NGS Data Analysis and Galaxy

University of Cape Town
Cape Town, South Africa
21-25 October 2013

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

Gerrit Botha
Computational Biology
University of Cape Town
http://cbio.uct.ac.za/

CBIO Computational Biology @ UCT

AFRICAN CENTRE FOR GENE TECHNOLOGIES

Galaxy

amazon web services™

UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

South Africa Galaxy Workshop Tour

This Week

Monday	Welcome, Project Intro, Basic Galaxy Usage NGS QualityControl
Tuesday	RNA-Seq - Mapping and Transcript Prediction RNA-Seq: Differential expression and Alternative Pipelines; SNP & Variant Analysis
Wednesday	SNP & Variant Analysis Chip-Seq Analysis
Thursday	Genome Assembly Install your own Galaxy on Amazon Cloud
Friday	Customizing Galaxy, Galaxy Tool Shed, and Wrapping Tools for Galaxy

Monday Agenda

9:00 Welcome and Intro
9:40 Basic Analysis with Galaxy
11:00 Break + an Excerise
12:00 Basic Analysis into Reusable Workflows
13:00 Lunch
14:00 NGS Data Quality Control I
15:30 Break
16:00 NGS Data Quality Control II
16:30 Galaxy Community
17:00 Done

Goals

1. Introduce Galaxy
2. Introduce bioinformatics concepts and formats
3. Hands-on experience
    - Load and integrate data
    - Perform bioinformatic analysis
    - Evaluate different analysis options
    - Save, repeat, share describe and publish analyses
    - Visualize your results
    - Set up a Galaxy server in the cloud

This workshop will not cover details of how tools are implemented, or new algorithm designs, or which assembler or mapper or ... is best for you.

Introductions

In 50 seconds or less tell us

- your name
- your affiliation(s)
- something about your research
- something about your goals for today

What is Galaxy?

- **A free (for everyone) web service**
- **Open source software**
- These options result in several **ways to use Galaxy**

http://galaxyproject.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

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
  - Run jobs on existing compute clusters
- Requires a computational resource on which to be deployed

http://getgalaxy.org

Got your own cluster?

- Galaxy works with any DRMAA compliant cluster job scheduler (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

  http://usegalaxy.org

- As open source software

  http://getgalaxy.org

- *On the Cloud*

  http://wiki.galaxyproject.org/Cloud

  We are using this right now, and you will set up your own Galaxy on AWS on Thursday

http://aws.amazon.com/education

Galaxy is available ...

- As a free (for everyone) web service
- As open source software
- On the Cloud
- *With Commercial Support*

A ready-to-use appliance (BioTeam)

Cloud-based solutions (Appistry, ABgenomica, AIS)

Consulting & Customization (Arctix, Deena Bioinformatics)

Galaxy Project: Further reading & Resources

http://galaxyproject.org

http://usegalaxy.org

http://getgalaxy.org

http://wiki.galaxyproject.org/Cloud

http://bit.ly/gxychoices

Monday Agenda

9:00 Welcome and Intro
9:40 Basic Analysis with Galaxy
11:00 Break + an Excerise
12:00 Basic Analysis into Reusable Workflows
13:00 Lunch
14:00 NGS Data Quality Control I
15:30 Break
16:00 NGS Data Quality Control II
16:30 Galaxy Community
17:00 Done

Basic Analysis

Which genes have most overlapping Repeats?

HG19, chr22

http://cloud1.galaxyproject.org/
http://cloud2.galaxyproject.org/
http://cloud3.galaxyproject.org/

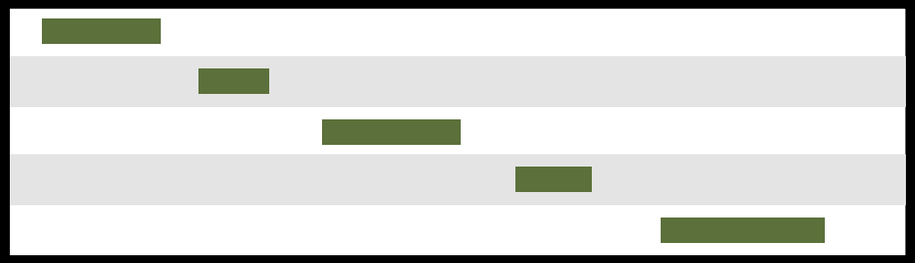
(~ http://usegalaxy.org/galaxy101)

Genes & Repeats: A General Plan

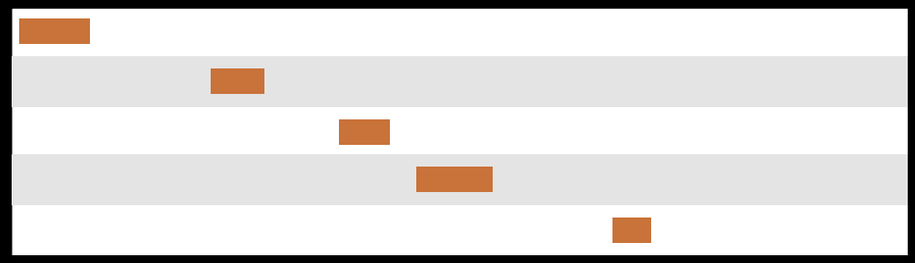
- Get some data
  - Get Data → UCSC Table Browser
- Identify which genes/exons have Repeats
- Count Repeats per exon
- Visualize, save, download, ... exons with most Repeats

http://cloud1.galaxyproject.org/
http://cloud2.galaxyproject.org/
http://cloud3.galaxyproject.org/

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

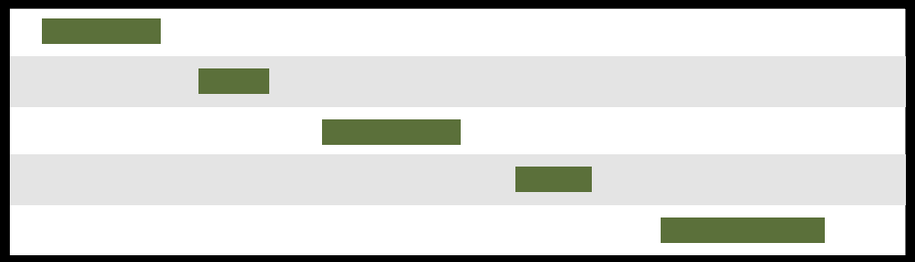


Exons

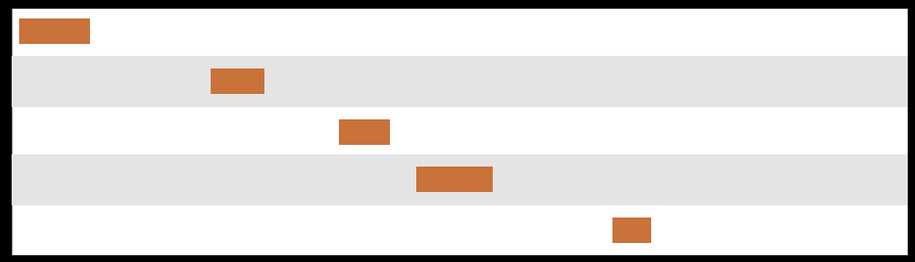


Repeats

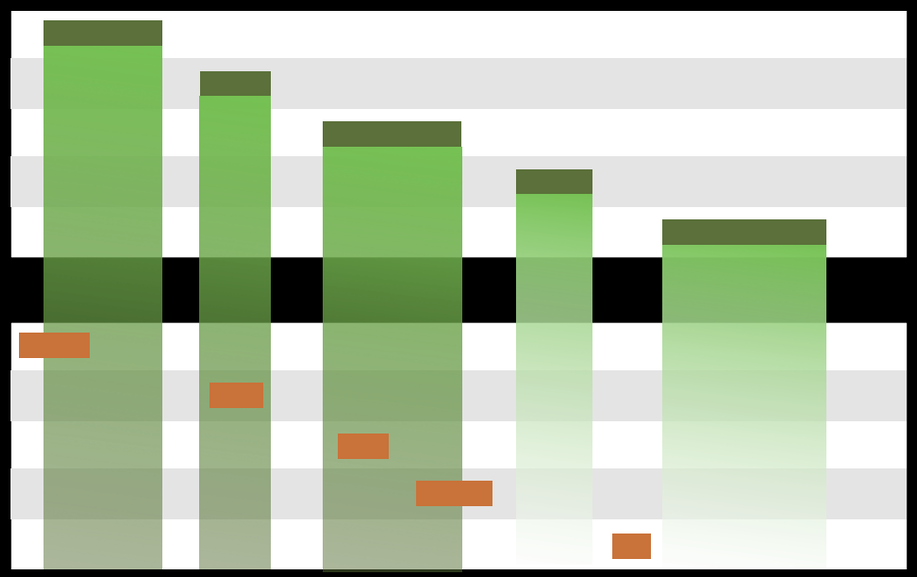
(Identify which genes/exons have Repeats)



