distributed Resource Management Application API — www.drmaa.org

Galaxy is available ...

- **As a free (for everyone) web service** integrating a wealth of tools, compute resources, terabytes of reference data and permanent storage
- **As open source software** that makes integrating your own tools and data and customizing for your own site simple
- **On the Cloud**

http://usegalaxy.org/cloud

Galaxy CloudMan
http://usegalaxy.org/cloud

- 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.

http://aws.amazon.com/education

Step by Step Instructions on the Wiki for Amazon

Galaxy Wiki

Login | Search:

CloudMan/AWS/GettingStarted

Getting Started with Galaxy CloudMan

This page provides a step-by-step instructions on how to start your own instance of Galaxy on Amazon Web Services (AWS) Elastic Compute Cloud (EC2). More general information and instructions about Galaxy CloudMan (GC) can be found here.

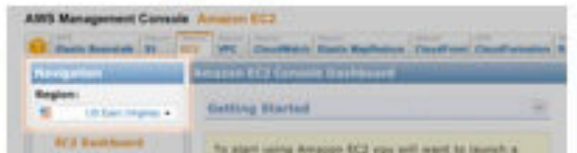
AWS
Get Started
Capacity Planning
AMIs
↑ CloudMan

Contents

Step 1: One Time Amazon Setup

1. Because AWS services implement pay-as-you-go access model for compute resources, it is necessary for every user of the service to *register with Amazon*. You will need a credit card to register. (You can apply for a AWS Education Grant after you register).
2. Once your account has been approved by Amazon (note that this may take up to

Step 1 Screenshots



Instant CloudMan

Launch a CloudMan instance directly from Main

Galaxy Community & Resources: Mailing Lists

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)

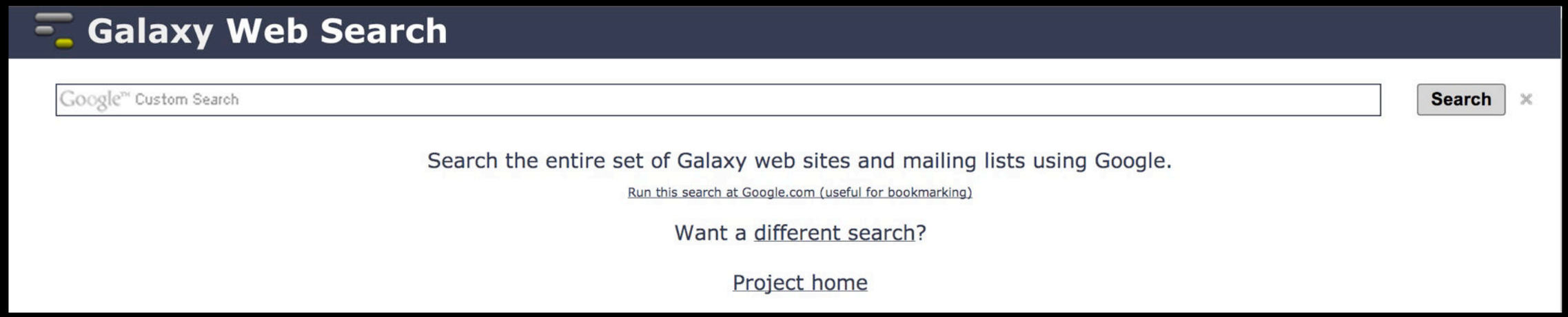
Galaxy-Dev

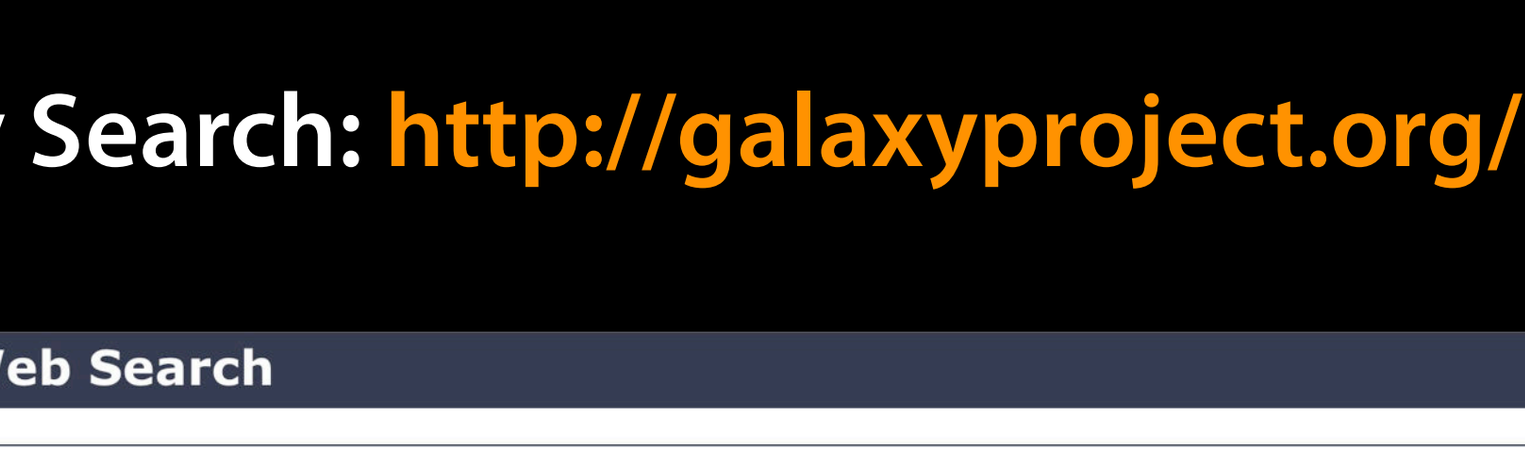
Questions about developing for and deploying Galaxy

