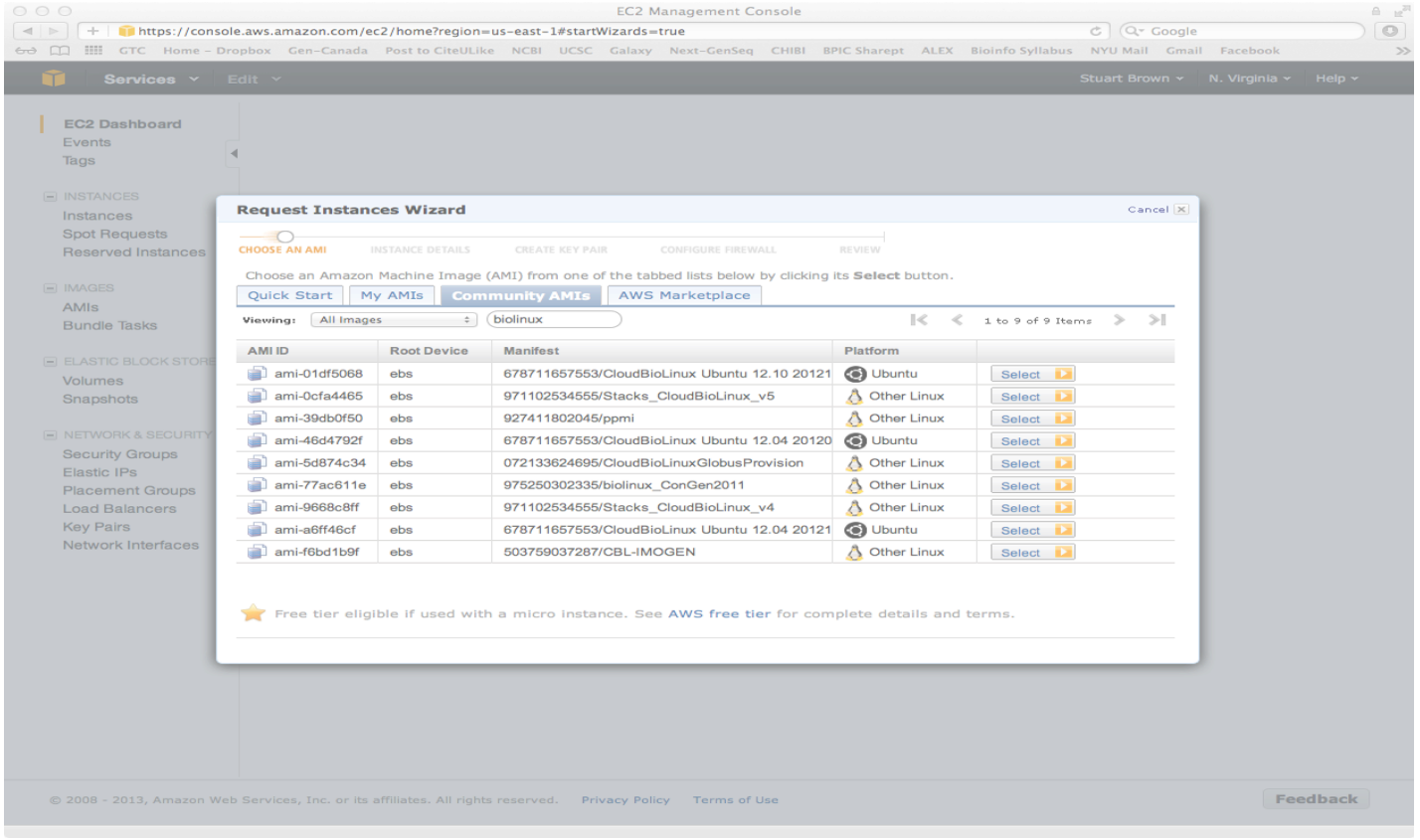
Using Cloud Computing for RNA-seq



Stuart M. Brown, Ph.D.

Center for Health Informatics & Bioinformatics

NYU School of Medicine

NYU Langone MEDICAL CENTER

RNA-seq Measures Gene Expression

- Takes advantage of the rapidly dropping cost of Next-Generation DNA sequencing
- Measures gene expression in true genome-wide fashion (all the RNA)
- Also enables detection of mutations (SNPs), alternative splicing, allele specific expression, and fusion genes
- More accurate and better dynamic range than Microarray
- Can be used to detect miRNA, ncRNA, and other non-coding RNA

RNA-seq is very compute intensive

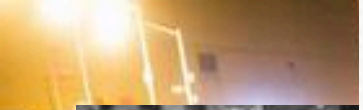
- Billions of reads
- Large file sizes (tens of GB)
- Alignment to complete reference genomes
- Spliced alignment
- Like most genomics research institutes, NYU has purchases substantial High Performance Computing (HPC) resources to support our NGS lab.
    - Cluster of servers
    - Machines with large amount of RAM
    - Data storage and backup system

Cloud = Renting Computers

- Instead of buying a High Performance Computing system, rent time on one from a vendor
- Amazon EC2 has simplified this process
- Scalable: Pay just for the computing you need, only when you need it.
- Also has benefits to move and share data among many users at different institutions with different security policies


NYULMC 1st Ave, NYC


13 foot storm surge at NYUMC





AHEAD
301

Computers Destroyed > go to Cloud

- All HPC computers used for NGS at NYUMC were destroyed by flood water
- Off site data backup was secure (!)
- NGS wet laboratory also damaged by flood
- By mid-November, NYUMC resumed sequencing using outsourced labs (NY Genome Center, MSKCC)
- We established a **cloud** server for NGS alignment, RNA-seq, data exchange with sequencing labs and investigators scattered all over NY City.

Amazon EC2

- Makes the use of Cloud computing easy
- Pre-configured “Amazon Machine Images” with useful sets of software installed.
- Can customize and save your own AMI’s
- Large data storage on S3
- Active data (reference genomes, sample sheets, configuration files) storage on EBS volumes.
- Can also use public data stored in “snapshots”

CloudBioLinux

- We used AMI's created by the (JCVI) CloudBioLinux project: http://cloudbiolinux.org

(as well as some generic Linux AMI's configured with our own software choices)

- Most of the popular NGS software is pre-installed (fastqc, bwa, bowtie, samtools, picard, tophat/cufflinks, etc.)