Exons



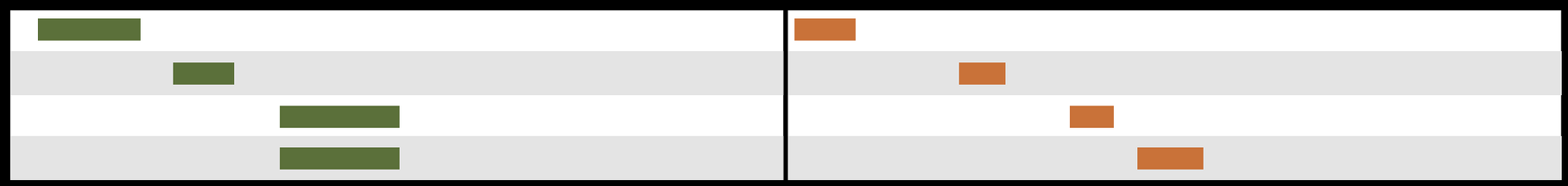
Repeats



Exons

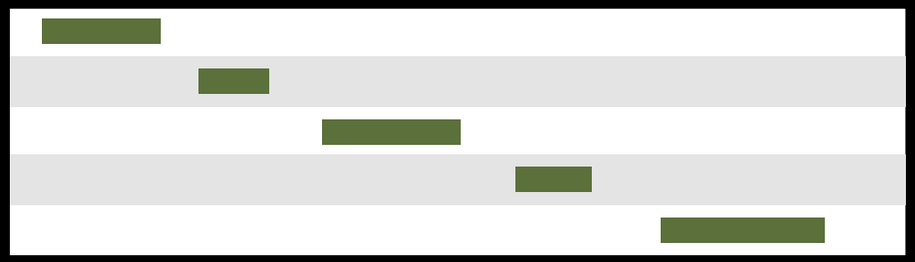
Repeats

Overlap pairings

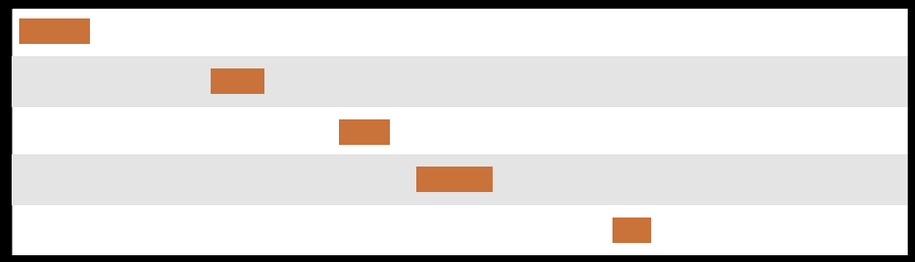


Operate on Genomic Intervals → **Join**

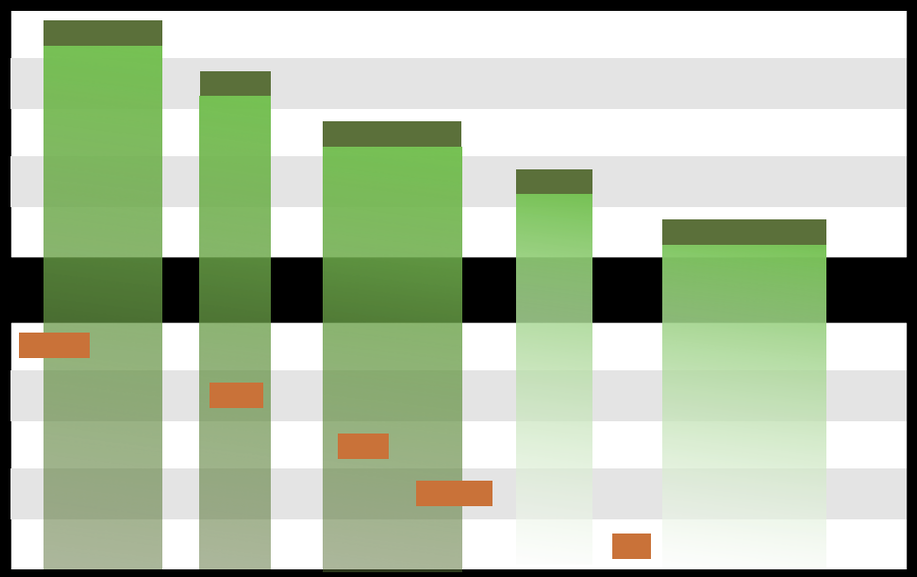
(Identify which genes/exons have Repeats)



Exons



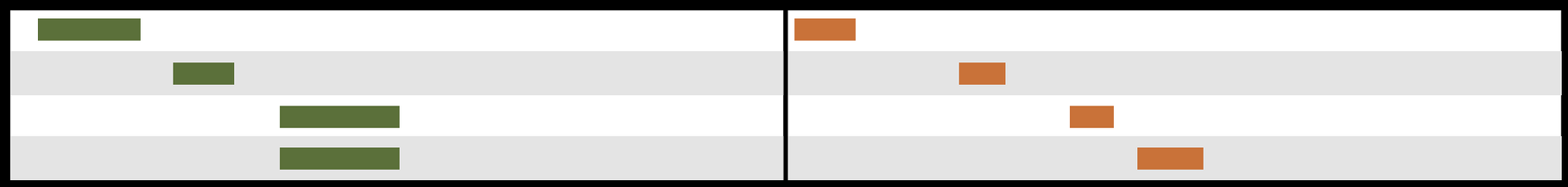
Repeats



Exons

Repeats

Overlap pairings

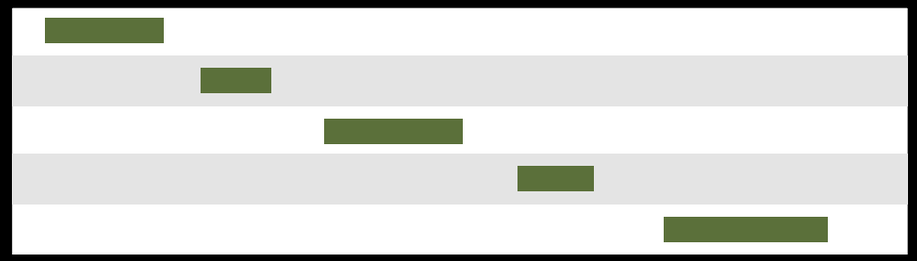


Exon overlap counts

Join, Subtract, and Group → Group

(Count Repeats per exon)

Exon overlap counts

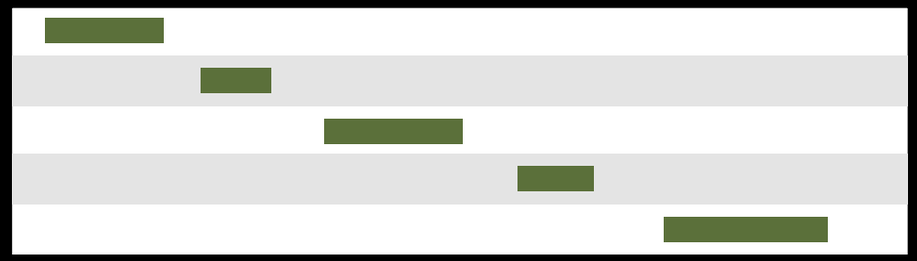


Exons

We’ve answered our question, but we can do better.

Incorporate the overlap count with rest of Exon information

Exon overlap counts



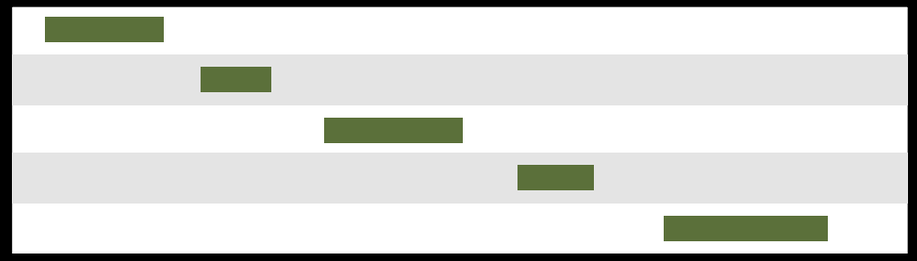
Exons

Join on exon name

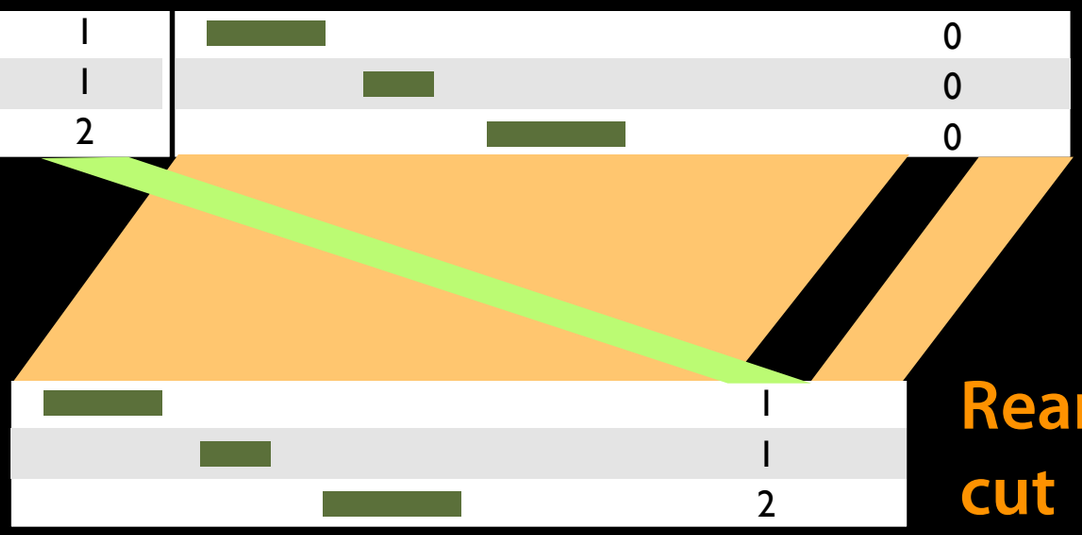
Join, Subtract, and Group → Join

(Incorporate the overlap count with rest of Exon information)

Exon overlap counts



Exons



Join on exon name

Rearrange columns w/ cut

Text Manipulation → Cut

(Incorporate the overlap count with rest of Exon information)

Basic Analysis: Further reading & Resources

http://usegalaxy.org/galaxy101

https://vimeo.com/76343659

Monday Agenda

9:00 Welcome and Intro
9:40 Basic Analysis with Galaxy
11:00 Break + an Excerise
12:00 Basic Analysis into Reusable Workflows
13:00 Lunch
14:00 NGS Data Quality Control I
15:30 Break
16:00 NGS Data Quality Control II
16:30 Galaxy Community
17:00 Done

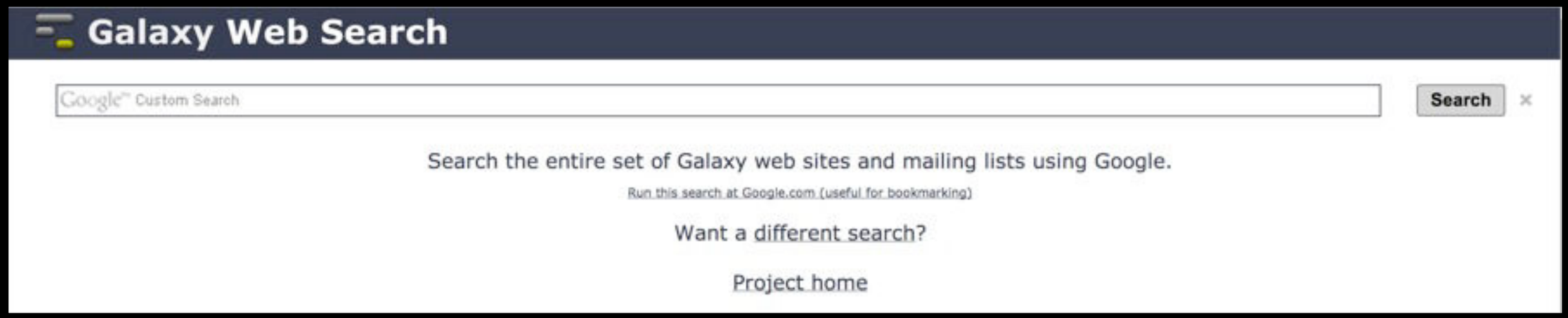
Genes & Repeats: Exercise

Include genes/exons with no overlaps in final output. Set the score for these to 0.