High volume (4500 posts, 850 members in 2012)

http://galaxyproject.org/wiki/MailingLists

Galaxy 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

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 **TODAY**

(GIGA)n SCIENCE

GCC2013 Training Day

http://wiki.galaxyproject.org

Galaxy Wiki DaveClements Settings Logout | Search: Titles Text

FrontPage Edit History Actions

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!**

GCC2013 Training Day

Topic voting now open!

Galaxy web search

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

Galaxy Wiki — DaveClements Settings Log

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@glaxyproject.org .

Upcoming Events

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

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.

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.

Visualization

Send data results to **external** genome browsers:

UCSC, Ensembl, GBrowse, IGV

Trackster: Galaxy's genome browser

Trackster

View your data from within Galaxy

- No data transfers to external site
- Use it locally, even without internet access

Supports common filetypes

- BAM, BED, GFF/GTF, WIG

Unique features

- custom genomes
- highly interactive

Galaxy | Published Visualization | GCC2011-1: Viewing and Navigating

http://main.g2.bx.psu.edu/u/jeremy/v/gcc2011-1-viewing-and-navigating

Google

Galaxy | Analyze Data | Workflow | Shared Data | Visualization | Help | User

Published Visualizations | jeremy | GCC2011-1: Viewing and | chr19 | 625,719 - 682,581

630,000 | 640,000 | 650,000 | 660,000 | 670,000 | 680,000

UCSC Main on Human: knownGene (chr19) — Auto (Squish)

UCSC Main on Human: all_est (chr19) — Dense

UCSC Main on Human: phyloP46wayPrimates (chr19) — Histogram

1

-1

h1-hESC Tophat Mapped Reads — Auto (Squish)

630,000 | 640,000 | 650,000 | 660,000 | 670,000 | 680,000

Display a menu

Galaxy | Published Visualization | GCC2011-1: Viewing and Navigating

http://main.g2.bx.psu.edu/u/jeremy/v/gcc2011-1-viewing-and-navigating

Google

Galaxy | Analyze Data | Workflow | Shared Data | Visualization | Help | User

Published Visualizations | jeremy | GCC2011-1: Viewing and chr19 663,032 - 663,110

663,040 663,050 663,060 663,070 663,080 663,090 663,100

g g c c c g g g c c T C A C C G G C A G G C G C G G G A C G A T C T C C A C G G A G C A G C A G T G G C A G A A G T A C C G T C C G G G A T G C G G C G A C

UCSC Main on Human: knownGene (chr19) Auto (Pack)

UCSC Main on Human: all_est (chr19) Dense

UCSC Main on Human: phyloP46wayPrimates (chr19) Histogram

1

-1

h1-hESC Tophat Mapped Reads Auto (Pack)