- Good instructions:
  "Getting started with CloudBioLinux", Bela Tiwari, 2011

https://github.com/chapmanb/cloudbiolinux/blob/master/doc/intro/gettingStarted_CloudBioLinux.pdf?raw=true

Using Cloud Computing Infrastructure with CloudBioLinux, CloudMan, and Galaxy.

Afgan et al. Curr Protoc Bioinform. 38: 11.9.1-11.9.20 (2012).

http://onlinelibrary.wiley.com/doi/10.1002/0471250953.bi1109s38/full

- *# List of custom software installed*

bio_general:
- emboss
- pgdspider

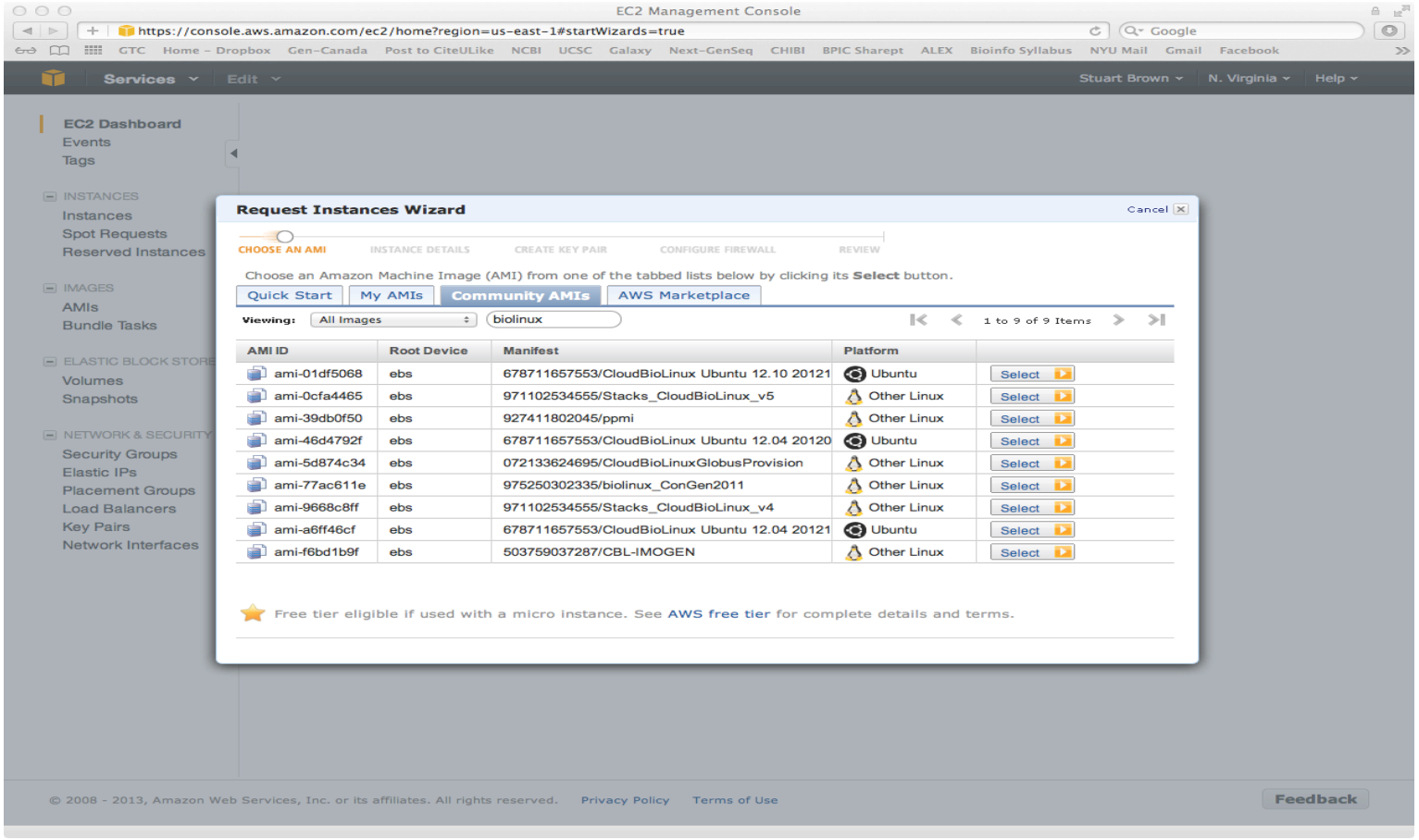
bio_nextgen:
- alignment:
  - bwa
  - bowtie
  - bowtie2
  - gmap
  - lastz
  - mosaik
  - snap
  - stampy
- utilitites:
  - bamtools
  - bedtools
  - cram
  - dwgsim
  - fastqc
  - fastq_screen
  - fastx_toolkit
  - ogap
  - *# - solexaqa*
  - varianttools
  - vcftools
- analysis:
  - cufflinks
  - picard
  - samtools
  - sambamba
  - shrec
  - tophat
  - vep

- assembly:
  - abyss
  - cortex_var
  - ray
  - transabyss
  - trinity
  - velvet
- sv:
  - hydra
- variant:
  - bcbio_variation
  - crisp
  - freebayes
  - gatk
  - gemini
  - grabix
  - tabix
  - tassel
  - snpeff
  - stacks
  - varscan
  - vcflib
- chip:
  - macs
- needs_64bit:
  - ucsc_tools
  - bfast
  - novoalign
  - novosort
  - plink_seq

- bio_proteomics:
  - -transproteomic_pipeline
  - omssa
  - mzmine
  - myrimatch
  - directag
  - tagrecon
  - idpqonvert
  - pepitome
  - percolator
- bio_proteomics_wine
  proteomics_wine_env
  - *#- multiplierz*
  - proteowizard
- cloudman:
  - nginx
  - proftpd
  - sge
  - novnc
- galaxy:
  - galaxy_webapp
  - galaxy_tools
- galaxyp:
  - protkgem
  - protvis
  - proteomics_tools
- distributed:
  - gnu_parallel
  - pydoop
  - seal
- system:
  - s3fs

- python:
  - bx-python
  - netsa-python
  - rpy
- java:
  - leiningen
- phylogeny:
  - tracer
  - beast
- *#Viral Cloud Resource (VCR) - JCVI's Viral Genomic Pipelines on the cloud*
- vcr:
  - viralassembly
  - viralassembly_cleanall
  - viralvigor