Everything you need will be in the toolboxes we used in the first Gene/Exon-Repeats exercise.

http://cloud1.galaxyproject.org/
http://cloud2.galaxyproject.org/
http://cloud3.galaxyproject.org/

Your Friend: http://galaxyproject.org/search



Find

Everything on …

Tools for …

Email about …

Source code for …

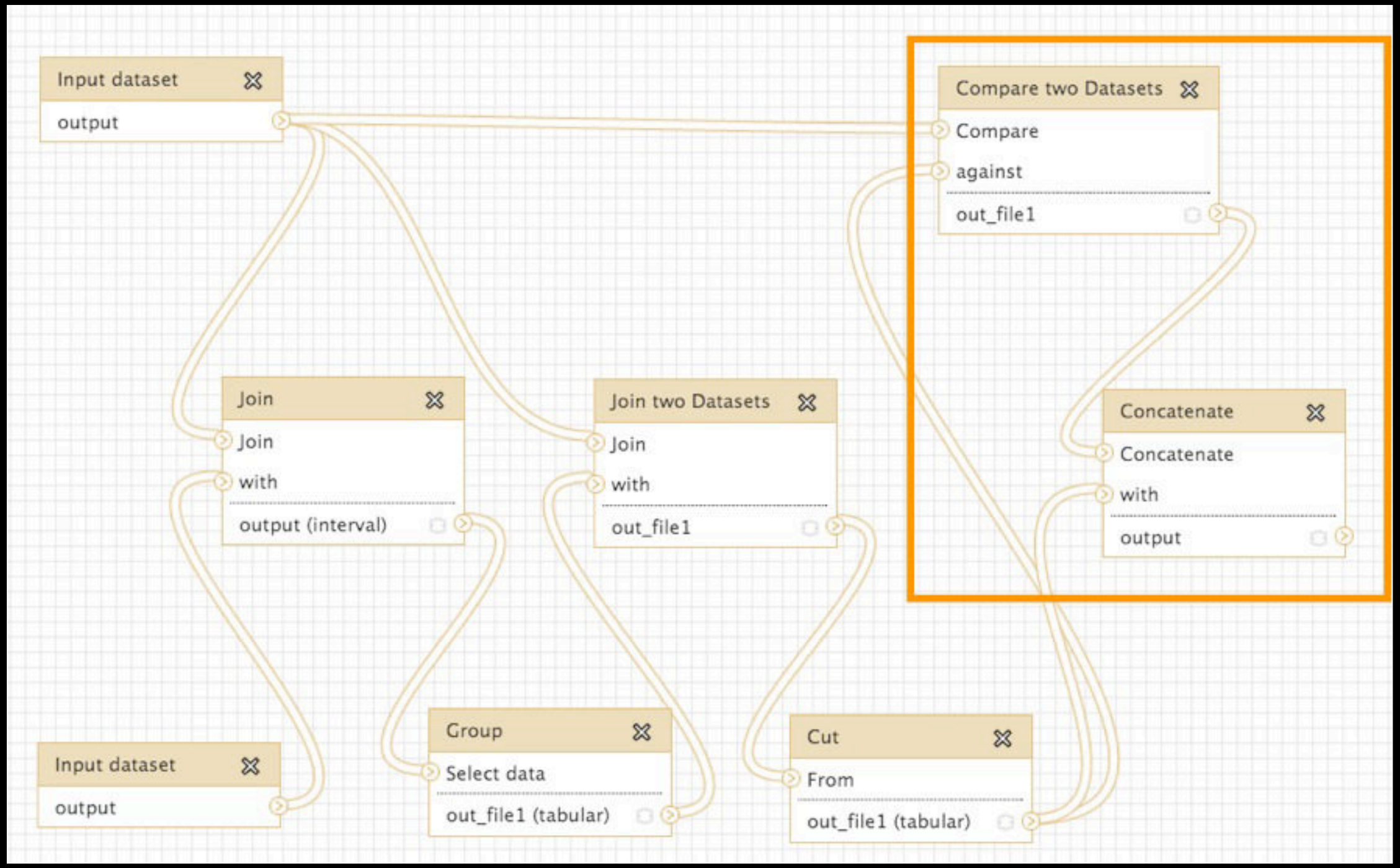
Published Histories, Pages, Workflows, about …

Documentation on …

Papers using Galaxy for …

Related feature requests

One Possible Solution



Solution from Stanford Kwenda and Caron Griffiths in Pretoria.

Takes advantage of the fact that Exons already have 0 scores.

Monday Agenda

9:00 Welcome and Intro
9:40 Basic Analysis with Galaxy
11:00 Break + an Excerise
12:00 Basic Analysis into Reusable Workflows
13:00 Lunch
14:00 NGS Data Quality Control I
15:30 Break
16:00 NGS Data Quality Control II
16:30 Galaxy Community
17:00 Done

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

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

- The analysis we just finished was about
  - Human chr22
  - Overlap between exons and Repeats
- But, ...
  - there is nothing inherent in the analysis about humans, 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

Guided: rerun with same inputs

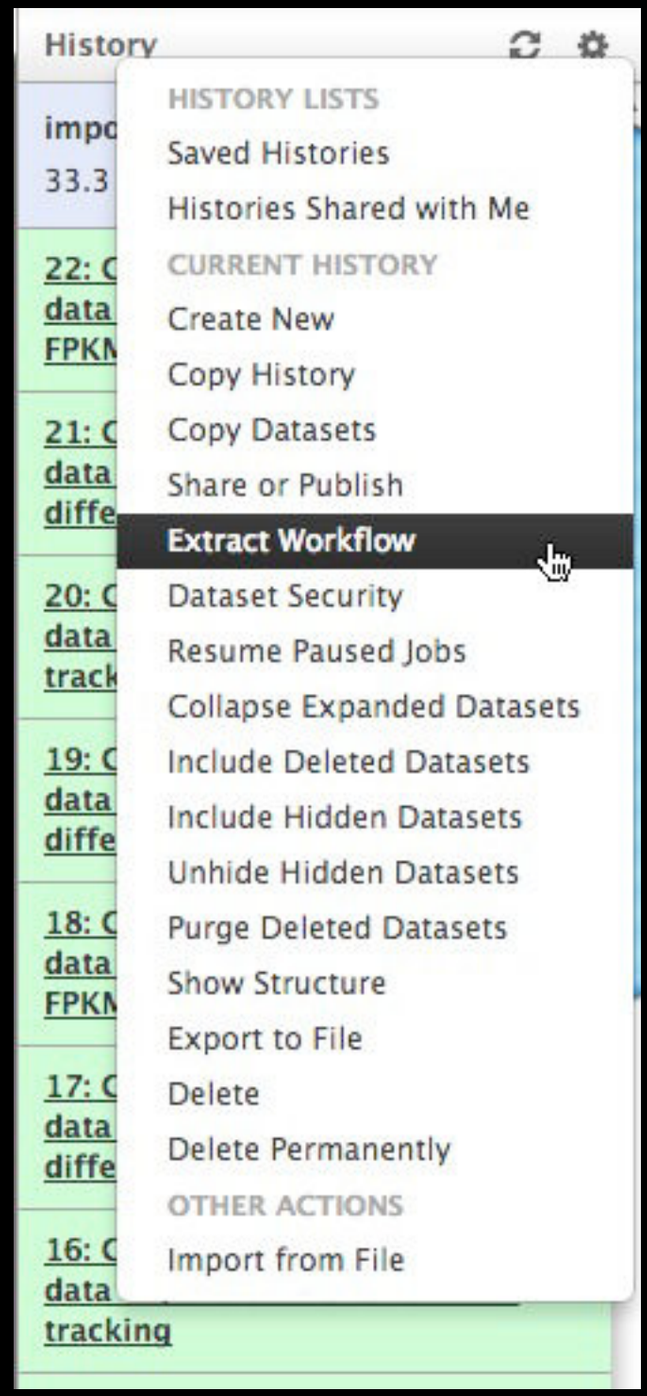
Did that work?

On your own:

Count # of exons in each Repeat

Did that work? *Why not?*

Edit workflow: doc assumptions



Monday Agenda

9:00 Welcome and Intro
9:40 Basic Analysis with Galaxy
11:00 Break + an Excerise
12:00 Basic Analysis into Reusable Workflows
13:00 Lunch
14:00 NGS Data Quality Control I
15:30 Break
16:00 NGS Data Quality Control II
16:30 Galaxy Community
17:00 Done

Monday Agenda

9:00 Welcome and Intro
9:40 Basic Analysis with Galaxy
11:00 Break + an Excerise
12:00 Basic Analysis into Reusable Workflows
13:00 Lunch
14:00 NGS Data Quality Control I
15:30 Break
16:00 NGS Data Quality Control II
16:30 Galaxy Community
17:00 Done

NGS Data Quality Control

- Introduce FASTQ format
- Examine quality in an RNA-Seq dataset
- Trim/filter as we see fit
- Fix some things we just broke