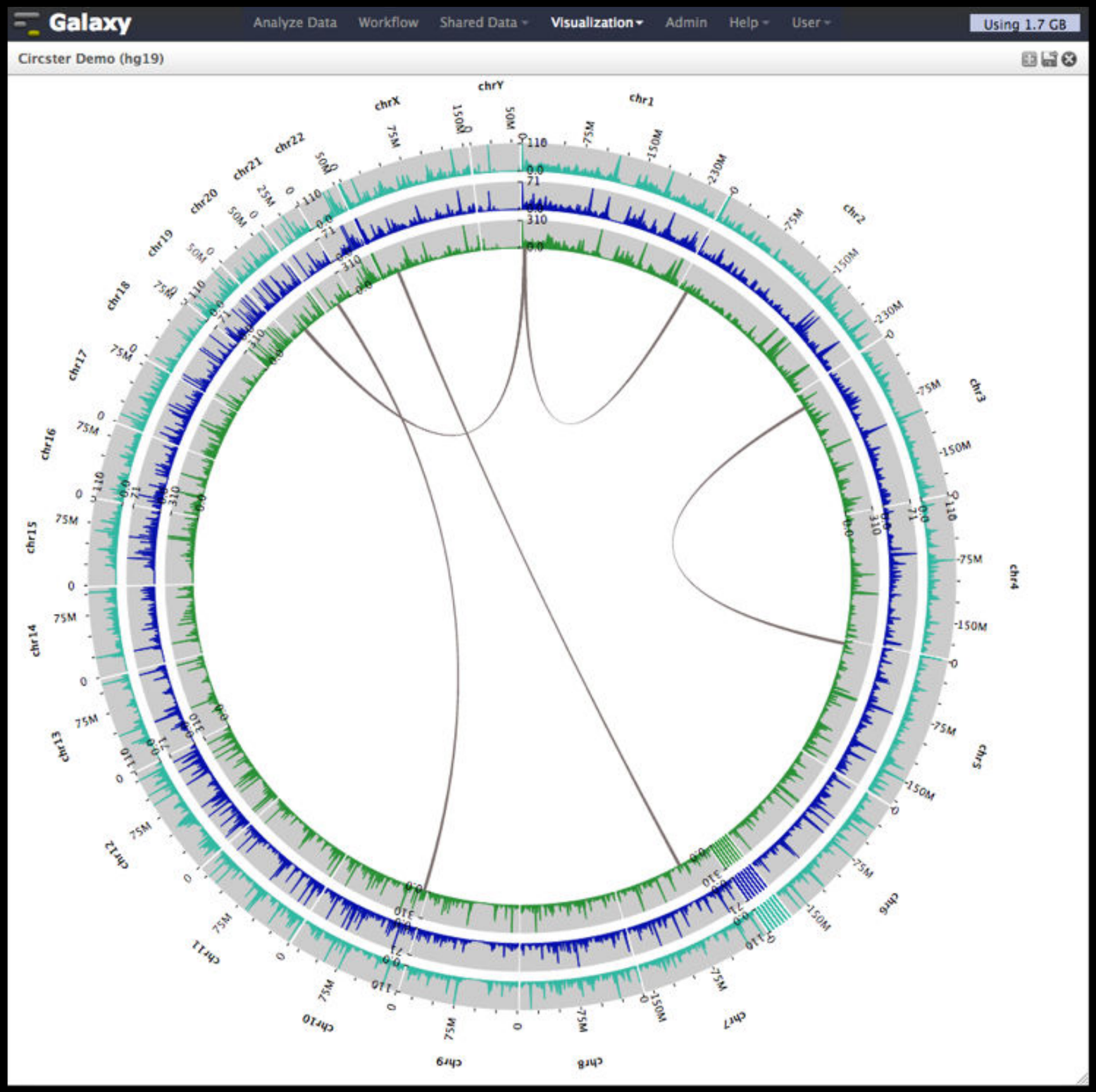
HWI-EAS229_1:2:7:78:211

h1-hESC Cufflinks assembled transcripts Auto (Pack)

663,040 663,050 663,060 663,070 663,080 663,090 663,100

Canceled opening the page

Visualization: Circster



Circster: Circos style visualizations

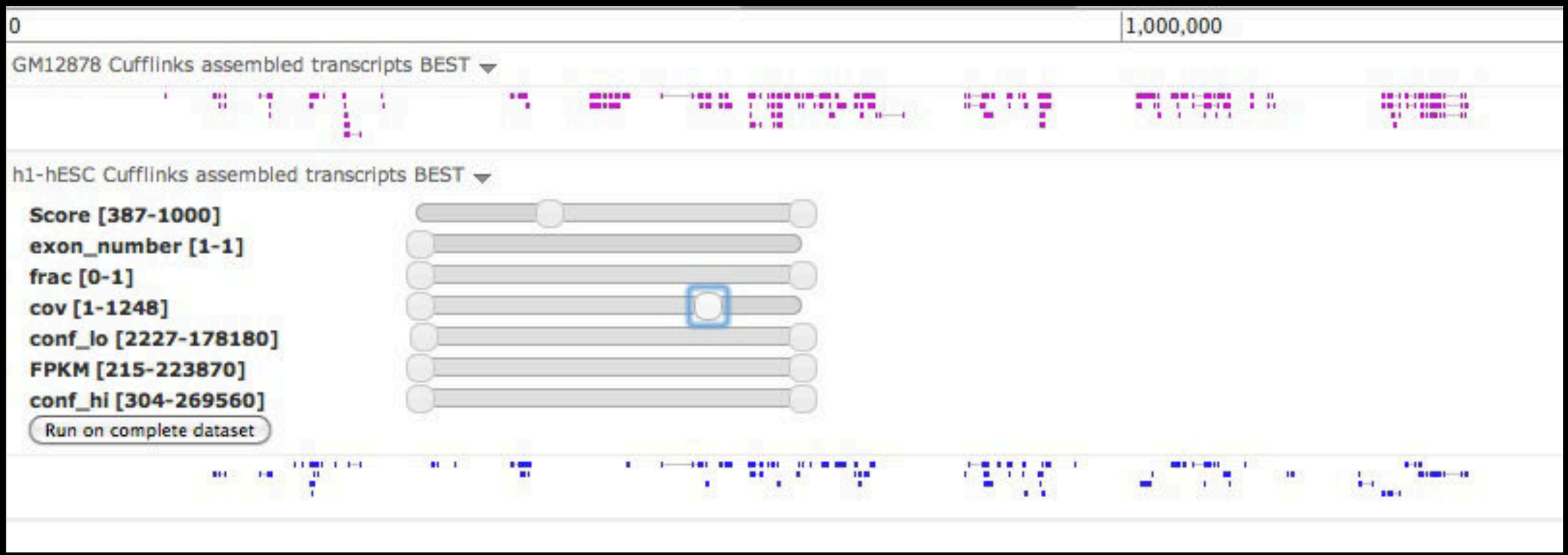
But really, why *another* genome browser