Create an Instance

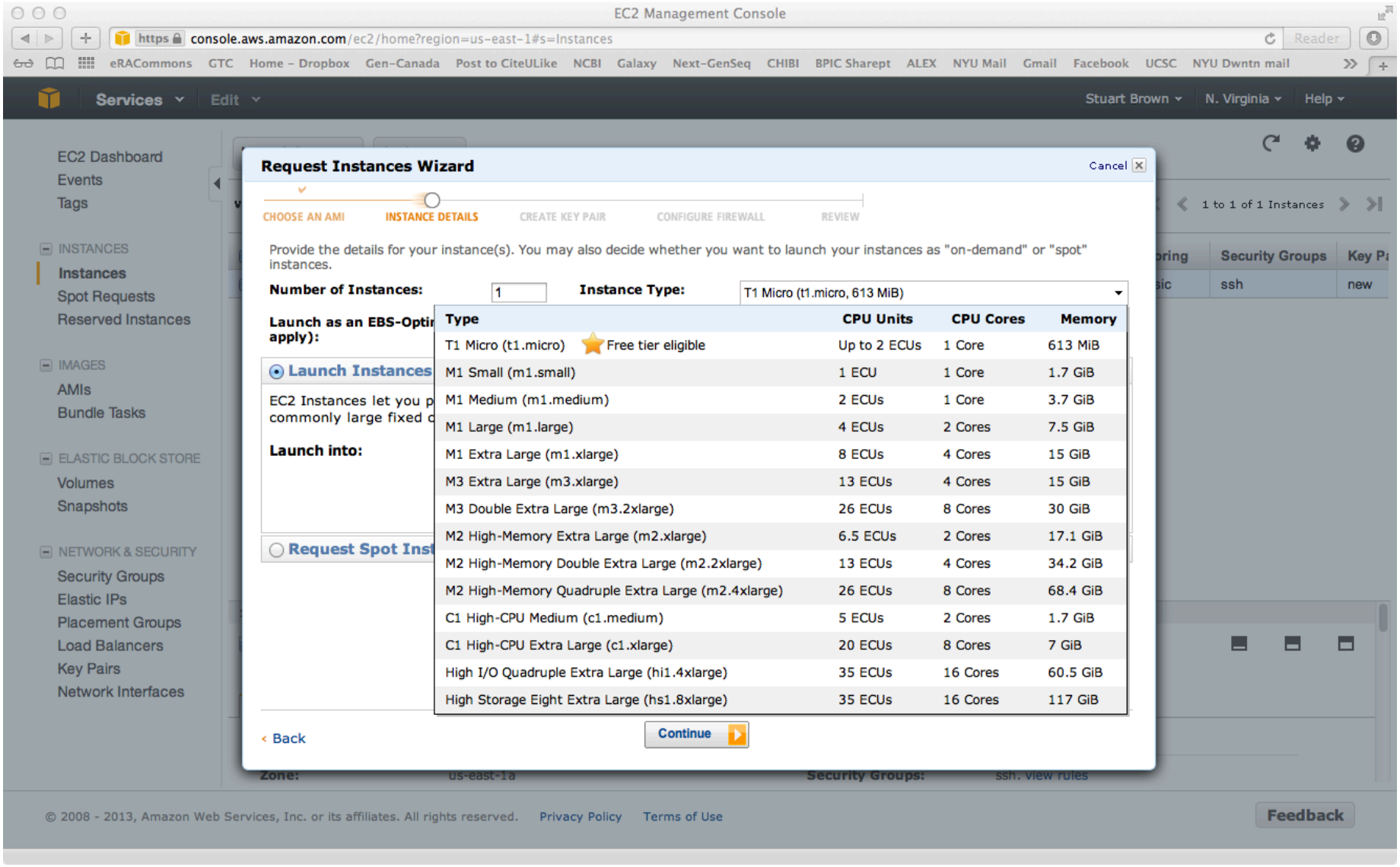


NYU Langone MEDICAL CENTER

Computer Configurations

- The speed of computing tasks are generally limited by 3 things:
  - CPU speed
  - RAM
  - Speed to read stored data (I/O speed)
- All Bioinformatics work does not have one ideal computer configuration
  - different tasks, software, even changes in parameters/options within one program can make different demands from computer hardware
- Amazon EC2 setup allows for many different options in CPU speed, RAM, and data access.
- The same “Image” can be launched on different types of virtual machines
- You may need to experiment to find an optimal setup for your workflow
- We have had success with the “m3.xlarge” setup to process RNA-seq data from HiSeq-2000 runs

