http://bit.ly/glaxy2015slides

Introduction to Galaxy

University of Glasgow
8-9 June 2015

Dave Clements
Galaxy Project
Johns Hopkins University

Mani Mudaliar
Glasgow Polyomics
University of Glasgow

Graham Hamilton
Glasgow Polyomics
University of Glasgow

http://bit.ly/glaxy2015slides

Agenda: Day 2

9:00 Welcome

9:15 RNA-Seq Example Part 1

10:30 Break

10:45 RNA-Seq Example Part 2

12:00 Lunch (on your own)

13:00 SNP & Variant Calling Part 1

15:30 Break

15:45 SNP & Variant Calling Part 2

16:15 Galaxy on the Cloud

17:00 Done

http://bit.ly/glaxy2015slides

Agenda: Day 2

9:00 Welcome

9:15 RNA-Seq Example Part 1

10:30 Break

10:45 RNA-Seq Example Part 2

12:00 Lunch (on your own)

13:00 SNP & Variant Calling Part 1

15:30 Break

15:45 SNP & Variant Calling Part 2

16:15 Galaxy on the Cloud

17:00 Done

http://bit.ly/glaxy2015slides

RNA-Seq Analysis: Get the Data

Create new history

 (cog) → Create New

Import:

Shared Data → Data Libraries → Training → RNA-Seq*

→ Raw Reads

→ Select all

→ Reference

→ Select all

UCDAVIS Genome Center Bioinformatics Core

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

NGS Quality Control *Revisted*

Do QC, or rely on bad data not to map?

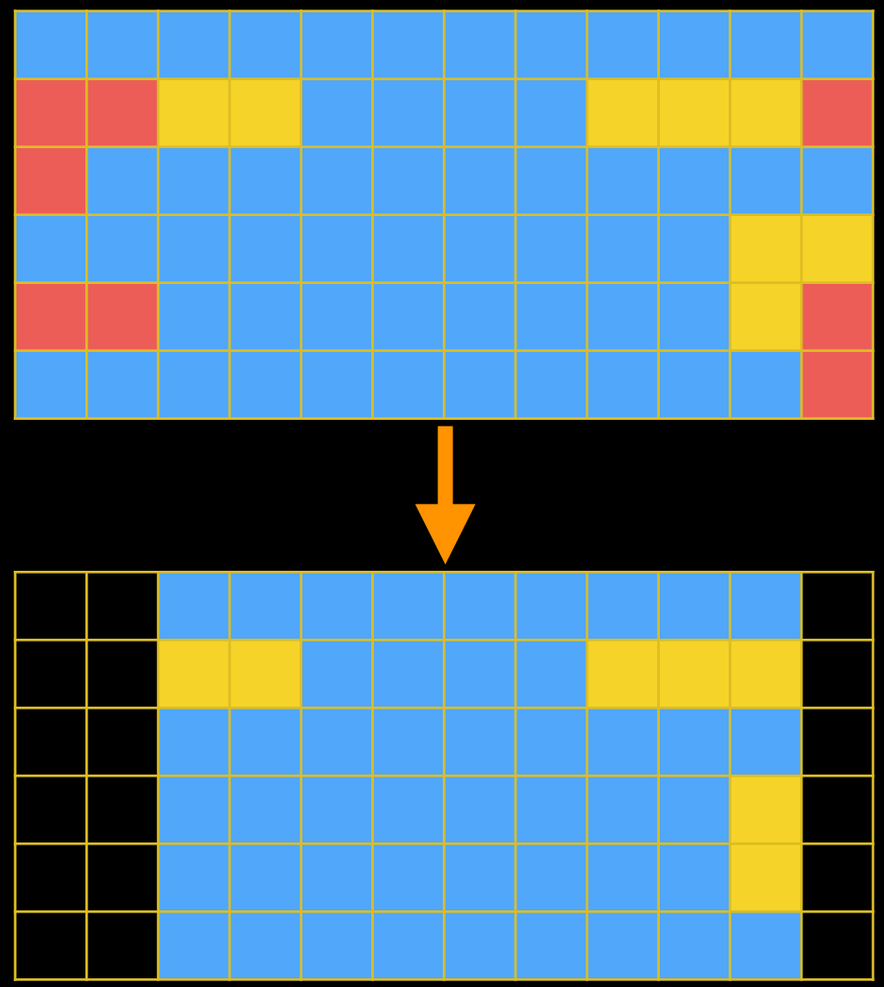
NGS Data Quality: Assessment tools

Same tools available as yesterday:

FastQC, Sliding window, Trimmer by Column, by quality score and length

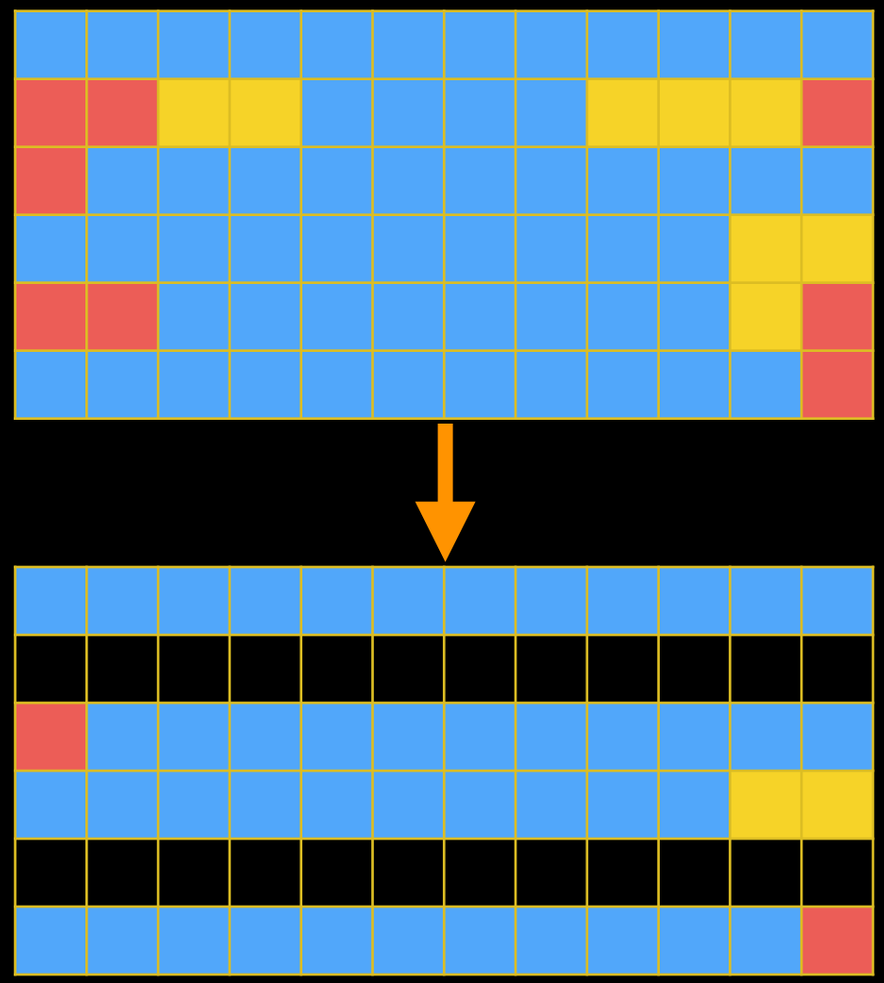
NGS Data Quality: Trim as we see fit

- 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



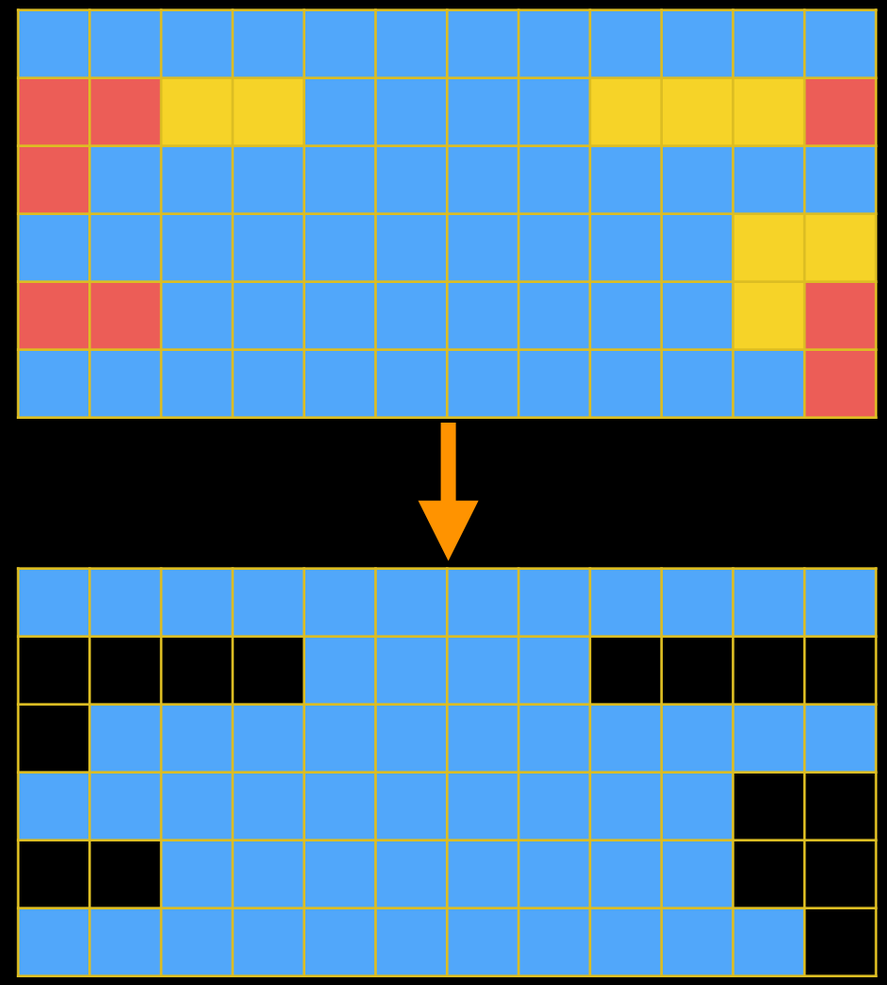
NGS Data Quality: Base Quality Trimming

- ~~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



Trim? *As we see fit?*

- 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

Keeping paired ends paired: Things to Try

- Don't bother.
- Run a workflow (try the "Re-Pair Paired ends after QC may have broken them" workflow on usegalaxy.org) that removes any unpaired reads before mapping:
- Run the Picard Paired Read Mate Fixer after mapping reads.

RNA-Seq Analysis: Restore Pairings

If your QC filters might have broken pairings, then you may want to restore them.

Shared Data → Published Workflows
→ Re-Pair Paired ends after QC may have broken them
→ Import

Workflows
→ Re-Pair Paired ends after QC may have broken them
→ Run

Re-Pair Paired ends after QC may have broken them

Workflow takes 4 inputs

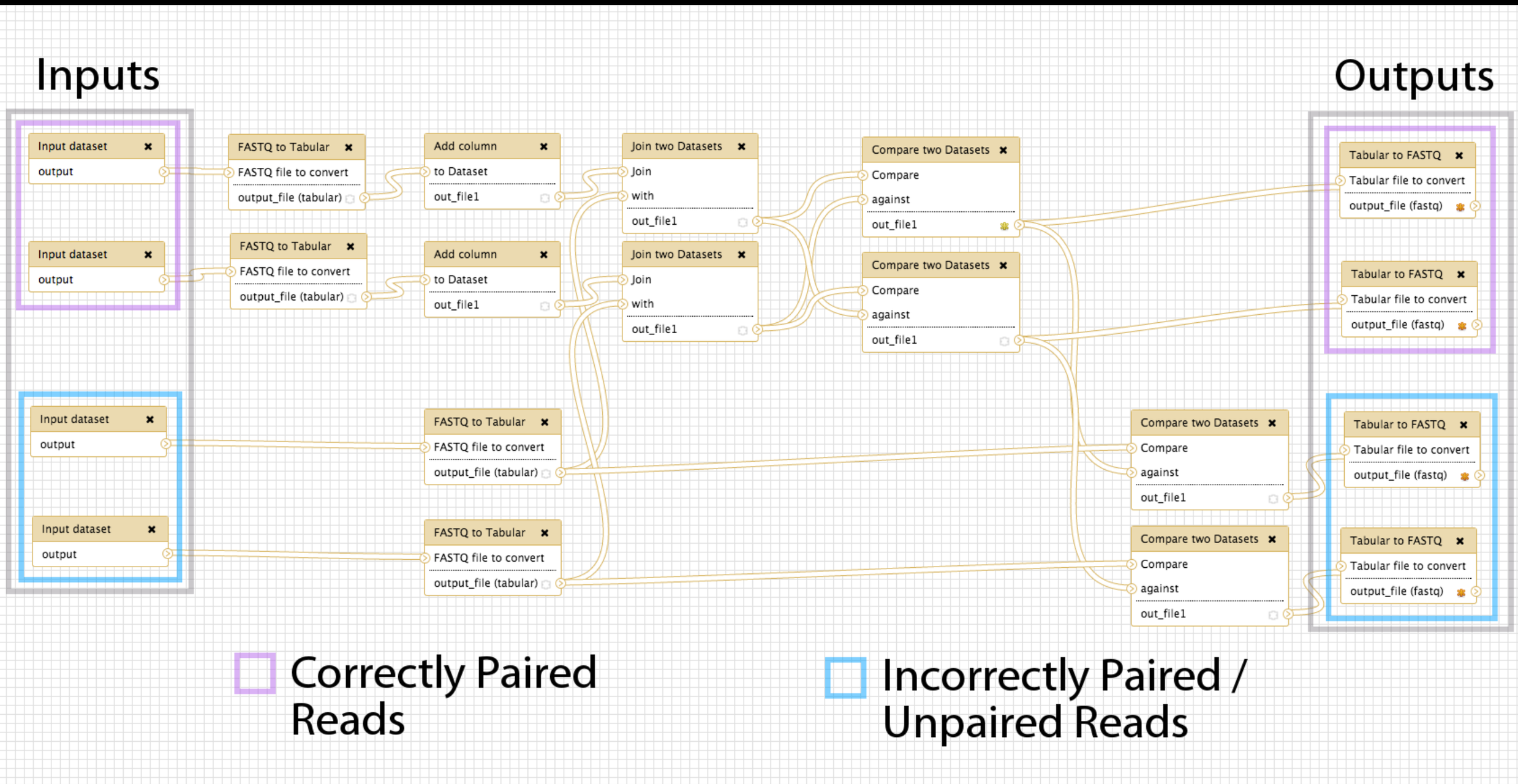
- Forward Reads, before QC
- Reverse Reads, before QC
- Forward Reads, after QC
- Reverse Reads, after QC