From static browsing to visual analysis

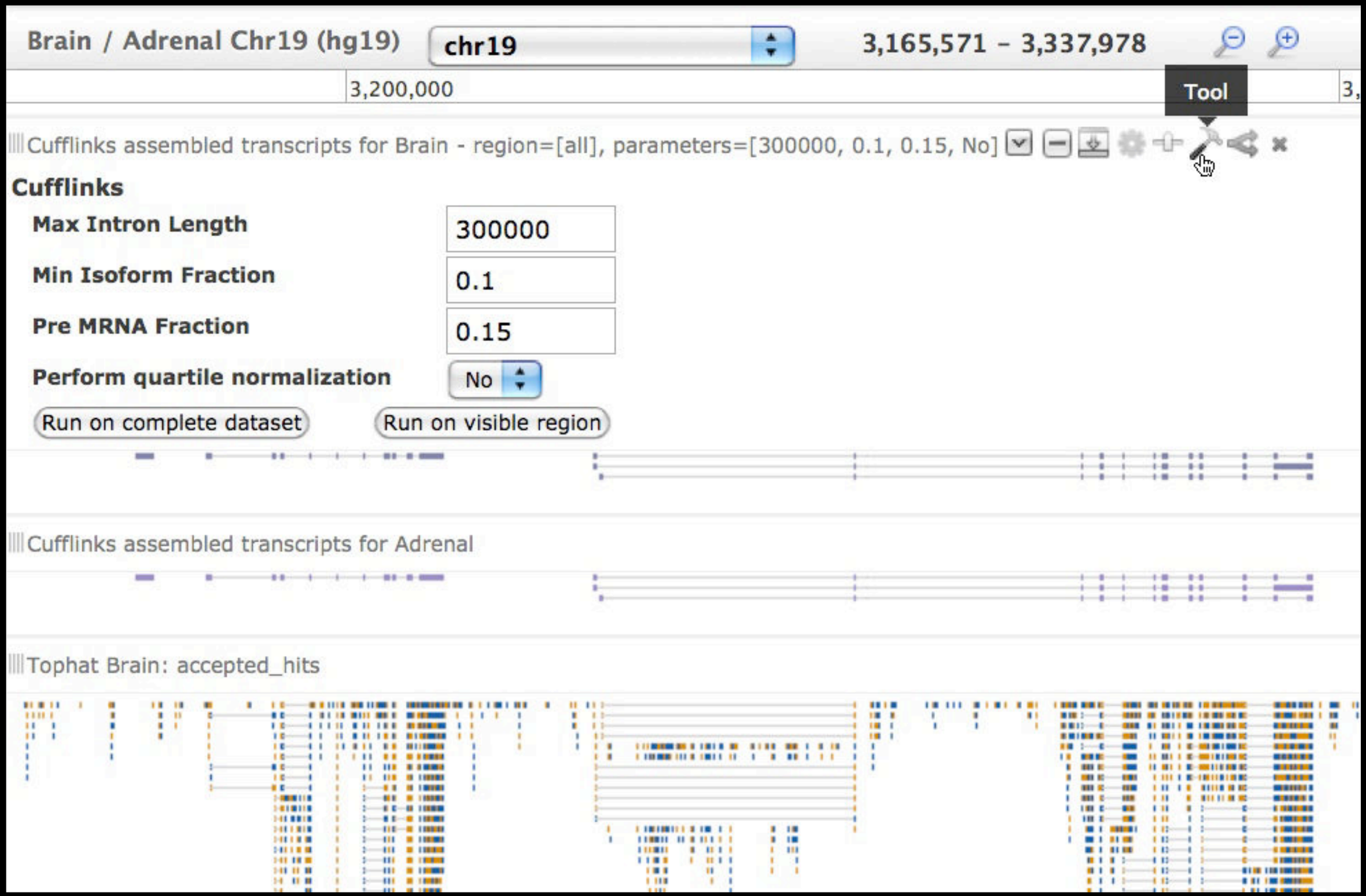
Visual feedback and experimentation needed for complex tools with many parameters

Leverage Galaxy strengths: a very sound model for abstracting interfaces to analysis tools and already integrates an enormous number