EC2 computer sizes

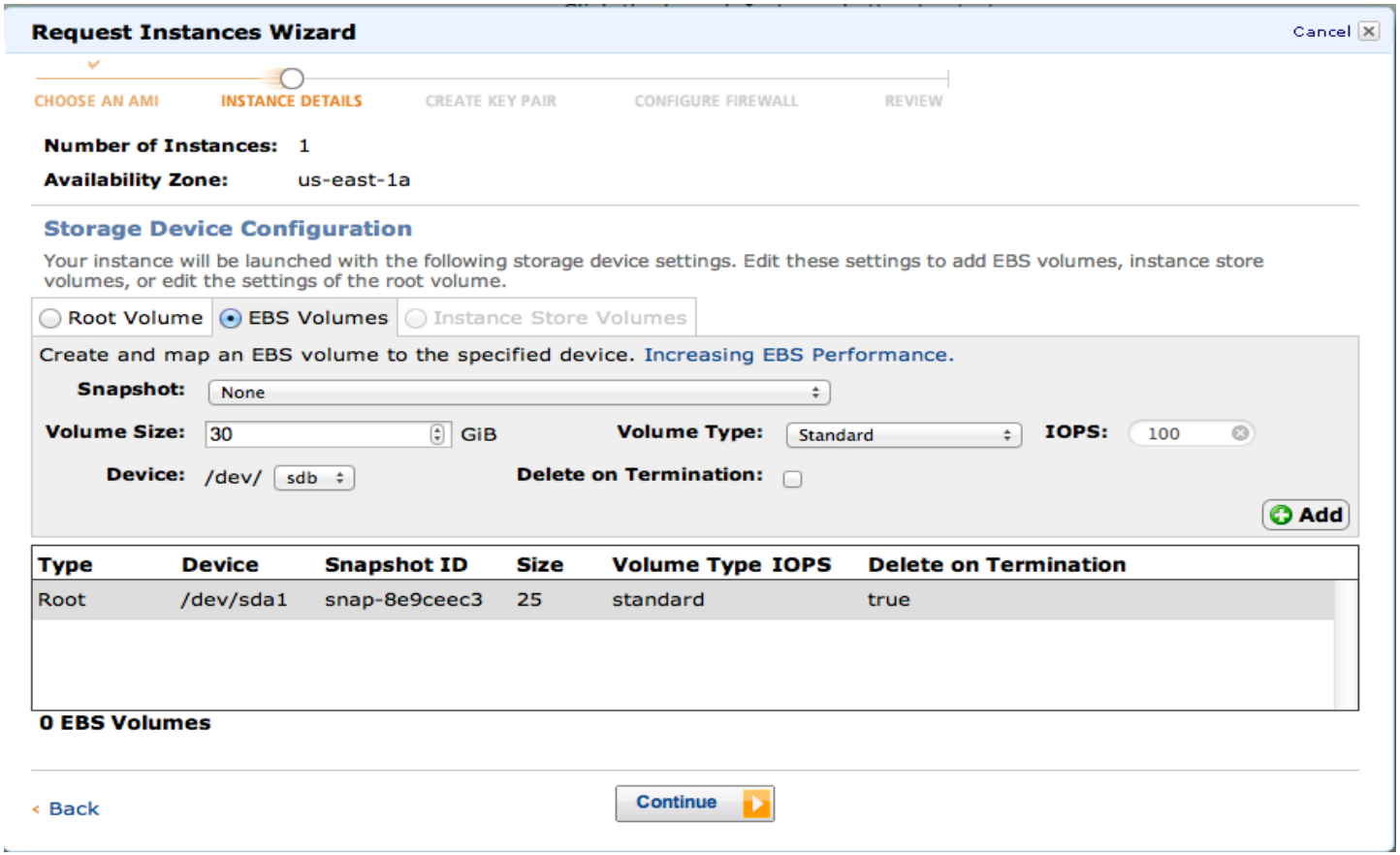


Type	CPU Units	CPU Cores	Memory
T1 Micro (t1.micro) Free tier eligible	Up to 2 ECUs	1 Core	613 MiB
M1 Small (m1.small)	1 ECU	1 Core	1.7 GiB
M1 Medium (m1.medium)	2 ECUs	1 Core	3.7 GiB
M1 Large (m1.large)	4 ECUs	2 Cores	7.5 GiB
M1 Extra Large (m1.xlarge)	8 ECUs	4 Cores	15 GiB
M3 Extra Large (m3.xlarge)	13 ECUs	4 Cores	15 GiB
M3 Double Extra Large (m3.2xlarge)	26 ECUs	8 Cores	30 GiB
M2 High-Memory Extra Large (m2.xlarge)	6.5 ECUs	2 Cores	17.1 GiB
M2 High-Memory Double Extra Large (m2.2xlarge)	13 ECUs	4 Cores	34.2 GiB
M2 High-Memory Quadruple Extra Large (m2.4xlarge)	26 ECUs	8 Cores	68.4 GiB
C1 High-CPU Medium (c1.medium)	5 ECUs	2 Cores	1.7 GiB
C1 High-CPU Extra Large (c1.xlarge)	20 ECUs	8 Cores	7 GiB
High I/O Quadruple Extra Large (hi1.4xlarge)	35 ECUs	16 Cores	60.5 GiB
High Storage Eight Extra Large (hs1.8xlarge)	35 ECUs	16 Cores	117 GiB

Costs

- Your mileage may vary. We spent about $5000 per month to support our sequencing lab.
- Surprisingly, most of the cost was for compute usage, and not for data storage or transfer.