And produces 4 outputs

- Forward reads, re-paired
- Reverse reads, re-paired
- Forward reads, singletons
- Reverse reads, singletons

Workflow assumes pre-QC reads are correctly paired

Re-Pair Paired ends after QC may have broken them



NGS Data Quality: Sequencing Artifacts

Repeat this process with MeOH Rep1 R2 (the reverse reads)

... and now we notice a problem in Overrepresented sequences:

Overrepresented sequences

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

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

But this will break pairings (if we still have them).

Or, can rely on mapper to just not map them.

RNA-seq Exercise: Mapping with Tophat2

- Tophat looks for best place(s) to map reads, and best places to insert introns
- *Imagine pages and pages of discussion on the intricacies and pitfalls of RNA-seq mapping here.*

Mapping with Tophat: mean inner distance

Expected distance between paired end reads

- Determined by sample prep
- We'll use 90* for mean inner distance
- We'll use 50 for standard deviation

✳ The library was constructed with the typical Illumina TruSeq protocol, which is supposed to have an average insert size of 200 bases. Our reads are 55 bases (R1) plus 55 bases (R2). So, the Inner Distance is estimated to be 200 - 55 - 55 = 90

From the 2013 UC Davis Bioinformatics Short Course

Mapping with Tophat: Use Existing Annotations?

You can bias Tophat towards known annotations

- Use Own Junctions → Yes
  - Use Gene Annotation → Yes
    - Gene Model Annotation → genes_chr12.gtf
- Use Raw Junctions → Yes (tab delimited file)
- Only look for supplied junctions → Yes

Mapping with Tophat: Make it quicker?

Warning: Here be dragons!

- Allow indel search → No
- Use Coverage Search → No (wee dragons)

TopHat generates its database of possible splice junctions from two sources of evidence. The first and strongest source of evidence for a splice junction is when two segments from the same read (for reads of at least 45bp) are mapped at a certain distance on the same genomic sequence or when an internal segment fails to map - again suggesting that such reads are spanning multiple exons. With this approach, "GT-AG", "GC-AG" and "AT-AC" introns will be found *ab initio*. The second source is pairings of "coverage islands", which are distinct regions of piled up reads in the initial mapping. Neighboring islands are often spliced together in the transcriptome, so TopHat looks for ways to join these with an intron. We only suggest users use this second option (--coverage-search) for short reads (< 45bp) and with a small number of reads (<= 10 million). This latter option will only report alignments across "GT-AG" introns

TopHat Manual

Mapping with Tophat: Max # of Alignments Allowed

Some reads align to more than one place equally well.

For such reads, how many should Tophat include?

If more than the specified number, Tophat will pick those with the best mapping score.

Tophat breaks ties randomly.

Tophat assigns equal fractional credit to all *n* mappings

Instructs TopHat to allow up to this many alignments to the reference for a given read, and choose the alignments based on their alignment scores if there are more than this number. The default is 20 for read mapping. Unless you use --report-secondary-alignments, TopHat will report the alignments with the best alignment score. If there are more alignments with the same score than this number, TopHat will randomly report only this many alignments. In case of using --report-secondary-alignments, TopHat will try to report alignments up to this option value, and TopHat may randomly output some of the alignments with the same score to meet this number.

TopHat Manual

Mapping With Tophat: Cleanup

Use only the good stuff!

NGS BAM Tools → Filter

Mapping Quality → **>=20**

Insert Filter → **isProperPair: Yes**

Insert Filter → **reference: chr12**

RNA-Seq Mapping With Tophat: Resources

RNA-Seq Concepts, Terminology, and Work Flows
by Monica Britton

Aligning PE RNA-Seq Reads to a Genome
by Monica Britton

both from the UC Davis 2013 Bioinformatics Short Course

RNA-Seq Analysis with Galaxy
by Jeroen F.J. Laros, Wibowo Arindrarto, Leon Mei

from the GCC2013 Training Day

RNA-Seq Analysis with Galaxy
by Curtis Hendrickson, David Crossman, Jeremy Goecks

from the GCC2012 Training Day

Agenda: Day 2

9:00 Welcome

9:15 RNA-Seq Example Part 1

10:30 Break

10:45 RNA-Seq Example Part 2

12:00 Lunch (on your own)

13:00 SNP & Variant Calling Part 1

15:30 Break

15:45 SNP & Variant Calling Part 2

16:15 Galaxy on the Cloud

17:00 Done

http://bit.ly/glaxy2015slides

Agenda: Day 2

9:00 Welcome

9:15 RNA-Seq Example Part 1

10:30 Break

10:45 RNA-Seq Example Part 2

12:00 Lunch (on your own)

13:00 SNP & Variant Calling Part 1

15:30 Break

15:45 SNP & Variant Calling Part 2

16:15 Galaxy on the Cloud

17:00 Done

http://bit.ly/glaxy2015slides

RNA-Seq Analysis: Transcript Prediction

Import:

Shared Data → Data Libraries → Training → RNA-Seq*

→ Mapped Filtered Reads

→ Select all

UCDAVIS Genome Center Bioinformatics Core

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

RNA-Seq Analysis: Transcript Prediction

Run StringTie

NGS RNA Analysis → StringTie

Use GFF to guide? → No

Use defaults

RNA-Seq Analysis: Transcript Prediction

Run Cuffmerge

NGS RNA Analysis → Cuffmerge

Use reference Annotation? → Yes

Use sequence data? → Yes

Source for reference List → History, chr12.fa

Cuffdiff

- Part of the Tuxedo RNA-Seq Suite (as are Tophat, Bowtie, StringTie, Cufflinks, Cuffmerge, ...)
- Identifies differential expression between multiple datasets
- Widely used and widely installed on Galaxy

NGS: RNA Analysis → Cuffdiff

Cuffdiff

Cuffdiff previously used FPKM/RPKM as central statistic.
Total # mapped reads heavily influences FPKM/RPKM.
Can lead to challenges when you have very highly expressed genes in the mix.