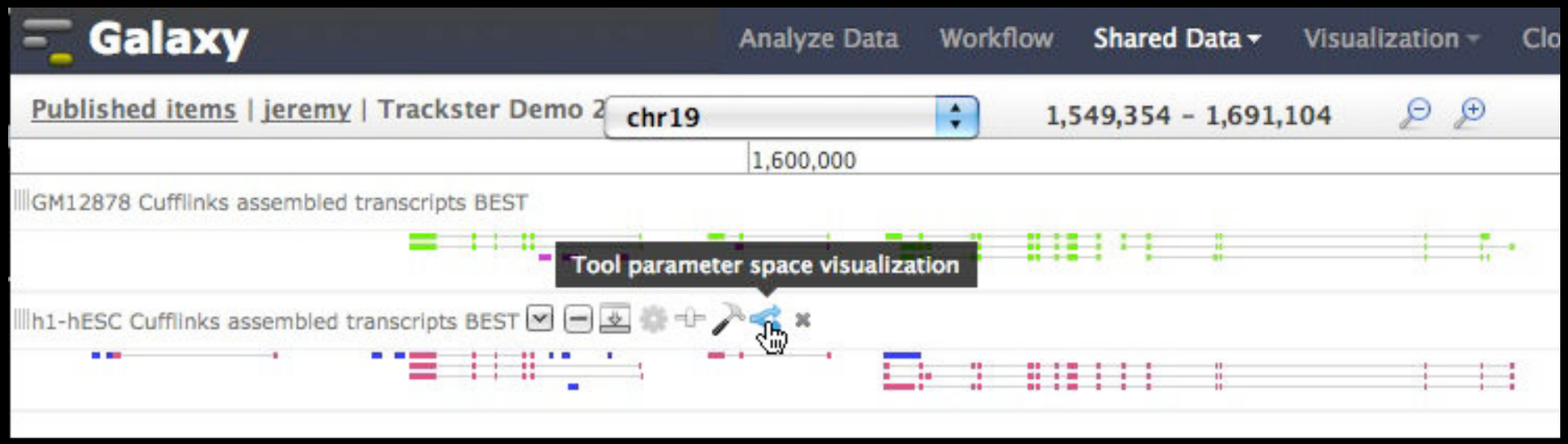
Dynamic Filtering

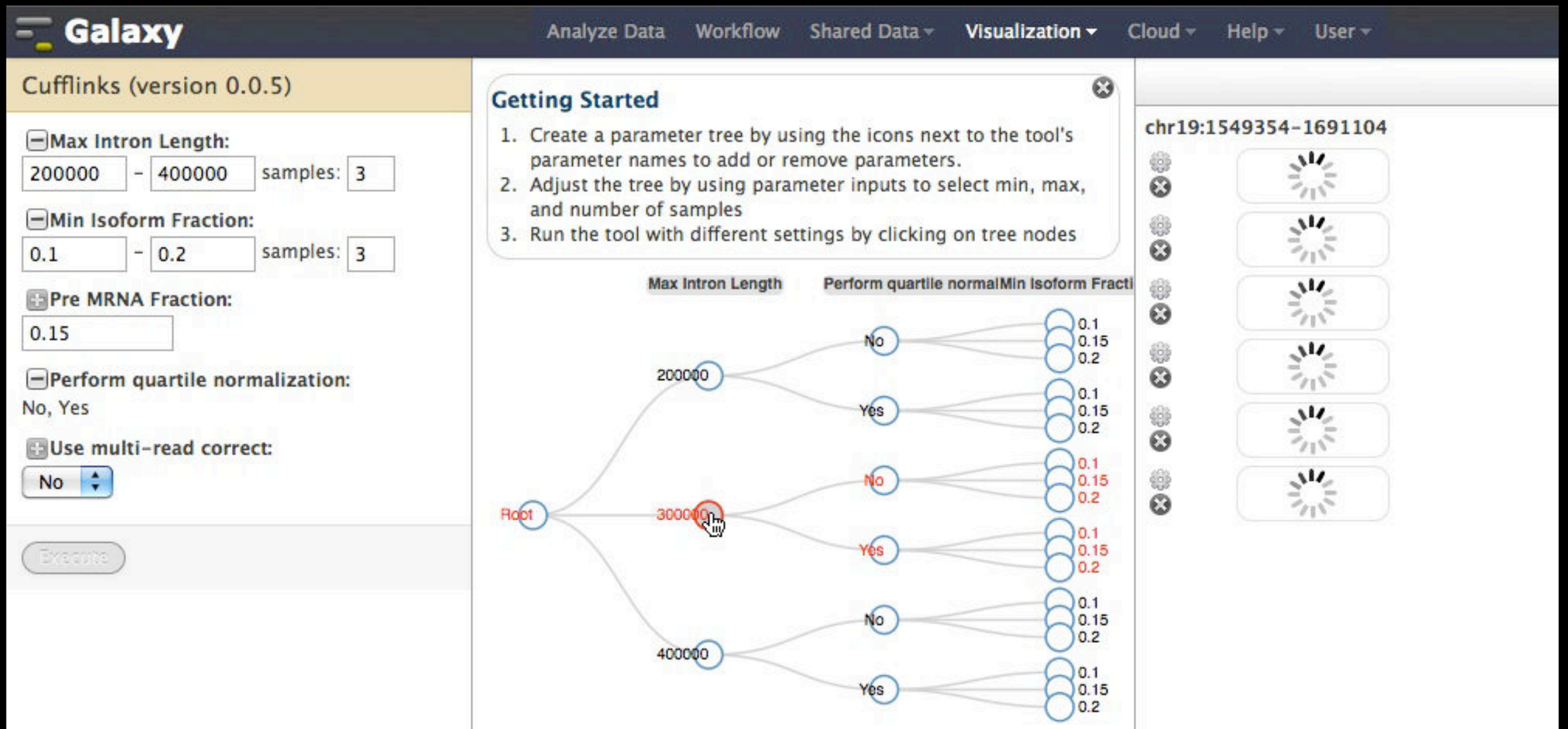


Integrating Tools and Visualization

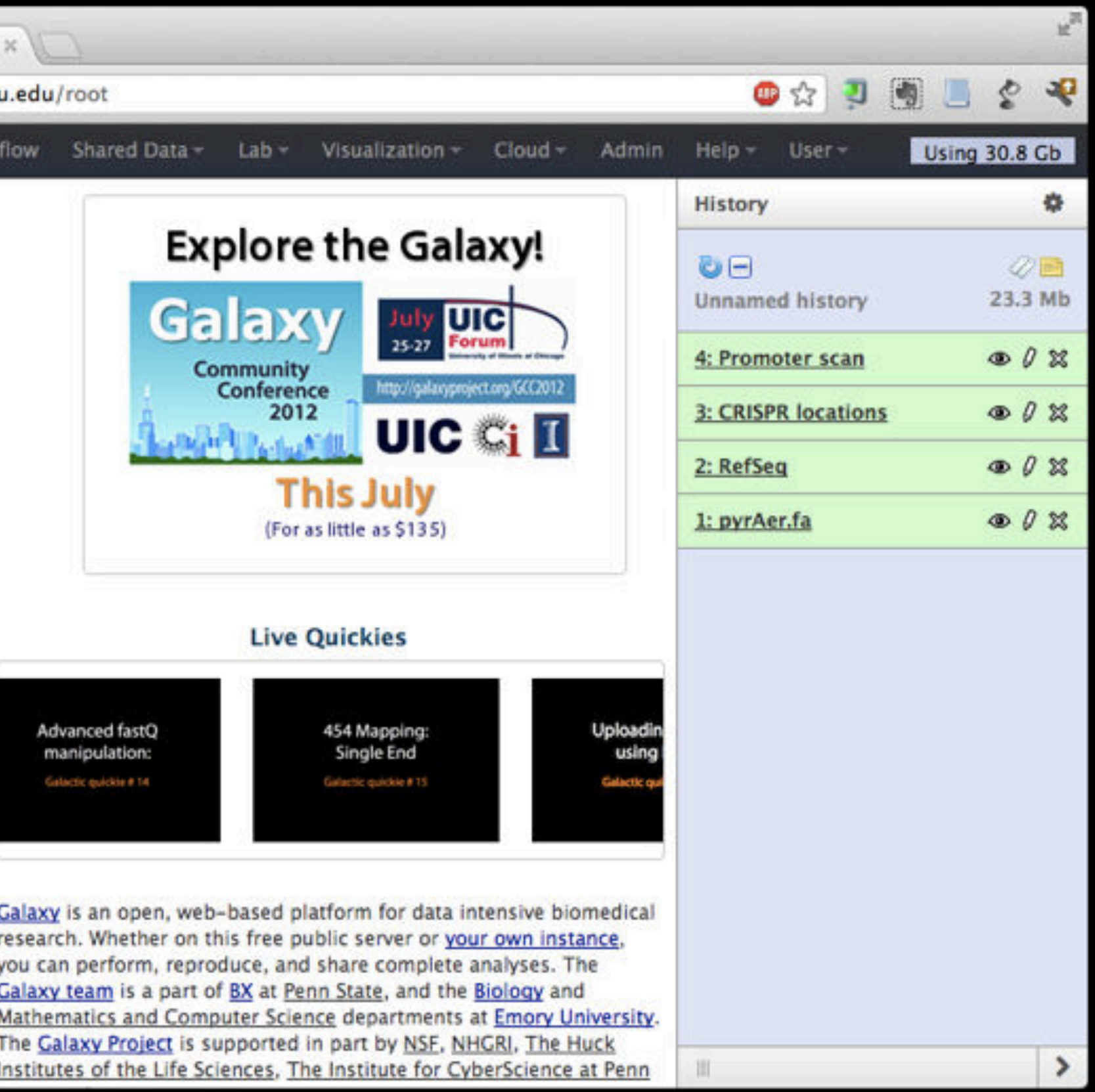


Exploring Parameter Space with Trackster





Building a new browser for your own genome



Promoter potential (WIG)

CRISPR elements (BED)

Genes (BED)

Genome (FASTA)

Galaxy

https://main.g2.bx.psu.edu/root

Galaxy | Analyze Data | Workflow | Shared Data | Lab | Visualization | Cloud | Admin | Help | User | Using 30.8 Gb

Current Custom Builds:

You currently have no custom builds.