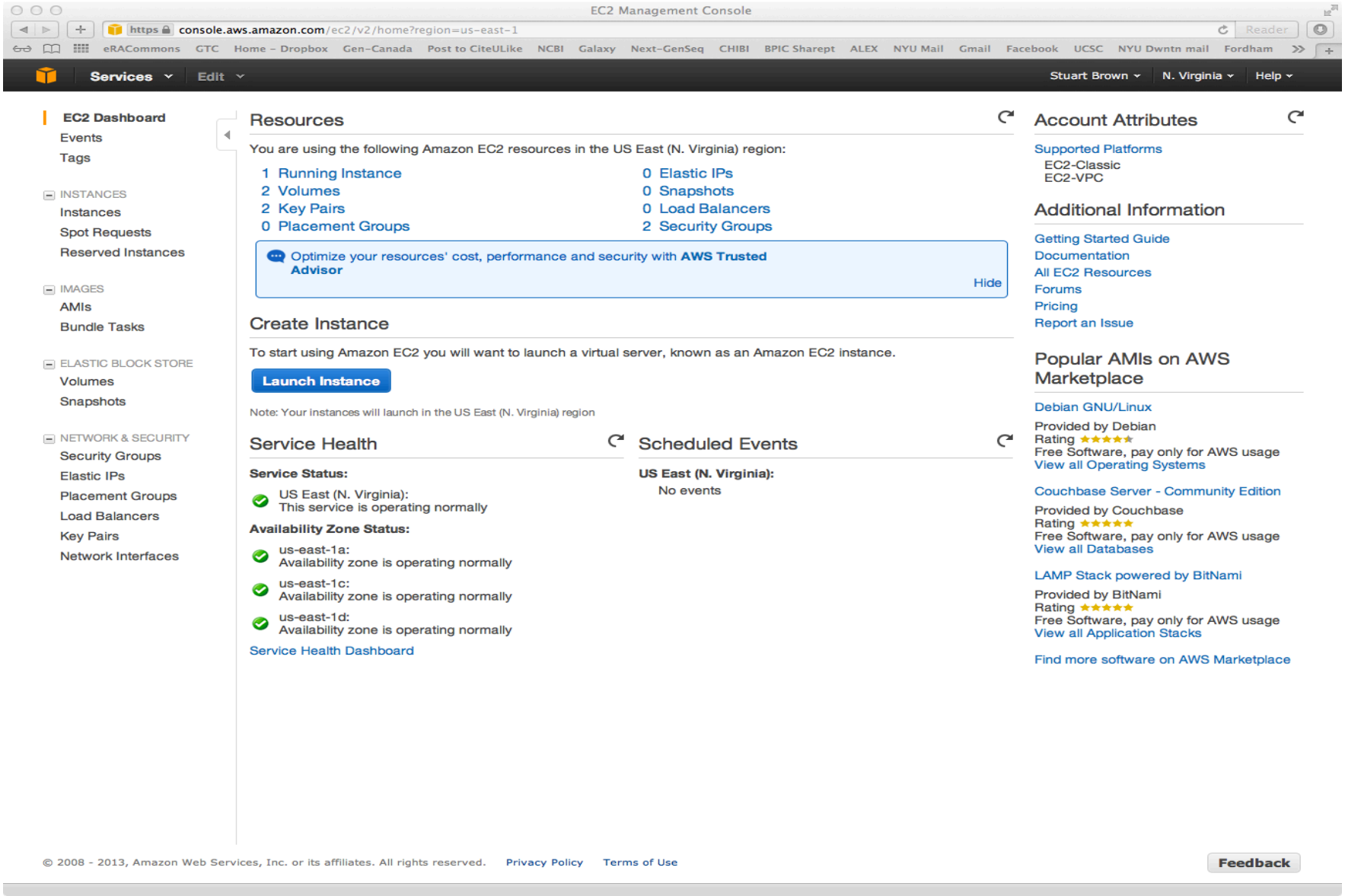
Add EBS Data Volume



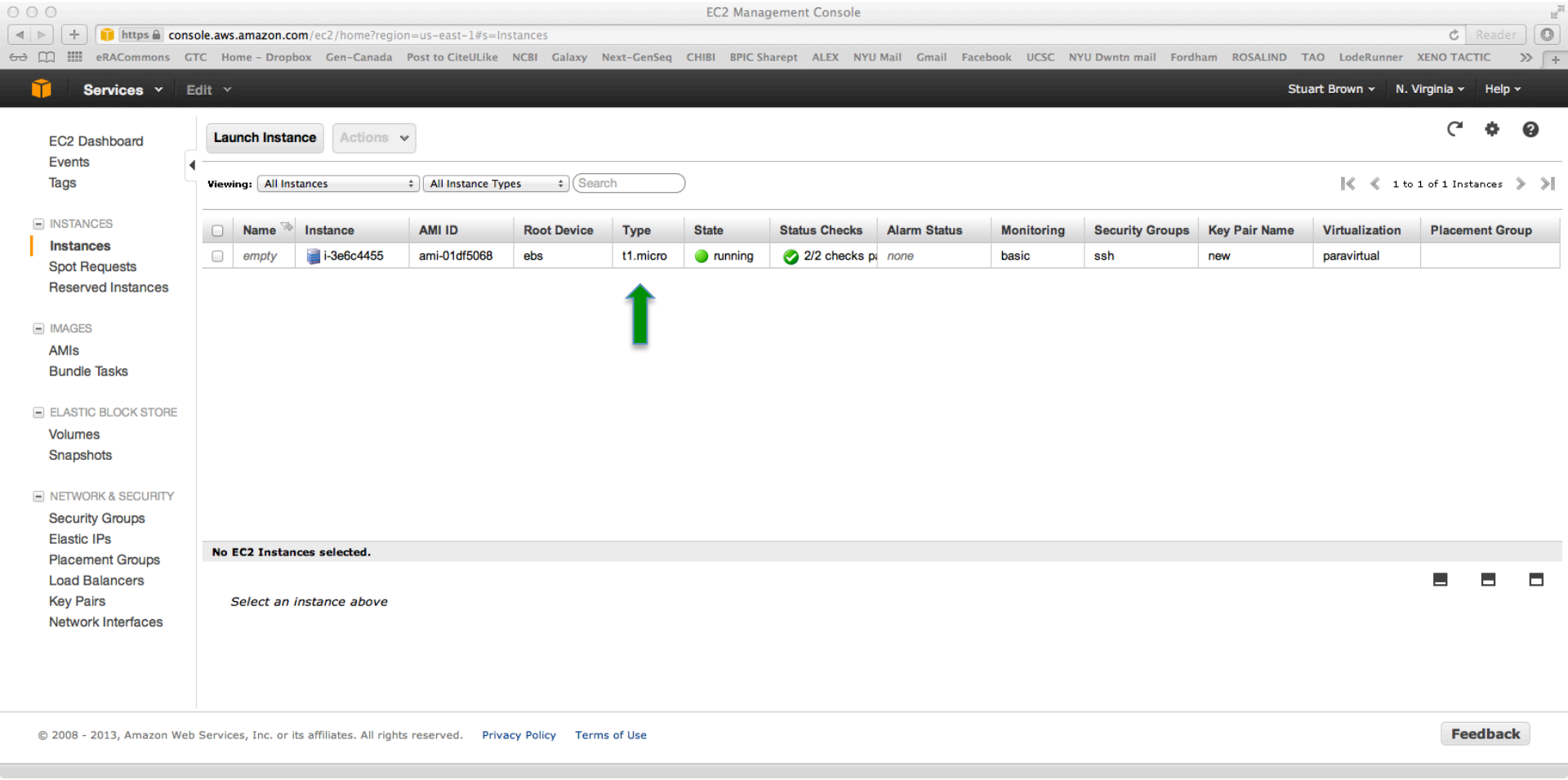
Permanent data storage attached as ‘volume’. Store your data, reference genomes, results. Faster, but more expensive than S3 storage. Use for smaller more frequently used data.