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!

```
  SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS
  ..............................IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
  ..........................XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
  !"#$%&'()*+,-./0123456789:;<=>?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz{|}~
  |                         |    |        |                              |                     |
 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

NGS Data Quality Exercise

Create new history


(cog) → Create New

Get some data

Shared Data → Data Libraries

→ UC Davis RNA-Seq Human*

→ Select MeOH_REP1_R1,
MeOH_REP1_R2 and then
Import to current history

RNA-Seq example datasets from the 2013 UC Davis Bioinformatics Short Course. http://bit.ly/ucdbsc2013

NGS Data Quality: Assessment tools

Options 1 & 2:

1. NGS QC and Manipulation → Compute Quality Statistics

   NGS QC and Manipulation → Draw quality score boxplot

   No control over how it is calculated or presented, statistics in text and graphic formats.

2. NGS QC and Manipulation → FastQ Summary Statistics,

   Graph / Display Data → Boxplot of quality statistics

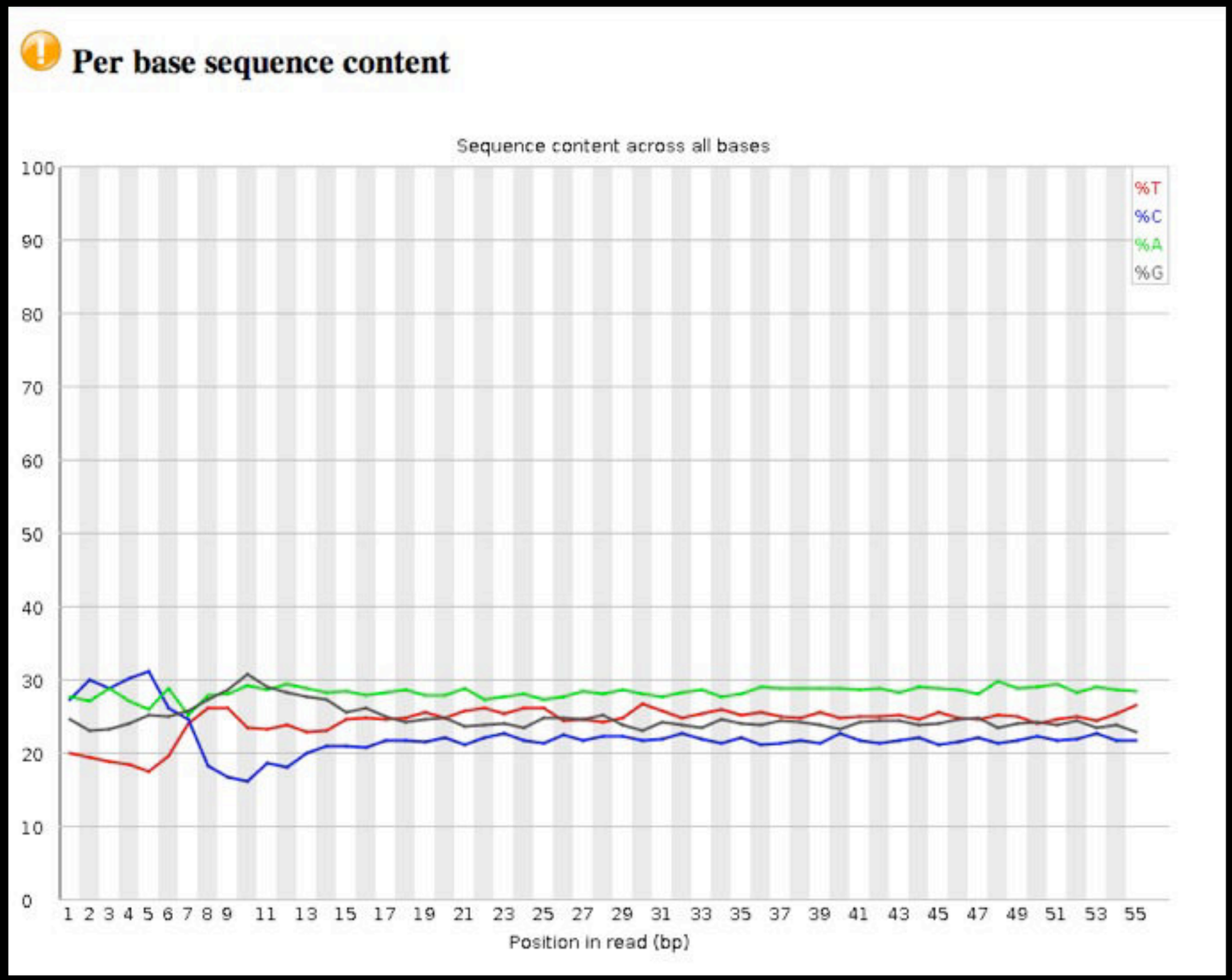
   Lots of control over what the box plot looks like, statistics in text and graphic formats

NGS Data Quality: Assessment tools

- Option 3
  - NGS QC and Manipulation → **FastQC**
  - Gives you a lot a lot more information but little control over how it is calculated or presented.

http://bit.ly/FastQCBoxPlot

NGS Data Quality: Sequence bias at front of reads?

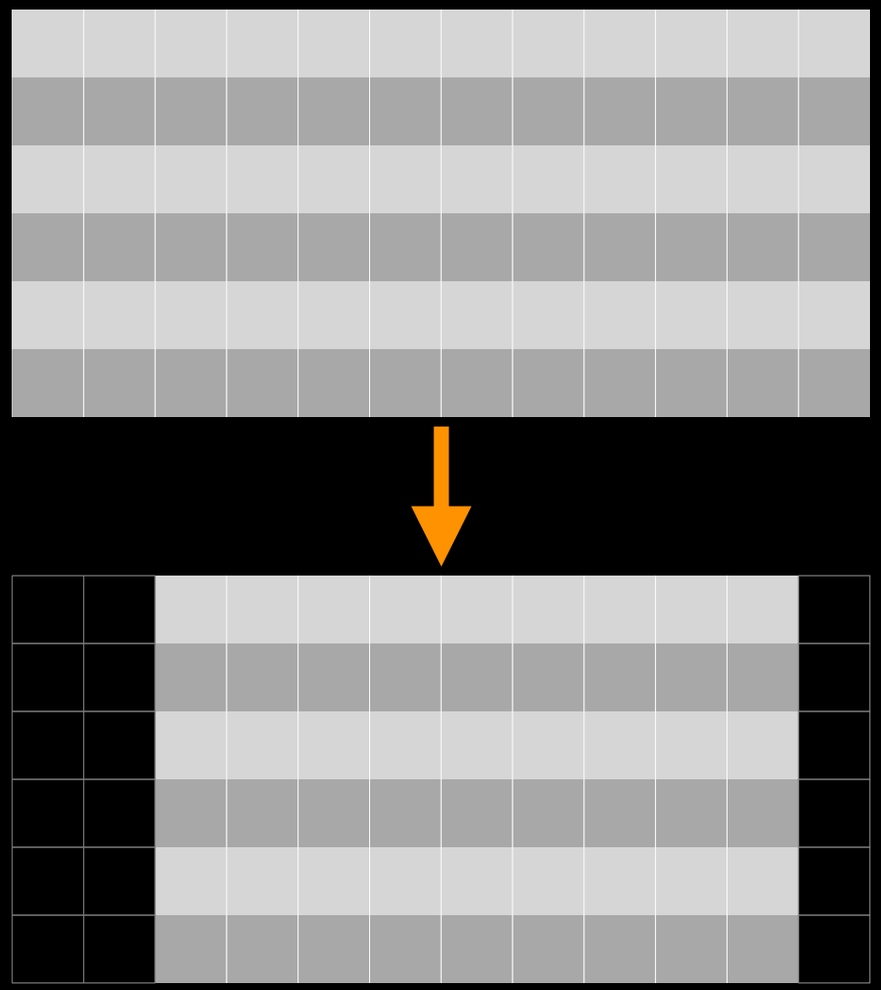


From a sequence specific bias that is caused by use of random hexamers in library preparation.

Hansen, *et al.*, "Biases in Illumina transcriptome sequencing caused by random hexamer priming" *Nucleic Acids Research*, Volume 38, Issue 12 (2010)

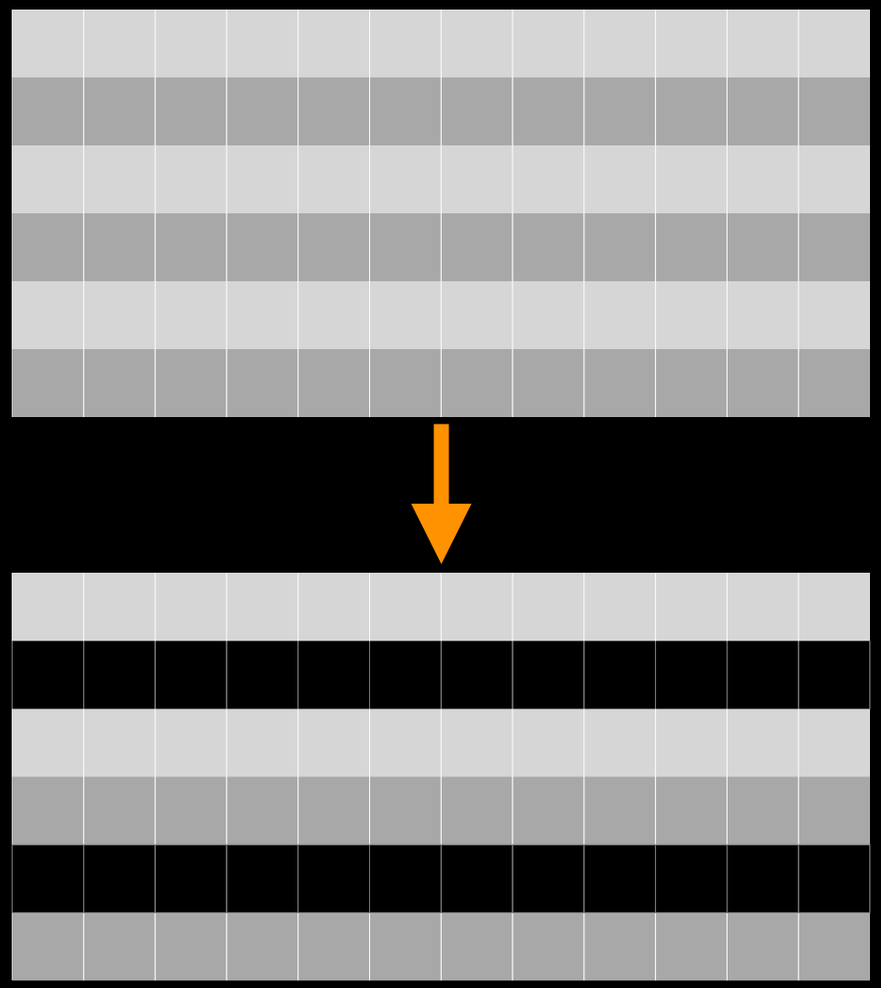
NGS Data Quality: Base Quality Trimming

- 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



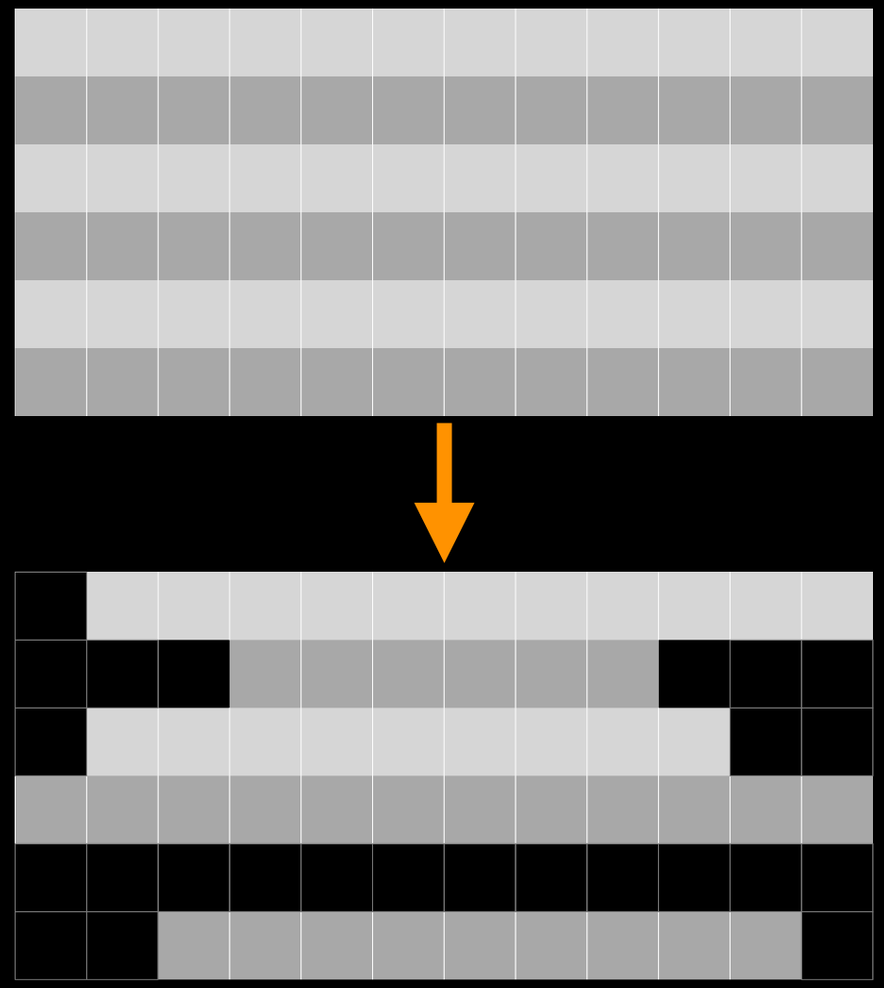
NGS Data Quality: Base Quality Trimming

- ~~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
  - Can have different thresholds for different regions of the reads.
  - Keeps original read length.

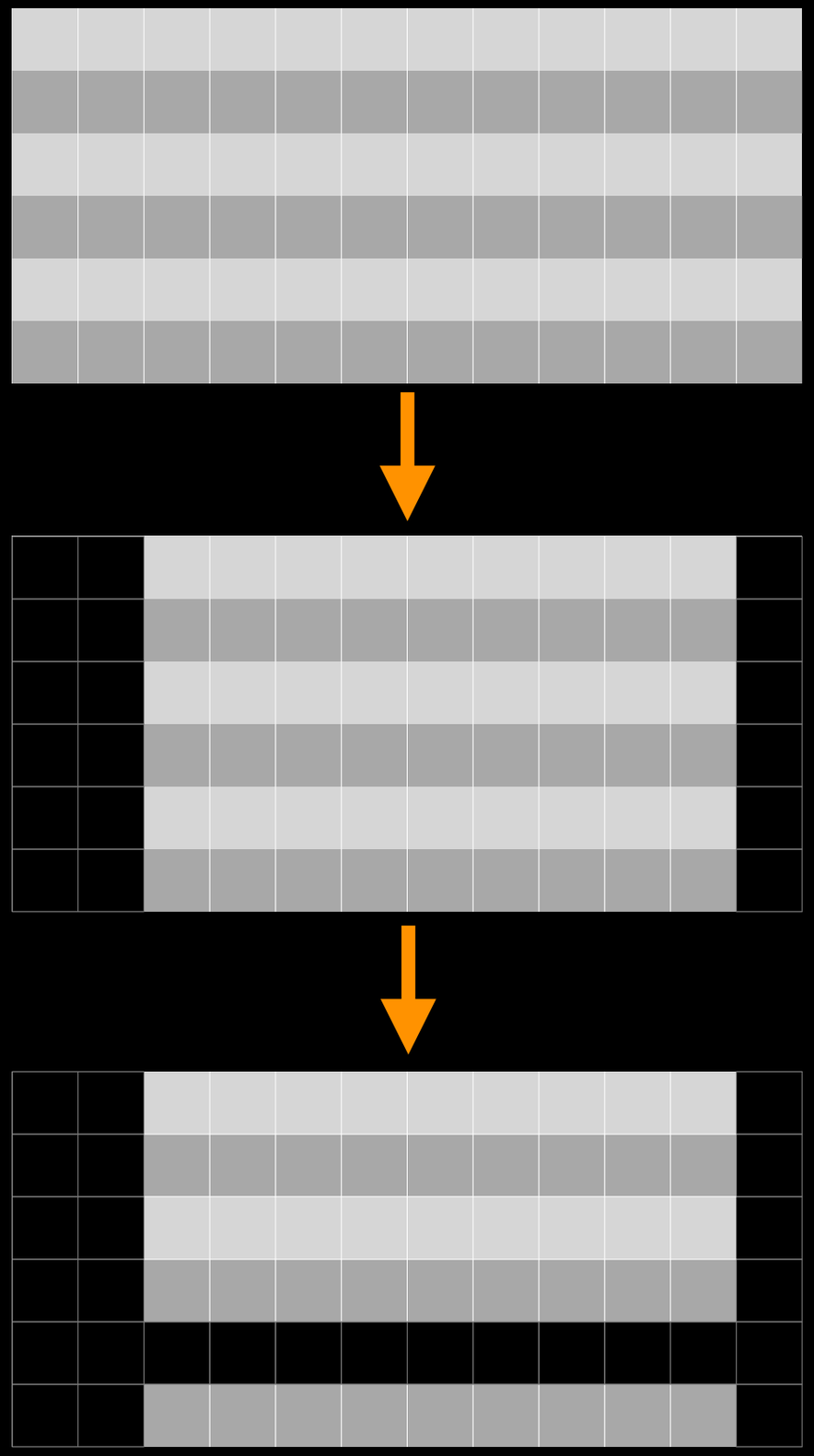


NGS Data Quality: Base Quality Trimming

- 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



Options are not mutually exclusive



Option 1

+

Option 2

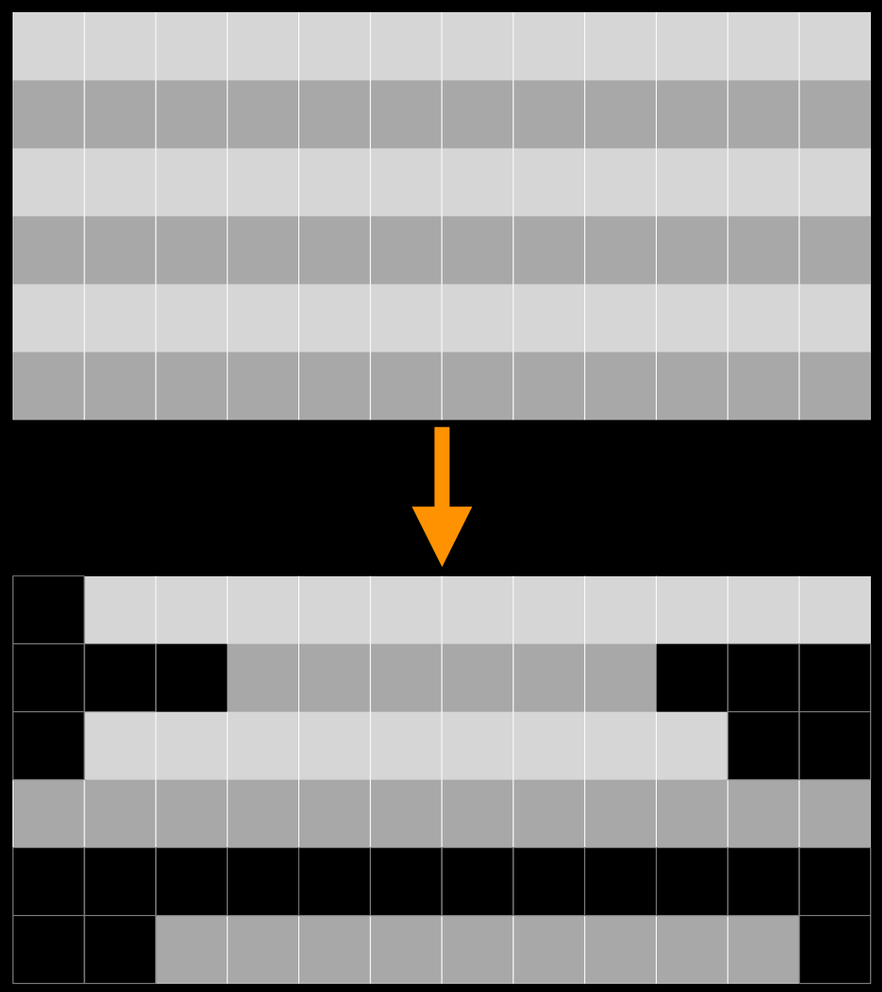
Trim? *As we see fit?*

- Introduced 3 options
  - One preserves original read length, two don't
  - One preserves number of reads, two don't
  - Two keep/make every read the same length, one does not
- One preserves pairings, two don't