Cuffdiff

- Which Transcript definitions to use?
  - Official (genes_chr12.gtf in our case)
  - MeOH or R3G Cufflinks transcripts
  - Results of Cuffmerge on MeOH & R3G Cufflinks transcripts
- Depends on what you care about
  - I'll use Cuffmerge on MeOH & R3G Cufflinks transcripts

Cuffdiff

- Running with 2 Groups: MeOH and R3G
- Each group has 3 replicates each

Cuffdiff

Produces many output files, all explained in doc

We'll focus on gene differential expression testing

test_id	gene_id	gene	locus	sample_1	sample_2	status	value_1	value_2	log2(fold_change)	test_stat	p_value	q_value	significant
A2M	A2M	A2M	chr12:9217772-9268558	MeOH	R3G	NOTEST	3.32147	3.13694	-0.0824644	0	1	1	no
A2M-AS1	A2M-AS1	A2M-AS1	chr12:9217772-9268558	MeOH	R3G	NOTEST	7.45797	13.9413	0.902515	0	1	1	no
A2ML1	A2ML1	A2ML1	chr12:8975149-9029381	MeOH	R3G	NOTEST	4.83055	7.79884	0.691072	0	1	1	no
A2MP1	A2MP1	A2MP1	chr12:9381128-9386803	MeOH	R3G	NOTEST	2.49656	0	-inf	0	1	1	no
AAAS	AAAS	AAAS	chr12:53701239-53715412	MeOH	R3G	OK	269.035	159.23	-0.756683	-2.22857	0.0005	0.00194017	yes
AACS	AACS	AACS	chr12:125549924-125627871	MeOH	R3G	NOTEST	29.2933	35.0339	0.258178	0	1	1	no
ABCB9	ABCB9	ABCB9	chr12:123405497-123451056	MeOH	R3G	NOTEST	4.68869	1.7732	-1.40283	0	1	1	no
ABCC9	ABCC9	ABCC9	chr12:21950323-22089628	MeOH	R3G	OK	553.247	487.261	-0.18323	-2.02806	0.0004	0.00162143	yes
ABCD2	ABCD2	ABCD2	chr12:39945021-40013843	MeOH	R3G	OK	86.1377	172.795	1.00435	4.3436	5e-05	0.000246739	yes
ACACB	ACACB	ACACB	chr12:109577201-109706030	MeOH	R3G	NOTEST	8.45306	15.5772	0.881885	0	1	1	no
ACAD10	ACAD10	ACAD10	chr12:112123856-112194911	MeOH	R3G	NOTEST	21.8237	27.8326	0.350882	0	1	1	no
ACADS	ACADS	ACADS	chr12:121163570-121177811	MeOH	R3G	NOTEST	38.644	16.1739	-1.25658	0	1	1	no
ACRBP	ACRBP	ACRBP	chr12:6747241-6756580	MeOH	R3G	NOTEST	2.96987	3.26939	0.138621	0	1	1	no
ACSM4	ACSM4	ACSM4	chr12:7456927-7480969	MeOH	R3G	NOTEST	0	0	0	0	1	1	no
ACSS3	ACSS3	ACSS3	chr12:81471808-81649582	MeOH	R3G	NOTEST	0	0	0	0	1	1	no
ACTR6	ACTR6	ACTR6	chr12:100593864-100618202	MeOH	R3G	OK	475.594	421.324	-0.174799	-0.797581	0.1588	0.258406	no
ACVR1B	ACVR1B	ACVR1B	chr12:52345450-52390863	MeOH	R3G	NOTEST	32.5737	38.3075	0.233922	0	1	1	no
ACVRL1	ACVRL1	ACVRL1	chr12:52301201-52317145	MeOH	R3G	NOTEST	1.27713	2.16161	0.759201	0	1	1	no
ADAM1A	ADAM1A	ADAM1A	chr12:112336866-112339706	MeOH	R3G	NOTEST	30.0162	55.2154	0.879331	0	1	1	no
ADAMTS20	ADAMTS20	ADAMTS20	chr12:43748011-43945724	MeOH	R3G	NOTEST	0.453322	0.502067	0.147346	0	1	1	no
ADCY6	ADCY6	ADCY6	chr12:49159974-49182820	MeOH	R3G	NOTEST	9.32722	17.6743	0.922135	0	1	1	no
ADIPOR2	ADIPOR2	ADIPOR2	chr12:1800246-1897845	MeOH	R3G	OK	207.468	179.333	-0.210248	-1.02392	0.09	0.158988	no
AEBP2	AEBP2	AEBP2	chr12:19592607-19675173	MeOH	R3G	OK	143.039	128.293	-0.156957	-0.688267	0.2254	0.344537	no
AGAP2	AGAP2	AGAP2	chr12:58118075-58135944	MeOH	R3G	OK	98.2385	116.302	0.243511	0.935119	0.11475	0.198086	no
AICDA	AICDA	AICDA	chr12:8754761-8765442	MeOH	R3G	NOTEST	78.1514	63.4313	-0.301077	0	1	1	no
AKAP3	AKAP3	AKAP3	chr12:4724675-4754343	MeOH	R3G	NOTEST	6.12385	7.89626	0.366731	0	1	1	no
ALDH1L2	ALDH1L2	ALDH1L2	chr12:105413561-105478341	MeOH	R3G	NOTEST	7.11374	8.11722	0.190377	0	1	1	no
ALDH2	ALDH2	ALDH2	chr12:112204690-112247789	MeOH	R3G	NOTEST	12.8033	8.05635	-0.668321	0	1	1	no
ALG10	ALG10	ALG10	chr12:34175215-34181236	MeOH	R3G	NOTEST	54.8575	59.3459	0.11346	0	1	1	no
ALG10B	ALG10B	ALG10B	chr12:38710556-38723528	MeOH	R3G	NOTEST	43.8157	63.0457	0.524952	0	1	1	no
ALKBH2	ALKBH2	ALKBH2	chr12:109525992-109531293	MeOH	R3G	OK	679.517	297.183	-1.19316	-3.34255	5e-05	0.000246739	yes
ALX1	ALX1	ALX1	chr12:85674035-85695561	MeOH	R3G	NOTEST	0	0	0	0	1	1	no

Cuffdiff: differentially expressed genes

Column	Contents
test_stat	value of the test statistic used to compute significance of the observed change in FPKM
p_value	Uncorrected P value for test statistic
q_value	FDR-adjusted p-value for the test statistic
status	Was there enough data to run the test?
significant	and, was the gene differentially expressed?