NYU Langone MEDICAL CENTER

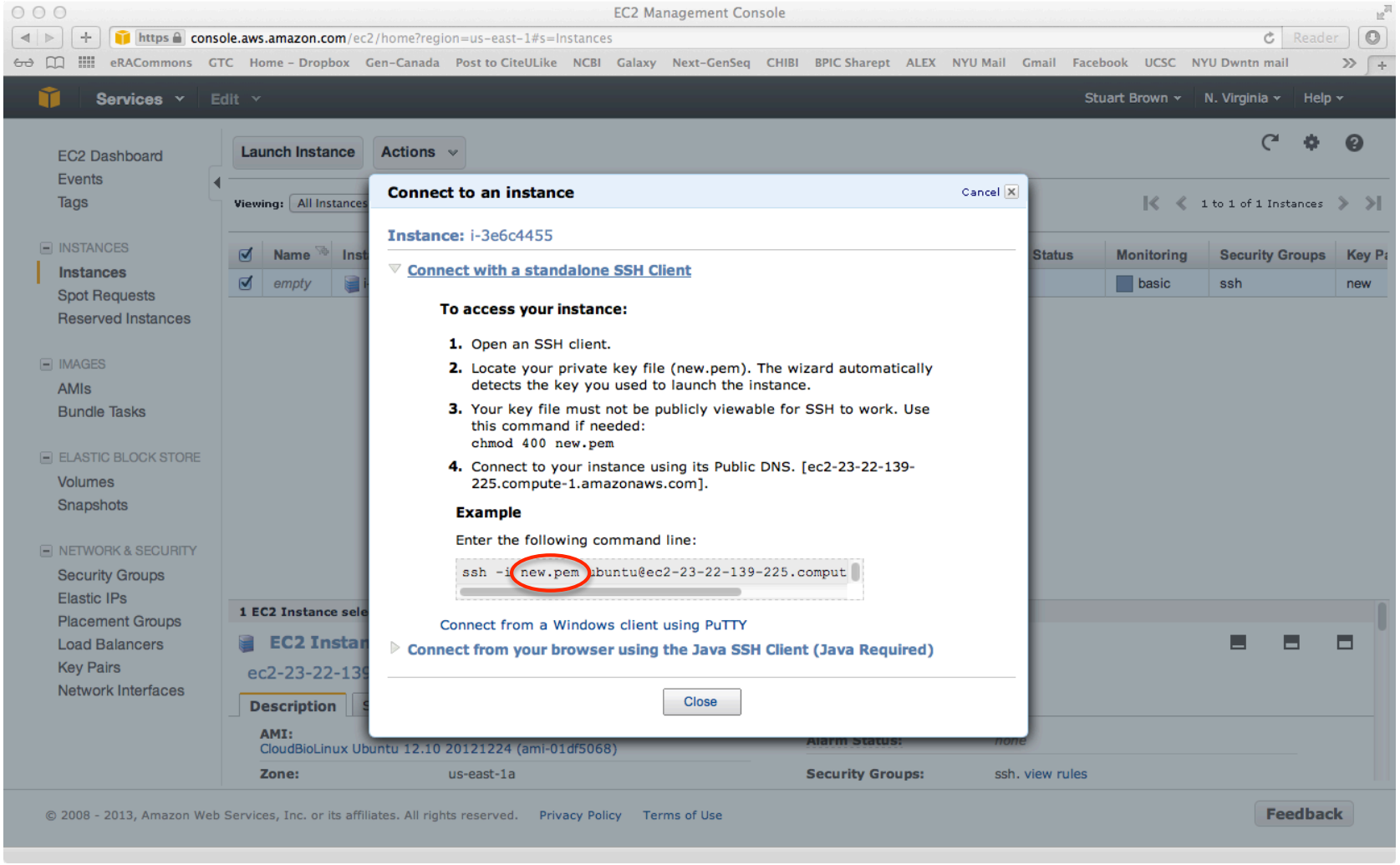
Dashboard


EC2 Management Console
https console.aws.amazon.com/ec2/v2/home?region=us-east-1
Reader
eRACommons GTC Home – Dropbox Gen-Canada Post to CiteULike NCBI Galaxy Next-GenSeq CHIBI BPIC Sharepit ALEX NYU Mail Gmail Facebook UCSC NYU Dwntn mail Fordham
Services Edit
Stuart Brown N. Virginia Help
EC2 Dashboard
Events
Tags
INSTANCES
Instances
Spot Requests
Reserved Instances
IMAGES
AMIs
Bundle Tasks
ELASTIC BLOCK STORE
Volumes
Snapshots
NETWORK & SECURITY
Security Groups
Elastic IPs
Placement Groups
Load Balancers
Key Pairs
Network Interfaces
Resources
You are using the following Amazon EC2 resources in the US East (N. Virginia) region:
1 Running Instance
2 Volumes
2 Key Pairs
0 Placement Groups
0 Elastic IPs
0 Snapshots
0 Load Balancers
2 Security Groups
Optimize your resources' cost, performance and security with AWS Trusted Advisor
Hide
Create Instance
To start using Amazon EC2 you will want to launch a virtual server, known as an Amazon EC2 instance.
Launch Instance
Note: Your instances will launch in the US East (N. Virginia) region
Service Health
Service Status:
US East (N. Virginia):
This service is operating normally
Availability Zone Status:
us-east-1a:
Availability zone is operating normally
us-east-1c:
Availability zone is operating normally
us-east-1d:
Availability zone is operating normally
Service Health Dashboard
Scheduled Events
US East (N. Virginia):
No events
Account Attributes
Supported Platforms
EC2-Classic
EC2-VPC
Additional Information
Getting Started Guide
Documentation
All EC2 Resources
Forums
Pricing
Report an Issue
Popular AMIs on AWS Marketplace
Debian GNU/Linux
Provided by Debian
Rating
Free Software, pay only for AWS usage
View all Operating Systems
Couchbase Server - Community Edition
Provided by Couchbase
Rating
Free Software, pay only for AWS usage
View all Databases
LAMP Stack powered by BitNami
Provided by BitNami
Rating
Free Software, pay only for AWS usage
View all Application Stacks
Find more software on AWS Marketplace
© 2008 - 2013, Amazon Web Services, Inc. or its affiliates. All rights reserved.
Privacy Policy
Terms of Use
Feedback

My Instance (micro)