  - Can also trim aggressively and then restore pairings

Trim? *As we see fit?*

- Choice depends on downstream tools
- Find out assumptions & requirements for downstream tools and make appropriate choice(s) now.
- How to do that?
  - http://biostars.org/
  - http://seqanswers.com/
  - http://galaxyproject.org/search

NGS Data Quality: Base Quality Trimming



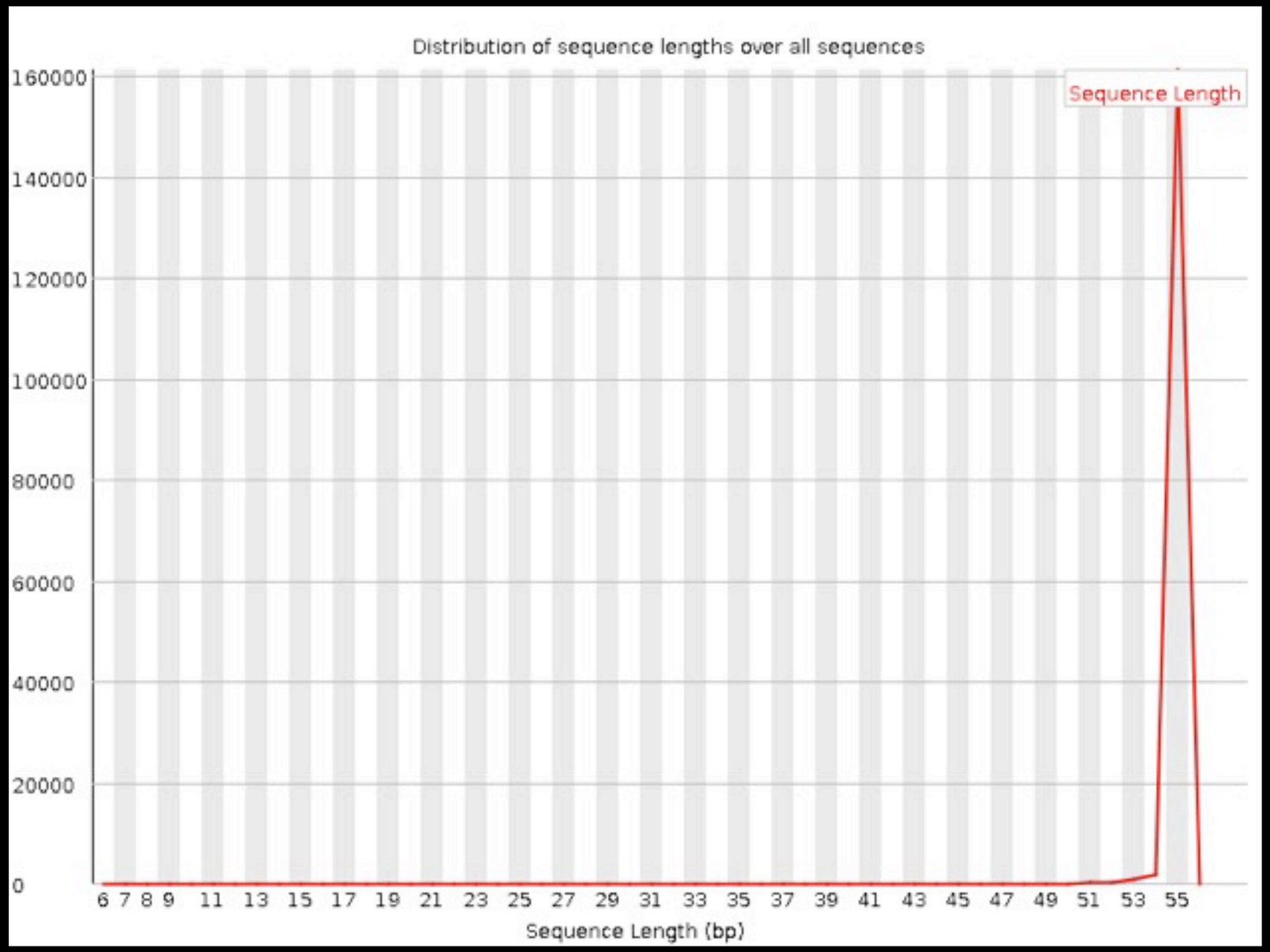
I'll use Option 3

- NGS QC and Manipulation → **FASTQ Quality Trimmer by sliding window**

Run

- NGS QC and Manipulation → **FastQC** on trimmed dataset

NGS Data Quality: Base Quality Trimming



New Problem:

Now some reads are so short they are just noise and can't be meaningfully mapped

Option 2!

NGS QC and Manipulation →
Filter FASTQ reads by quality score and length
NGS QC and Manipulation → **FastQC** on trimmed dataset

NGS Data Quality: Sequencing Artifacts

Repeat this process with MeOH Rep R2 (the reverse reads)

... and there's a new problem in Overrepresented sequences:

Overrepresented sequences

Sequence	Count	Percentage	Possible Source
CTGTGTATTTGTCAATTTTCTTCTCCACGTTCTTCTCGGCCTGTTTCCGTAGCCT	590	0.3541692929220167	No Hit
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	342	0.2052981325073385	No Hit
CGGCCACAAATAAACACAGAAATAGTCCAGAATGTCACAGGTCCAGGGCAGAGGA	325	0.1950932545756871	9 No Hit
CTGCATTATAAAAAGGACAGCCAGATATCAACTGTTACAGAAATGAAATAAGACG	230	0.1380659955458709	3 No Hit
CGGCCGCAAATAAACACAGAAATAGTCCAGAATGTCACAGGTCCAGGGCAGAGGA	199	0.1194571004940361	4 No Hit
GTCAGCTCAACTTGTAGGCCCCAAAAGAAAACAGCGTCTTACTGGGGAGGGATAT	197	0.1182565266197242	2 No Hit

NGS QC and Manipulation → **Remove sequencing artifacts**

NGS Data Quality: Done with 1st Replicate!

Now, only ~~5~~ 3 more to go...

Exercise:

Load the MeOH_REP2, R3G_REP1, and R3G_REP2 replicates into your history, and

Create a workflow that runs a single FASTQ file through all the quality steps.

Or ...

Create a workflow that runs a pair of FASTQ files through all the quality steps

NGS Data Quality: Restoring Pairings

"Mixing paired- and single- end reads together is **not** supported."

Tophat manual

"Dang."

Dave C

Is that really true? Running Tophat on *no-longer-cleanly-paired* data *does map the reads*, ...

But, it no longer keeps track of read pairs in the SAM/BAM file.

Run the "Re-pair paired end reads after QC may have deleted some of them" workflow on each set of paired end reads.

Each workflow run takes the raw and trimmed versions of the forward and reverse reads for each replicate.

NGS Data Quality: Further reading & Resources

FastQC Documenation

Read Quality Assessment & Improvement

by Joe Fass

From the UC Davis 2013 Bioinformatics Short Course

Manipulation of FASTQ data with Galaxy

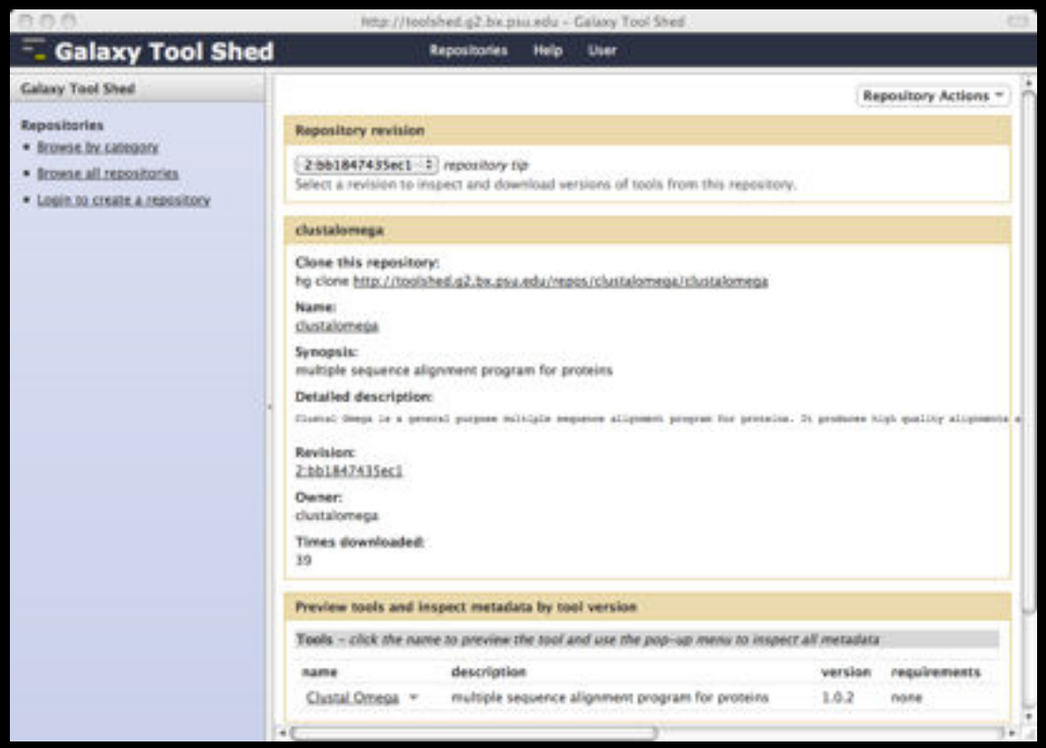