Cuffdiff

- Column 7 (“status”) can be FAIL, NOTEST, LOWDATA or OK
  - Filter and Sort → Filter
    - c7 == 'OK'
- Column 14 (“significant”) can be yes or no
  - Filter and Sort → Filter
    - c14 == 'yes'

Returns the list of genes with

1) enough data to make a call, and

2) that are called as differentially expressed.

Cuffdiff: Next Steps

Try running Cuffdiff with different normalization and dispersion estimation methods.

Compare the differentially expressed gene lists.

Which settings have what type of impacts on the results?

Are there any patterns to the identified genes?

RNA-Seq Differential Expression with Cuffdiff: Resources

RNA-Seq Concepts, Terminology, and Work Flows

by Monica Britton

from the UC Davis 2013 Bioinformatics Short Course

RNA-Seq Analysis with Galaxy

by Jeroen F.J. Laros, Wibowo Arindrarto, Leon Mei

from the GCC2013 Training Day

RNA-Seq Analysis with Galaxy

by Curtis Hendrickson, David Crossman, Jeremy Goecks

from the GCC2012 Training Day

Agenda: Day 2

9:00 Welcome

9:15 RNA-Seq Example Part 1

10:30 Break

10:45 RNA-Seq Example Part 2

12:00 Lunch (on your own)

13:00 SNP & Variant Calling Part 1

15:30 Break

15:45 SNP & Variant Calling Part 2

16:15 Galaxy on the Cloud

17:00 Done

http://bit.ly/glaxy2015slides

Agenda: Day 2

9:00 Welcome

9:15 RNA-Seq Example Part 1

10:30 Break

10:45 RNA-Seq Example Part 2

12:00 Lunch (on your own)

13:00 SNP & Variant Calling Part 1

15:30 Break

15:45 SNP & Variant Calling Part 2

16:15 Galaxy on the Cloud

17:00 Done

http://bit.ly/glaxy2015slides

Agenda: Day 2

9:00 Welcome

9:15 RNA-Seq Example Part 1

10:30 Break

10:45 RNA-Seq Example Part 2

12:00 Lunch (on your own)

13:00 SNP & Variant Calling Part 1

15:30 Break

15:45 SNP & Variant Calling Part 2

16:15 Galaxy on the Cloud

17:00 Done

http://bit.ly/glaxy2015slides

Galaxy is available ...

http://aws.amazon.com/education
http://globus.org/
http://wiki.galaxyproject.org/Cloud

AWS in Education Grants Program

http://aws.amazon.com/education

What is our path?

Today we will:

- Launch our own Galaxy server on AWS
- Make the server dynamically scalable in response to demand.
- Run some basic analysis on it.
- Make it go away.

Full Disclosure

To use AWS you must create an AWS account with a credit card associated with it.

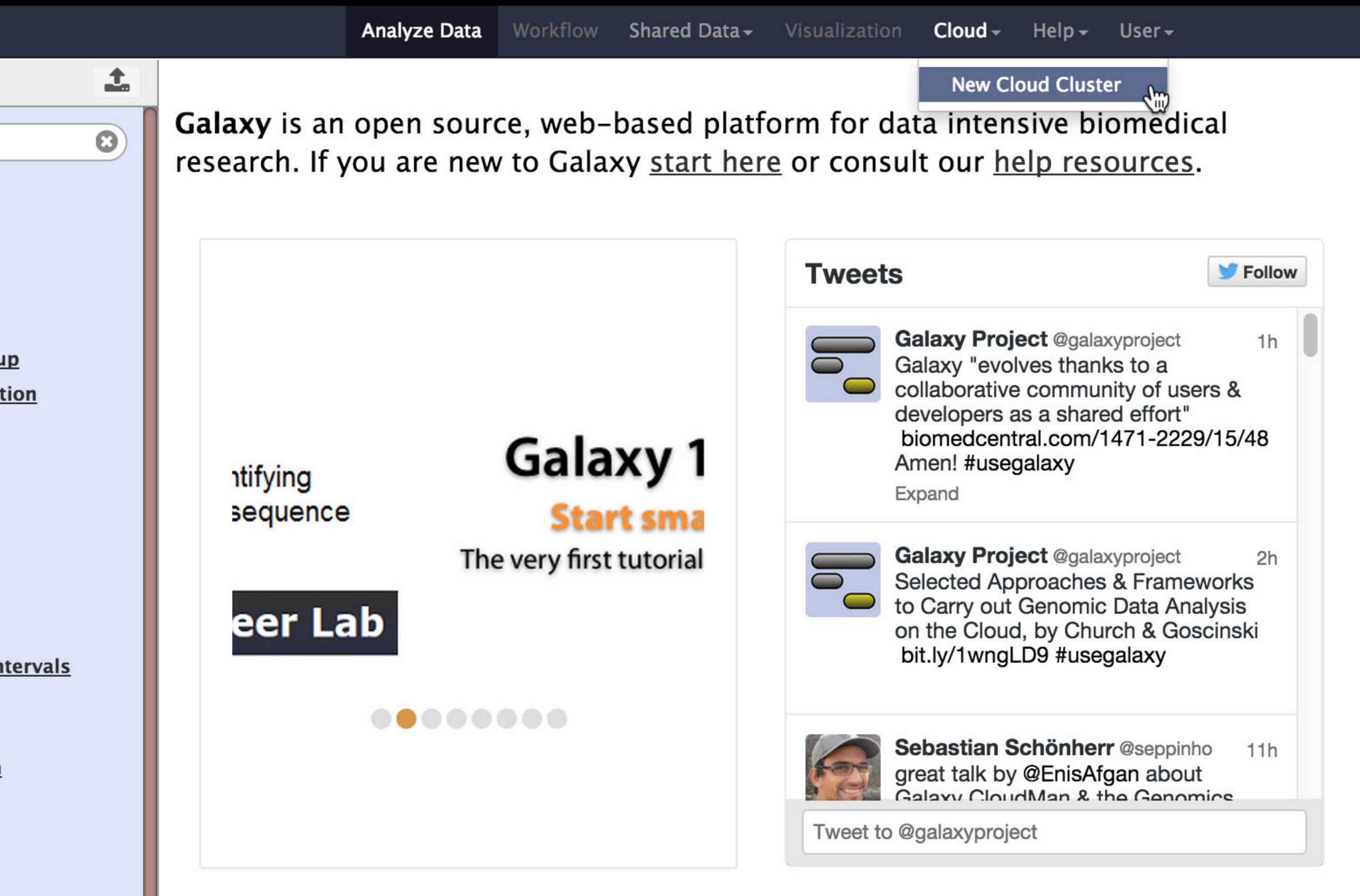
You must also have created a key pair.

We will use the IAM account for this workshop.

IAM Accounts

Imagine, a link to a list of accounts, and credentials, here.

CloudLaunch: From UseGalaxy.org



CloudLaunch

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

Launch a Galaxy Cloud Instance

To launch a Galaxy Cloud Cluster, enter your AWS Secret Key ID, and Secret Key. Galaxy will use these to present appropriate options for launching your cluster. Note that using this form to launch computational resources in the Amazon Cloud will result in costs to the account indicated above. See Amazon's pricing for more information.