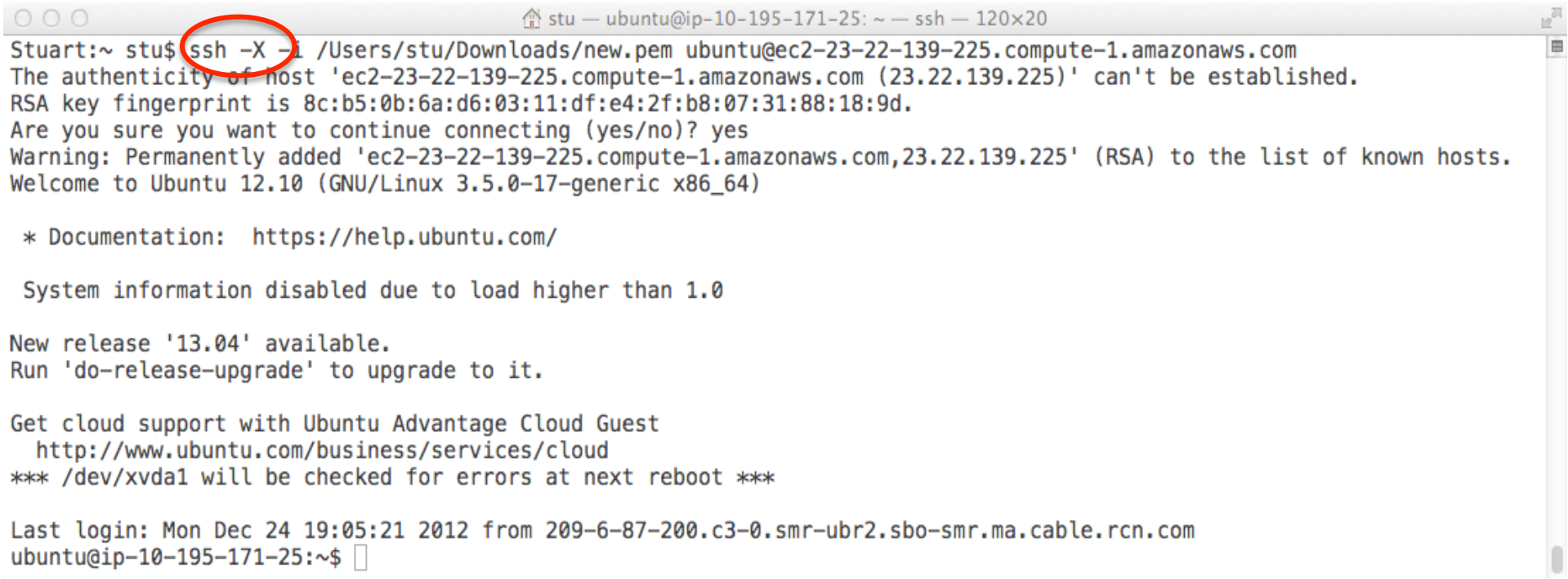


Actions > Connect > SSH Client

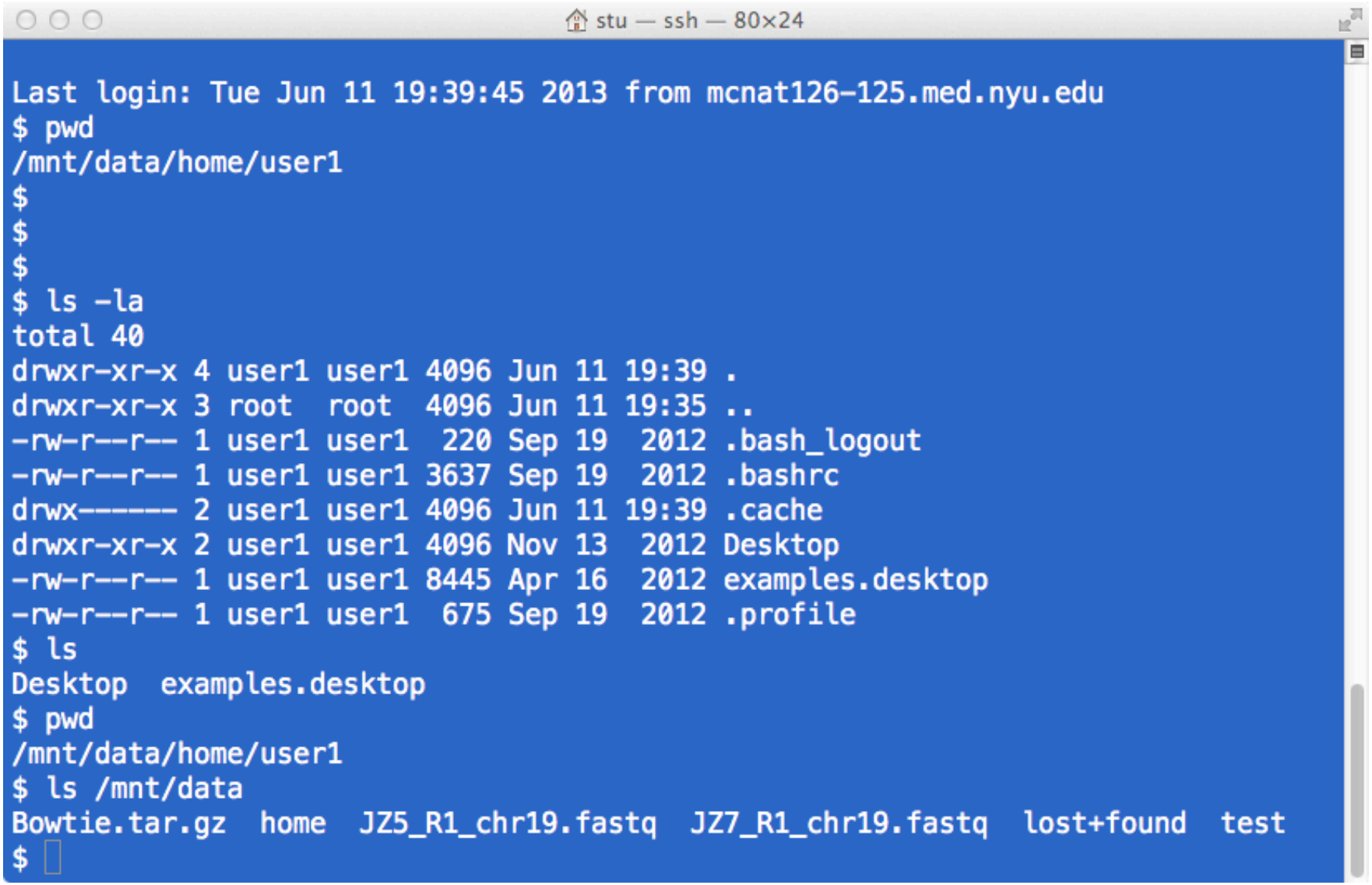


ssh login with Terminal (Mac) or PuTTY (Windows)

ssh -X option for X11 graphics on Mac



Look at files



FASTQ files

- $ cd /mnt/data
- $ more JZ5_R1_chr19.fastq

FASTQC for initial QC

- $ fastqc /mnt/data/JZ5_R1_chr19.fastq -o .
- $ unzip JZ5_R1_chr19_fastqc.zip
- $ more JZ5_R1_chr19_fastqc/summary.txt