by Blankenberg, *et al*.

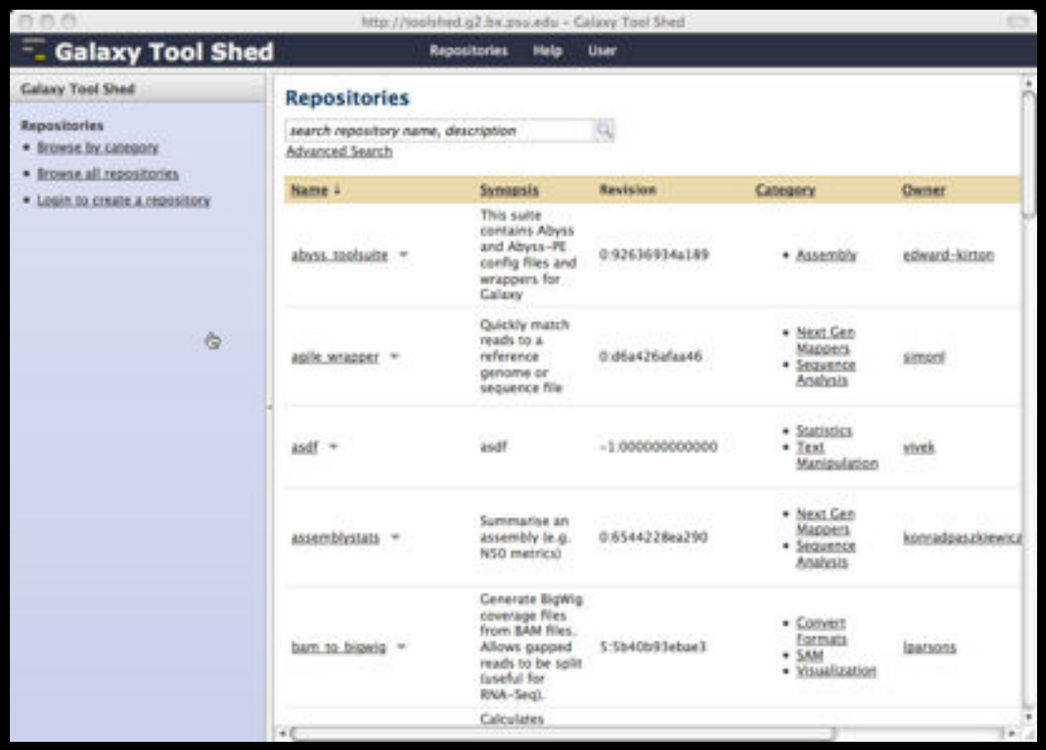
Monday Agenda

9:00 Welcome and Intro
9:40 Basic Analysis with Galaxy
11:00 Break + an Excerise
12:00 Basic Analysis into Reusable Workflows
13:00 Lunch
14:00 NGS Data Quality Control I
15:30 Break
16:00 NGS Data Quality Control II
16:30 Galaxy Community
17:00 Done

Community: 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

Community: Public Galaxy Instances

http://bit.ly/gxyServers

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

Plus many more

Galaxy Resources and Community: Mailing Lists

http://wiki.galaxyproject.org/MailingLists

Galaxy-Announce

Project announcements, low volume, moderated

Low volume (42 posts in 2012, 2100+ members)

Galaxy-User

Questions about using Galaxy and usegalaxy.org

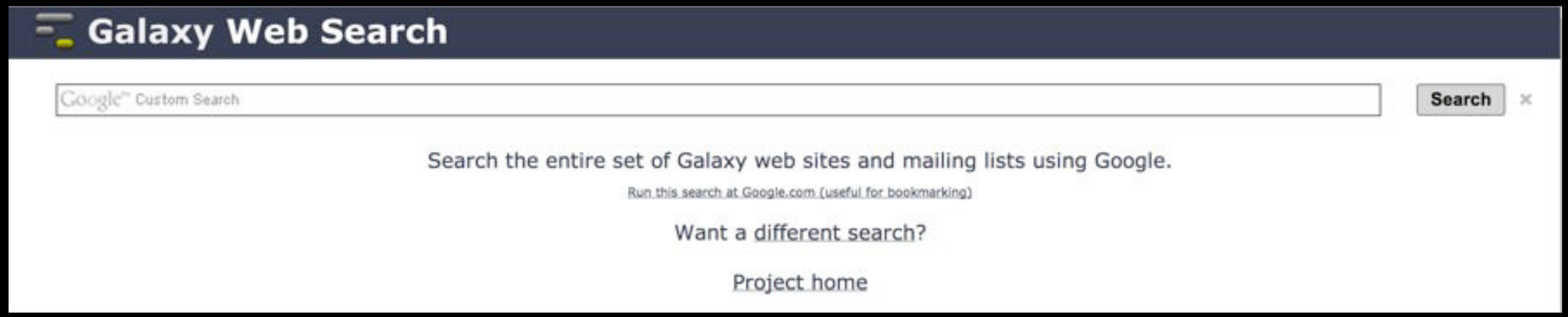
High volume (2900 posts in 2012, 2700+ members)

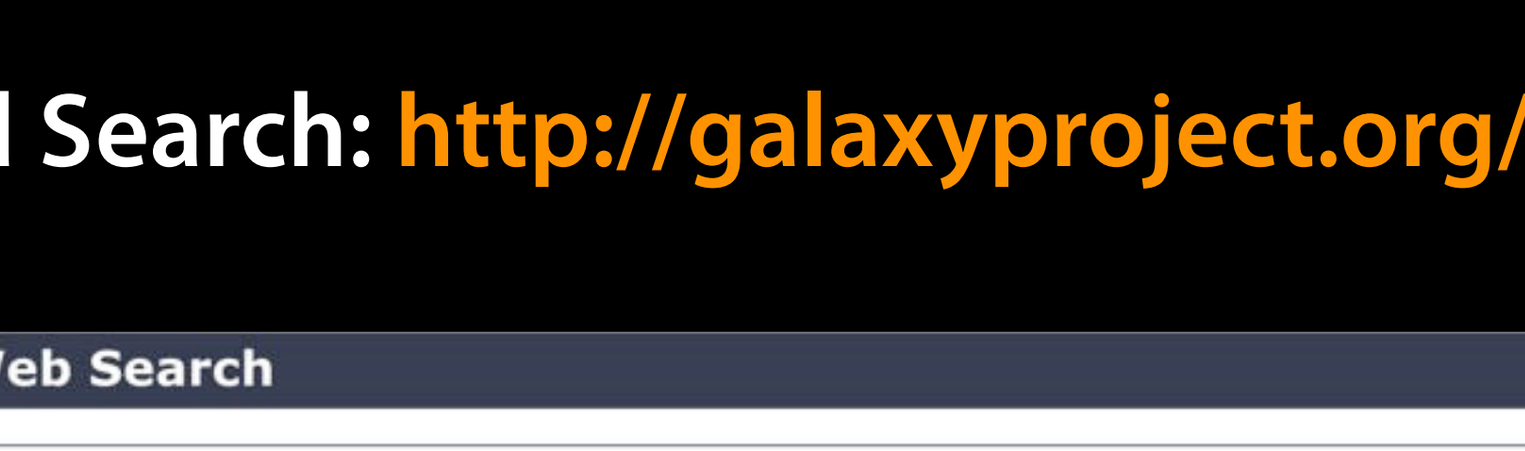
Galaxy-Dev

Questions about developing for and deploying Galaxy

High volume (4500 posts in 2012, 900+ members)

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





Find

Everything on …

Tools for …

Email about …

Source code for …

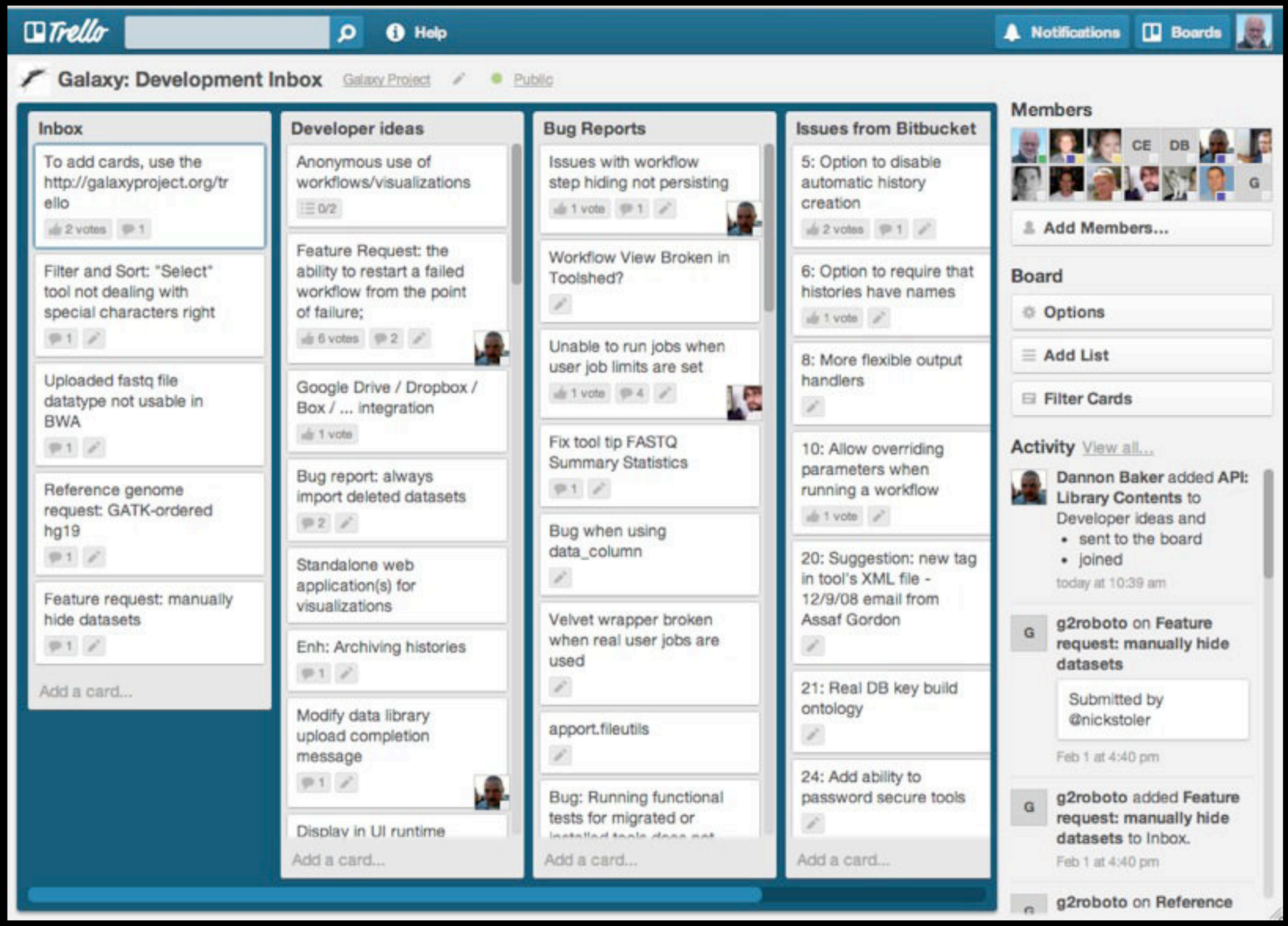
Published Histories, Pages, Workflows, about …

Documentation on …

Papers using Galaxy for …

Related feature requests

Community can create, vote and comment on issues



http://bit.ly/gxyissues

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
- Galaxy Appliance

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 web search

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

SLIPSTREAM APPLIANCE Galaxy made easy

Contribute

Tool Shed • Share
Issues & Requests
Support

Galaxy Project

Home • About
Community
Big Picture

Wiki

Help • All Pages

Events

Galaxy Wiki — DaveClements Settings Logout

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 .

Upcoming Events

Research Triangle Galaxy Workshop Tour

Date	Topic/Event	Venue/Location
July 18-23	*Introduction to Galaxy Workshop* National Institute of Environmental Health Sciences (NIEHS)	
	Introduction to Galaxy Workshop University of North Carolina, Chapel Hill	2013 Research Triangle Wo
	Galaxy Installation Tutorial **2013 GMOD Summer School**	Tour, North Carolina, United
	Introduction to Galaxy Workshop North Carolina State University	
July 19-23	**ISMB/ECCB, BOSC and MS SIG 2013** Talks, posters and workshops. Lots of them.	Berlin, Germany
July 21-25	*Experiences in building a Next-Generation Sequencing Analysis Service using Galaxy, Globus Online, and Amazon Web Services*	**XSEDE13**, San Diego, Calif
	A Sustainable National Gateway for Biological Computation	United States
	Supporting Genomics and other Biological Research	
September 28 - October 1	*Galaxy Workshop*	The Genomic Bioinformatics Workshop, Sydney, Australia