Key ID

This is the text string that uniquely identifies your account, found in the Security Credentials section of the AWS Console.

Secret Key

This is your AWS Secret Key, also found in the Security Credentials section of the AWS Console.

CloudLaunch

Launch a Galaxy Cloud Instance

To launch a Galaxy Cloud Cluster, enter your AWS Secret Key ID, and Secret Key. Galaxy will use these to present appropriate options for launching your cluster. Note that using this form to launch computational resources in the Amazon Cloud will result in costs to the account indicated above. See Amazon's pricing for more information.

Key ID

This is the text string that uniquely identifies your account, found in the Security Credentials section of the AWS Console.

Secret Key

This is your AWS Secret Key, also found in the Security Credentials section of the AWS Console.

Instances in your account

New Cluster

Cluster Name

This is the name for your cluster. You'll use this when you want to restart.

Cluster Password

Cluster Password - Confirmation

Key Pair

cloudman_keypair

Instance Type

Compute optimized Large (2 vCPU/4GB RAM)

Requesting the instance may take a moment, please be patient. Do not refresh your browser or navigate away from the page

Submit

CloudLaunch

https://usegalaxy.org/cloudlaunch

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

Launch a Galaxy Cloud Instance

Launch Pending, please be patient.

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

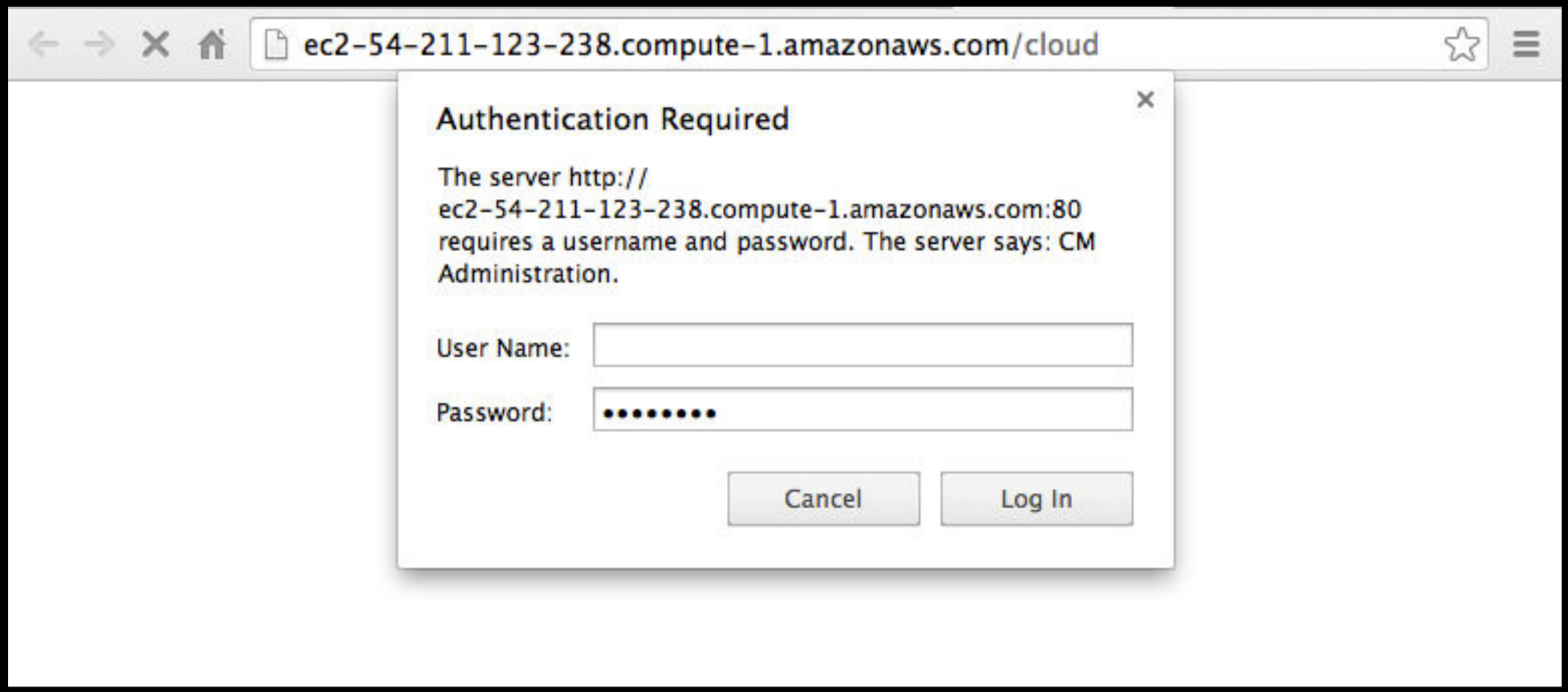
Launch a Galaxy Cloud Instance

Access Information

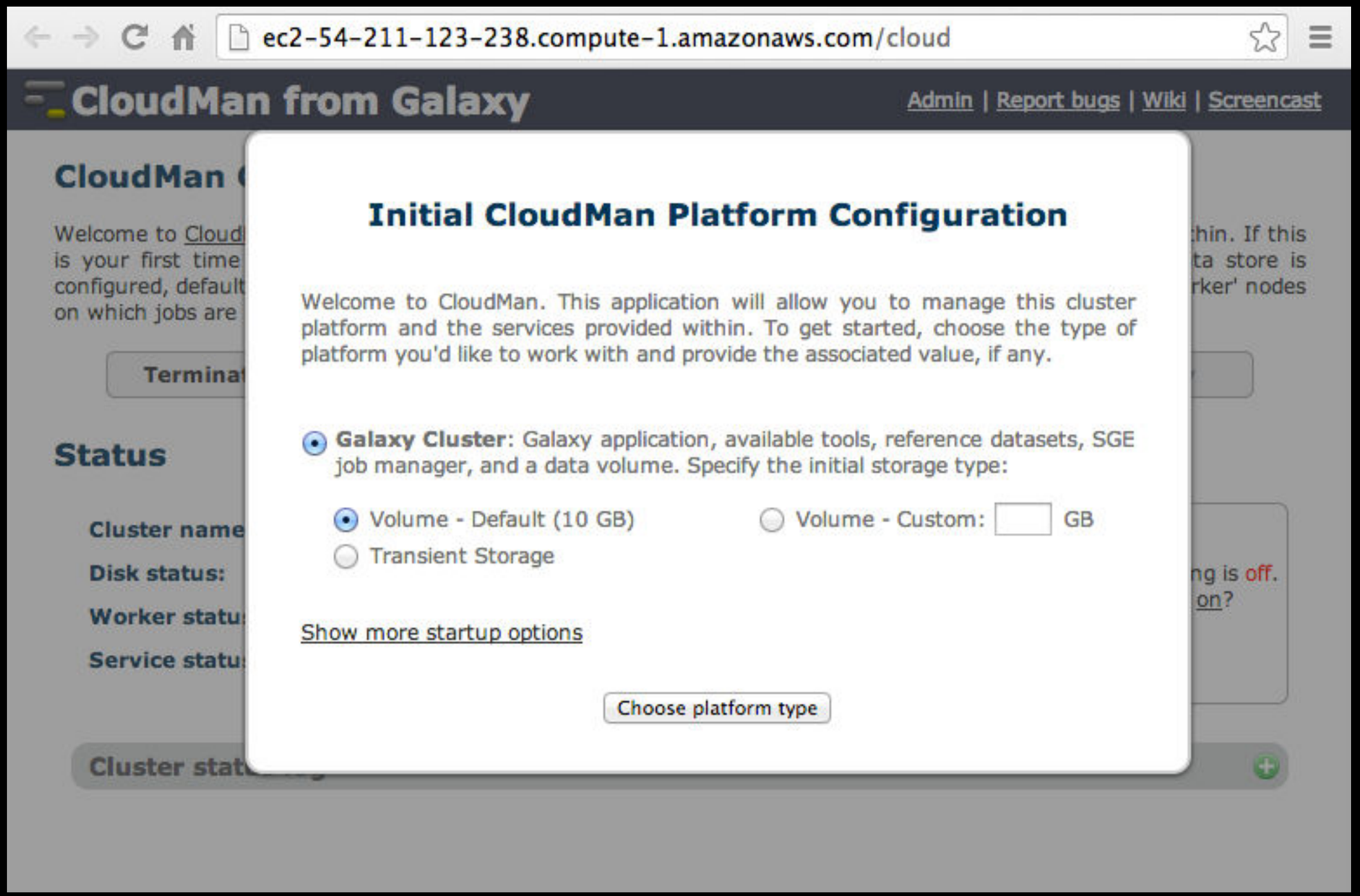
Your instance 'i-61503e9b' has been successfully launched using the 'ami-a7dbf6ce' AMI.
While it may take a few moments to boot, you will be able to access the cloud control panel at ec2-54-196-164-110.compute-1.amazonaws.com/cloud.
SSH access is also available using your private key. From the terminal, you would execute something like:

```
ssh -i cloudman_key_pair.pem ubuntu@ec2-54-196-164-110.compute-1.amazonaws.com
```