Show loaded, system-installed builds

Add a Custom Build

New Build

Name (eg: Hamster):

Pyrobaculum aeroph

Key (eg: hamster_v1):

pyroAero

Definition:

FASTA | Len File | Len Entry

1: pyrAer.fa

Submit

FASTA format

This is a multi-fasta file from your current history that provides the genome sequences for each chromosome/contig in your build.

Here is a snippet from an example multi-fasta file:

```
>chr1
ATTATATATAAGACCACAGAGAGAATATTTTGCCCGG...
>chr2
GGCGGCCGCGGCGATATAGAACTACTCATTATATATA...
...
```

History

Unnamed history — 23.3 Mb

4: Promoter scan

3: CRISPR locations

2: RefSeq

1: pyrAer.fa

Galaxy

https://main.g2.bx.psu.edu/root

Galaxy | Analyze Data | Workflow | Shared Data | Lab | Visualization | Cloud | Admin | Help | User | Using 30.8 Gb

Current Custom Builds:

Name	Key	Number of chroms/contigs	
Pyrobaculum aerophilum str. IM2	pyroAero	1	Delete

Show loaded, system-installed builds

Add a Custom Build

New Build

Name (eg: Hamster):

Key (eg: hamster_v1):

Definition:

FASTA | Len File | Len Entry

1: pyrAer.fa

Submit

FASTA format

This is a multi-fasta file from your current history that provides the genome sequences for each chromosome/contig in your build.

Here is a snippet from an example multi-fasta file:

```
>chr1
ATTATATATAAGACCACAGAGAGAATATTTTGCCCGG...
>chr2
GGCGGCCGCGGCGATATAGAACTACTCATTATATATA...
...
```

History

Unnamed history — 23.3 Mb

4: Promoter scan

3: CRISPR locations

2: RefSeq

1: pyrAer.fa

u.edu/root

flow Shared Data Lab Visualization Cloud Admin Help User Using 30.8 Gb

Software Engineer?
Post-Doc?

The Galaxy Project Wants You!

Live Quickies

Uploading Data using FTP
Galactic quickie # 17

Managing account histories
Galactic quickie # 19

Galaxy is an open, web-based platform for data intensive biomedical research. Whether on this free public server or your own instance, you can perform, reproduce, and share complete analyses. The Galaxy team is a part of BX at Penn State, and the Biology and Mathematics and Computer Science departments at Emory University. The Galaxy Project is supported in part by NSF, NHGRI, The Huck Institutes of the Life Sciences, The Institute for CyberScience at Penn

History

Unnamed history 23.3 Mb

4: Promoter scan

3: CRISPR locations

2: RefSeq
2,603 regions
format: bed, database: pyroAero

view in GeneTrack
display in IGB Local Web

1.Chrom	2.Start	3.End	4.Name	5	6.Str
chr	0	957	PAE0001	1	+
chr	953	1427	PAE0002	2	-
chr	1470	2397	PAE0005	3	-
chr	2639	3797	PAE0007	4	+
chr	3783	4143	PAE0008	5	+
chr	4315	4456	PAE0009	6	-

1: pyrAer.fa

Custom build

Visualize in trackster

Galaxy

https://main.g2.bx.psu.edu/tracks?default_dbkey=pyroAero&dataset_id=f6365eb5c96fc6d...

Galaxy | Analyze Data | Workflow | Shared Data | Lab | Visualization | Cloud | Admin | Help | User | Using 30.8 Gb

Demo (pyroAero) | chr | 0 - 2,222,429

0 | 1,000,000 | 2,000,000

RefSeq

0 | 1,000,000 | 2,000,000

Galaxy

https://main.g2.bx.psu.edu/tracks?default_dbkey=pyroAero&dataset_id=f6365eb5c96fc6d...

Galaxy | Analyze Data | Workflow | Shared Data | Lab | Visualization | Cloud | Admin | Help | User | Using 30.8 Gb