October 1-3	*Galaxy*	**Beyond the Genome 2013**, Francisco, California, United
October 7-8	*TBD*	NGS & Bioinformatics Su
	Using Galaxy to Provide a NGS Analysis Platform	Europe
October 9-11	*Galaxy Training Days*	GenoToul bioinformatics facil INRA, Toulouse Auzeville, Fra
October 22-26	*High Throughput Data Analysis and Visualization with Galaxy*	ASHG 2013, Boston, Massach United States
November 6-12	*Computational and Comparative Genomics Course* Application Deadline: July 15, 2013	Cold Spring Harbor Laborator York, United States

News

Galaxy Wiki — DaveClements Settings Logout | Search:

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

- Galaxy News Briefs
- Galaxy Updates
- Galaxy on Twitter
- Events
- Learn
- Support
- About the Galaxy Project

News Items

New CloudMan Release
SlipStream Appliance: Galaxy Edition
July 2013 Galaxy Update
1000th Galaxy CiteULike Paper
GCC2013 Registration Ends 14 June
June 3, 2013 Galaxy Distribution
June 2013 Galaxy Update
Software Carpentry Boot Camp: Oslo
GCC2013 Early Registration Ends 24 May
Duplicate Accounts on Main

News Archive

News Items

New CloudMan Release

We just released an update to Galaxy CloudMan. CloudMan offers an easy way to get a personal and completely functional instance of Galaxy in the cloud in just a few minutes, without any manual configuration.

CloudMan

IMPORTANT - please read

Any new cluster will automatically start using this version of CloudMan. Existing clusters will be given an option to do an automatic update once the main interface page is refreshed. Note that this upgrade is a major version upgrade and thus the migration is rather complicated. The migration process has been automated but will take a little while to complete. If you have made customizations to your cluster in terms of adding file systems, upgrading the database, or similar, we do not recommend you perform the upgrade. Note that this upgrade comes with (and requires) a new AMI (ami-118bfc78), which will automatically be used when starting an instance via CloudLaunch.

This update brings a large number of updates and new features, the most prominent ones being:

- Unification of galaxyfools and galaxydata file systems into a single galaxy filesystem. This change makes it possible to utilize the Galaxy Tool Shed when installing tools into Galaxy.
- Added initial support for Hadoop-type workloads
- Added initial support for cluster federation via HTCondor
- Added a new file system service for an instance's transient storage, allowing it to be used across the cluster over NFS
- Added a service for the Galaxy Reports webapp
- Added optional Loggly based off-site logging support
- Added tags to all resources utilized by CloudMan