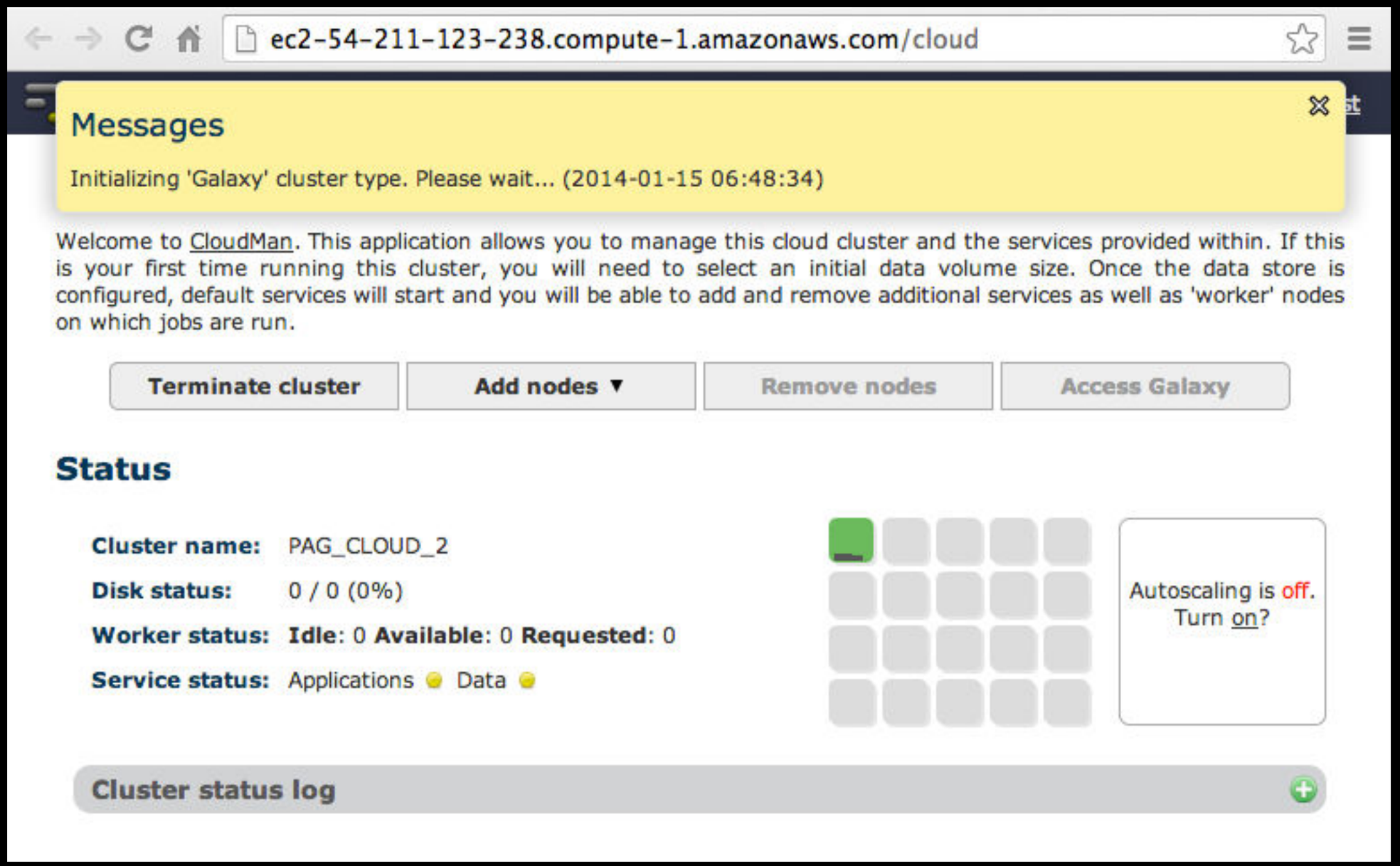
CloudLaunch



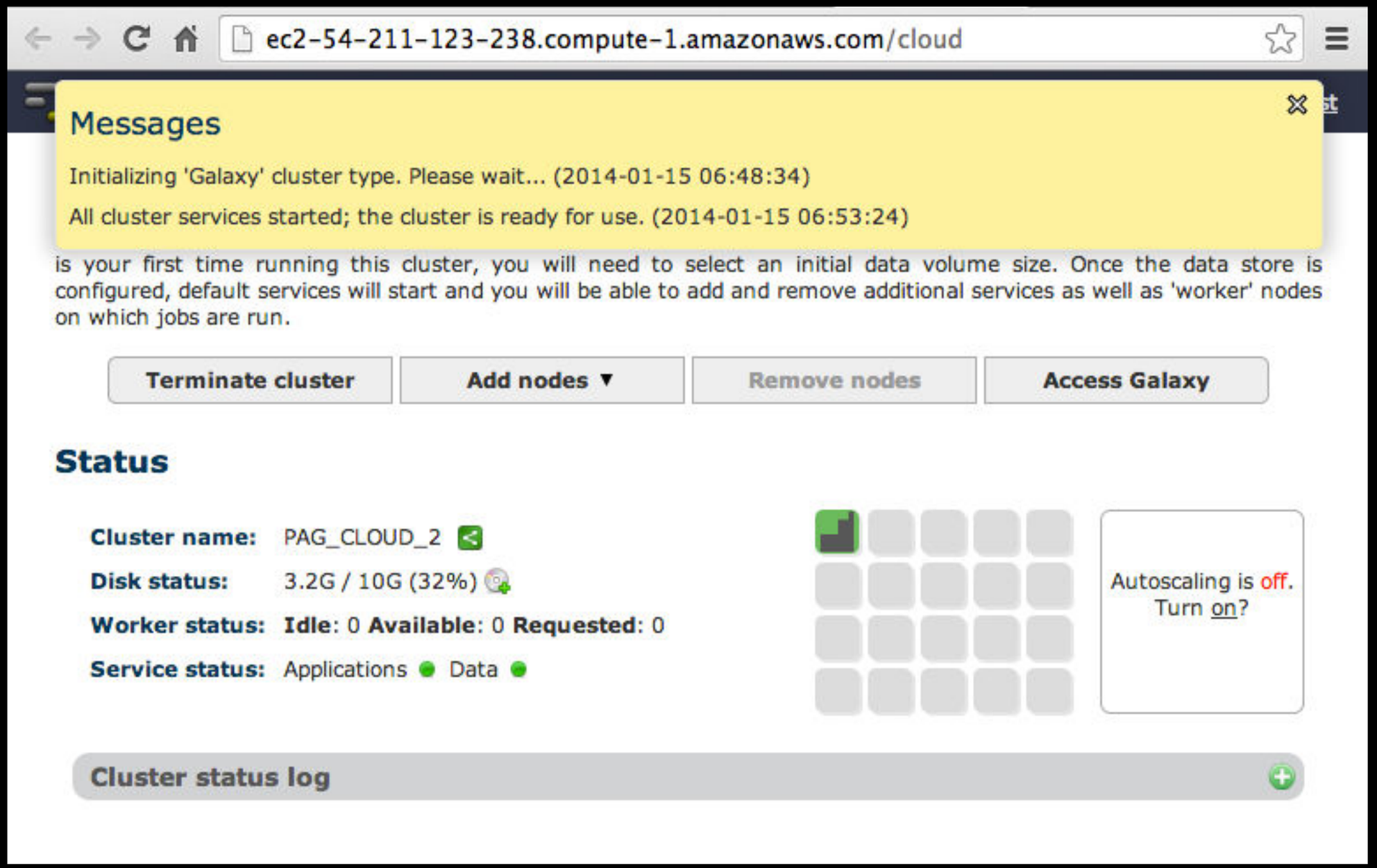
CloudLaunch



CloudLaunch



Cloud Launched



Cool things to do

- Create a login
- Become an admin
- Set up autoscaling
- Run ~ Galaxy 101
  - http://usegalaxy.org/galaxy101
- Shut it down

Agenda: Day 2

9:00 Welcome

9:15 RNA-Seq Example Part 1

10:30 Break

10:45 RNA-Seq Example Part 2

12:00 Lunch (on your own)

13:00 SNP & Variant Calling Part 1

15:30 Break

15:45 SNP & Variant Calling Part 2

16:15 Galaxy on the Cloud

17:00 Done (Almost)

http://bit.ly/glaxy2015slides

Feedback: We *Need* It!

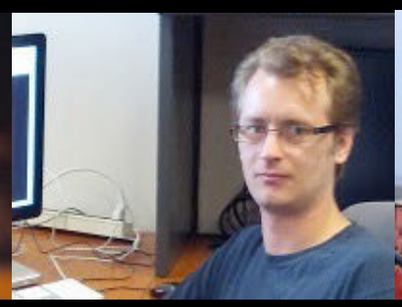
http://bit.ly/glaxy2015feedback

The Galaxy Team

Enis Afgan

Dannon Baker

Dan Blankenberg



Dave Bouvier

Marten Cech

John Chilton

Dave Clements

Nate Coraor

Carl Eberhard

Jeremy Goecks

Sam Guerler

Jen Jackson

Ross Lazarus

Anton Nekrutenko

Nick Stoler

James Taylor

Nitesh Turaga

http://wiki.galaxyproject.org/GalaxyTeam

http://bit.ly/glaxy2015feedback

Galaxy is hiring post-docs and software engineers

Please help.

http://wiki.galaxyproject.org/GalaxyIsHiring

Also thanks to

Mani Mudaliar

Graham Hamilton

Amy Cattanach

http://bit.ly/glaxy2015feedback

http://bit.ly/glaxy2015slides

Agenda: Day 2

9:00 Welcome

9:15 RNA-Seq Example Part 1

10:30 Break