Demo (pyroAero) | chr | 957,612 - 1,029,613

960,000 | 970,000 | 980,000 | 990,000 | 1,000,000 | 1,010,000 | 1,020,000

RefSeq

PAE1646 PAE1653 PAE1658 PAE1670 PAE1677 PAE1687 PAE1694 PAE1704 PAE1715 PAE1719 PAE1727 PAE1736 PAE1749 PAE1759
PAE1647 PAE1656 PAE1665 PAE1672 PAE1683 PAE1690 PAE1702 PAE1712 PAE1718 PAE1726 PAE1735 PAE1744 PAE1755
PAE1649 PAE1657 PAE1671 PAE1685 PAE1693 PAE1703 PAE1714 PAE1720 PAE1729 PAE1738 PAE1752 PAE1760
PAE1639 PAE1650 PAE1659 PAE1673 PAE1688 PAE1696 PAE1708 PAE1717a PAE1723 PAE1731 PAE1740 PAE1756
PAE1641 PAE1651 PAE1662 PAE1674 PAE1689 PAE1698 PAE1709 PAE1722 PAE1730 PAE1742 PAE1761
PAE1642 PAE1652 PAE1663 PAE1676 PAE1695 PAE1710 PAE1725 PAE1748
PAE1643 PAE1667 PAE1678 PAE1706 PAE1734
PAE1645 PAE1668 PAE1679 PAE1716
PAE1669 PAE1681 PAE1717

960,000 | 970,000 | 980,000 | 990,000 | 1,000,000 | 1,010,000 | 1,020,000

Galaxy

https://main.g2.bx.psu.edu/tracks?default_dbkey=pyroAero&dataset_id=f6365eb5c96fc6d...

Galaxy Visualization Using 30.8 Gb

Demo (pyroAero)

Select datasets for new tracks

Histories | Data Libraries

History 'Unnamed history'

search name and filetype

	Id ↑	Name	Filetype	Dbkey
☑	4	Promoter scan	wig	pyroAero
☑	3	CRISPR locations	bed	pyroAero
☐	2	RefSeq	bed	pyroAero

For 2 selected datasets:

Cancel Insert

Galaxy

https://main.g2.bx.psu.edu/tracks/browser?id=d518ad8573371ac2#chr:21663-122199

Galaxy | Analyze Data | Workflow | Shared Data | Lab | Visualization | Cloud | Admin | Help | User | Using 30.8 Gb

Demo (james:pyroAero) | chr | 21,663 - 122,199

100,000

RefSeq

PAE0055 PAE0064 PAE0075 PAE0087 PAE0100 PAE0111 PAE0120 PAE0131 PAE0148 PAE0165 PAE0178 PAE0198 PAE0208 PAE0218
PAE0056 PAE0064a PAE0077 PAE0089 PAE0103 PAE0114 PAE0121 PAE0134 PAE0150 PAE0166 PAE0180 PAE0199 PAE0209 PAE0219
PAE0057 PAE0065 PAE0078 PAE0090 PAE0105 PAE0115 PAE0122 PAE0137 PAE0152 PAE0168 PAE0181 PAE0200 PAE0210 PAE0220
PAE0061 PAE0068 PAE0082 PAE0096 PAE0109 PAE0118 PAE0128 PAE0146 PAE0164a PAE0201 PAE0211 PAE0221
PAE0062 PAE0069 PAE0084 PAE0098 PAE0112 PAE0124 PAE0138 PAE0154 PAE0172 PAE0202 PAE0212 PAE0222
PAE0063 PAE0073 PAE0086 PAE0101 PAE0116 PAE0126 PAE0139 PAE0157 PAE0173 PAE0204 PAE0213 PAE0224
PAE0051 PAE0067 PAE0079 PAE0091 PAE0106 PAE0117 PAE0129 PAE0151 PAE0170 PAE0205 PAE0214 PAE0225
47 PAE0070 PAE0088 PAE0108 PAE0119 PAE0133 PAE0153 PAE0171 PAE0207 PAE0215
49 PAE0080 PAE0097 PAE0135 PAE0158 PAE0175 PAE0216
049a PAE0081 PAE0141 PAE0160 PAE0176
PAE0142 PAE0163
PAE0143 PAE0164
PAE0144 PAE0167

Promoter scan

CRISPR locations

crispr1
crispr2
crispr3

100,000

Galaxy URLs to Remember

http://galaxyproject.org

http://usegalaxy.org

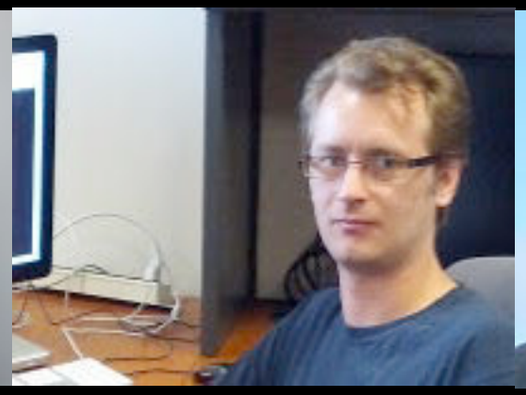
http://getgalaxy.org

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

Galaxy is hiring post-docs and software engineers
at both Emory and Penn State.

Please help.

http://wiki.galaxyproject.org/GalaxyIsHiring

Acknowledgements

Vicky Schneider-Gricar
Helen Tunney
Robert Davey

The Galaxy Team
You!

The Genome Analysis Centre

AWS Education Grant

NIH NSF Huck Institute
Penn State University Emory University

Thank you.