For more details on the new features, see the the CHANGELOG and for even more details see, *all 291 commit messages from 7 contributors*.

Enjoy and please let us know what you think,

Enis Afgan

Posted to the Galaxy News on 2013-07-08

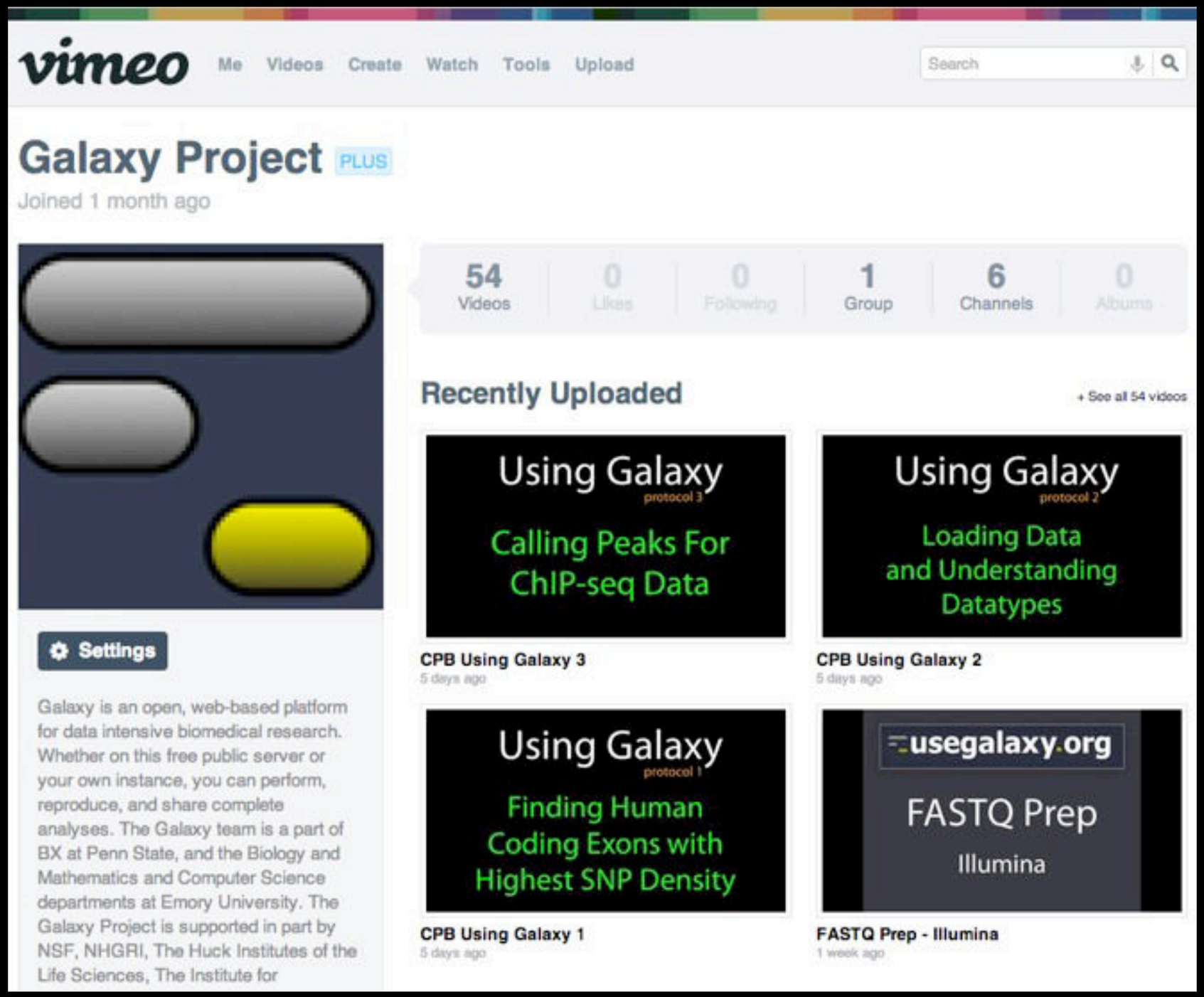
SlipStream Appliance: Galaxy Edition

GALAXY COMMUNITY CONFERENCE

BALTIMORE, MD | JUNE 30 - JULY 2, 2014

http://bit.ly/gcc2014

Galaxy Resources & Community: Videos



"How to" screencasts on using and deploying Galaxy

Talks from previous meetings.

http://vimeo.com/galaxyproject

Galaxy Resources & Community: CiteULike Group

Almost 1200 papers

17 different tags

http://bit.ly/gxycul

Share & Publish: 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

Sharing & Publishing enables **Reproducibility**

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

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)

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)

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
Generic workflow for performing a metagenomic analysis on NGS data.

Accessing the Data

Windshield Splatter datasets analyzed in this manuscript can be accessed through this Galaxy Library. From

About this Page

Author

aun1

Related Pages

All published pages
Published pages by aun1

Rating

Community
(6 ratings, 5.0 average)

Tags

Community:
paper galaxy megan

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

Sharing for Galaxy Administrators Too

Data Libraries

Make data easy to find

Genome Builds

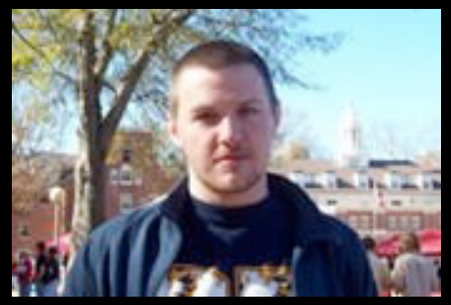
Care about a particular subset of life?

Galaxy Tool Shed

Wrapping tools and datatypes

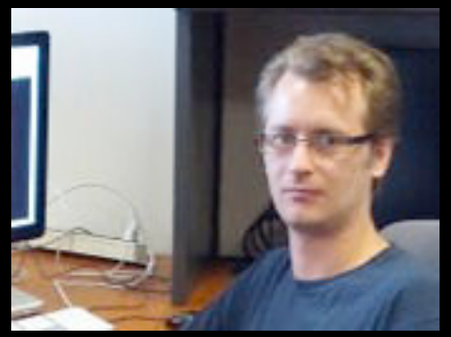
The Galaxy Team

Enis Afgan



Dannon Baker

Dan Blankenberg



Dave Bouvier

Marten Cech

Dave Clements

Nate Coraor

Carl Eberhard

Dorine Francheteau

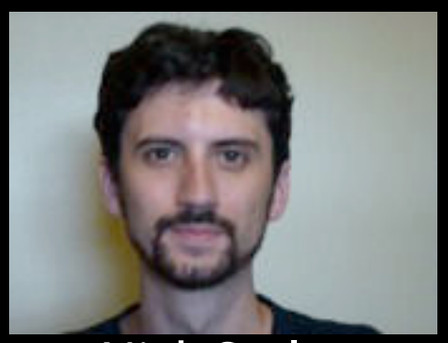
Jeremy Goecks

Sam Guerler

Jen Jackson

Greg von Kuster

Ross Lazarus



Nick Stoler

Anton Nekrutenko

James Taylor

http://wiki.galaxyproject.org/GalaxyTeam

Galaxy is hiring post-docs and software engineers

Please help.

http://wiki.galaxyproject.org/GalaxyIsHiring

Community: Further reading & Resources

http://toolshed.g2.bx.psu.edu
http://bit.ly/gxyServers
http://wiki.galaxyproject.org/MailingLists
http://galaxyproject.org/search
http://bit.ly/gxyissues
http://bit.ly/gcc2014
http://vimeo.com/galaxyproject
http://bit.ly/gxycul

Monday Agenda

9:00	Welcome and Intro
9:40	Basic Analysis with Galaxy
11:00	Break + an Excerise
12:00	Basic Analysis into Reusable Workflows
13:00	Lunch
14:00	NGS Data Quality Control I
15:30	Break
16:00	NGS Data Quality Control II
16:30	Galaxy Community
17:00	Done

Thanks

Dave Clements

Galaxy Project
Emory University

clements@galaxyproject.org