10:45 RNA-Seq Example Part 2

12:00 Lunch (on your own)

13:00 SNP & Variant Calling Part 1

15:30 Break

15:45 SNP & Variant Calling Part 2

16:15 Galaxy on the Cloud

17:00 Done

http://bit.ly/glaxy2015feedback

http://bit.ly/glaxy2015slides

Thanks

Dave Clements

Galaxy Project

Johns Hopkins University

clements@galaxyproject.org

http://bit.ly/glaxy2015feedback

http://bit.ly/glaxy2015slides

ChIP-Seq Analysis: Replicates

Shared Data → Data Libraries → Training → ChIP-Seq →

MACS Outputs → Peaks in BED format

Import files for

Nanog Rep 2

Pou5f1 Rep 1

Pou5f1 Rep 2

ChIP-Seq Analysis: Unify Replicates

Operate on Genomic Intervals → Concatenate

Concatenate Nanog Rep 1 and 2 peak files

Operate on Genomic Intervals → Cluster

Use default parameters

Rename the output dataset

Add the Nanog cluster output to your visualization

ChIP-Seq Analysis: Unify Replicates

Repeat for Pou5f1 replicates

Operate on Genomic Intervals → Concatenate

Concatenate Pou5f1 Rep 1 and 2 Peak files

Operate on Genomic Intervals → Cluster

Use default parameters

Rename the output dataset

Add the Pou5f1 cluster output to your visualization

ChIP-Seq Analysis: Differential binding

Operate on Genomic Intervals → Subtract

First dataset clustered → Pou5f1

Second dataset clustered → Nanog

Return → Intervals with no overlap