This dot is important, it puts the results back in your home directory

```
$ more summary.txt
PASS    Basic Statistics        JZ5_R1_chr19.fastq
PASS    Per base sequence quality       JZ5_R1_chr19.fastq
PASS    Per sequence quality scores     JZ5_R1_chr19.fastq
FAIL    Per base sequence content       JZ5_R1_chr19.fastq
FAIL    Per base GC content     JZ5_R1_chr19.fastq
WARN    Per sequence GC content JZ5_R1_chr19.fastq
PASS    Per base N content      JZ5_R1_chr19.fastq
PASS    Sequence Length Distribution    JZ5_R1_chr19.fastq
FAIL    Sequence Duplication Levels     JZ5_R1_chr19.fastq
PASS    Overrepresented sequences       JZ5_R1_chr19.fastq
WARN    Kmer Content    JZ5_R1_chr19.fastq
```

Align RNA-seq with TopHat/Bowtie

- To use TopHat, you will need the following programs in your PATH:
  - bowtie2 and bowtie2-align (or bowtie)
  - bowtie2-inspect (or bowtie-inspect)
  - bowtie2-build (or bowtie-build)
  - Samtools
  - You will also need Python **version 2.6 or higher**
  - Configure the package, specifying the install path and the library dependencies as needed:

    ./configure --prefix=<install_prefix> --with-boost=<boost_install_prefix> --with-bam=<samtools_install_prefix>

- Setting up a server to run Tophat is not trivial. The CloudBioLinux people **took care of all of this** in their EC2 Image.

TopHat uses Bowtie

- Tophat uses Bowtie to align reads to the genome
- Bowtie requires a pre-computed index of the genome (a k-mer hash)
- Minimal TopHat command syntax:

$ tophat2 bowtie_index data1.fastq

TopHat has lots of options and parameters

http://tophat.cbcb.umd.edu/manual.shtml

```
tophat:
TopHat maps short sequences from spliced transcripts to whole genomes.
Usage:
   tophat [options] <bowtie_index> <reads1[,reads2,...]> [reads1[,reads2,...]] \
                    [quals1,[quals2,...]] [quals1[,quals2,...]]
Options:
   -v/--version
   -o/--output-dir             <string>    [ default: ./tophat_out       ]
   --bowtie1                               [ default: bowtie2            ]
   -N/--read-mismatches        <int>       [ default: 2                  ]
   --read-gap-length           <int>       [ default: 2                  ]
   --read-edit-dist            <int>       [ default: 2                  ]
   --read-realign-edit-dist    <int>       [ default: "read-edit-dist" + 1 ]
   -a/--min-anchor             <int>       [ default: 8                  ]
   -m/--splice-mismatches      <0-2>       [ default: 0                  ]
   -i/--min-intron-length      <int>       [ default: 50                 ]
   -I/--max-intron-length      <int>       [ default: 500000             ]
   -g/--max-multihits          <int>       [ default: 20                 ]
   --suppress-hits
   -x/--transcriptome-max-hits <int>       [ default: 60                 ]
   -M/--prefilter-multihits                ( for -G/--GTF option, enable
                                           an initial bowtie search
                                           against the genome )
   --max-insertion-length      <int>       [ default: 3                  ]
   --max-deletion-length       <int>       [ default: 3                  ]
   --solexa-quals
   --solexa1.3-quals                       (same as phred64-quals)
   --phred64-quals                         (same as solexa1.3-quals)
   -Q/--quals
   --integer-quals
   -C/--color                              (Solid - color space)
   --color-out
   --library-type              <string>    (fr-unstranded, fr-firststrand,
                                           fr-secondstrand)
   -p/--num-threads            <int>       [ default: 1                  ]
   -R/--resume                 <out_dir>   ( try to resume execution )
   -G/--GTF                    <filename>  (GTF/GFF with known transcripts)
   --transcriptome-index       <bwtidx>    (transcriptome bowtie index)
   -T/--transcriptome-only                 (map only to the transcriptome)
   -j/--raw-juncs              <filename>
   --insertions                <filename>
   --deletions                 <filename>
```

```
Advanced Options:
   --report-secondary-alignments
   --no-discordant
   --no-mixed

   --segment-mismatches        <int>       [ default: 2                  ]
   --segment-length            <int>       [ default: 25                 ]

   --bowtie-n                              [ default: bowtie -v          ]
   --min-coverage-intron       <int>       [ default: 50                 ]
   --max-coverage-intron       <int>       [ default: 20000              ]
   --min-segment-intron        <int>       [ default: 50                 ]
   --max-segment-intron        <int>       [ default: 500000             ]
   --no-sort-bam                           (Output BAM is not coordinate-sorted)
   --no-convert-bam                        (Do not output bam format.
                                           Output is <output_dir>/accepted_hit.sam)
   --keep-fasta-order
   --allow-partial-mapping

Bowtie2 related options:
  Preset options in --end-to-end --b2-very-fast
   --b2-fast
   --b2-sensitive
   --b2-very-sensitive

  Alignment options
   --b2-N                      <int>       [ default: 0                  ]
   --b2-L                      <int>       [ default: 20                 ]
   --b2-i                      <func>      [ default: S,1,1.25           ]
   --b2-n-ceil                 <func>      [ default: L,0,0.15           ]
   --b2-gbar                   <int>       [ default: 4                  ]

  Scoring options
   --b2-mp                     <int>,<int> [ default: 6,2                ]
   --b2-np                     <int>       [ default: 1                  ]
   --b2-rdg                    <int>,<int> [ default: 5,3                ]
   --b2-rfg                    <int>,<int> [ default: 5,3                ]
   --b2-score-min              <func>      [ default: L,-0.6,-0.6        ]
```

```
Fusion related options:
   --fusion-search
   --fusion-anchor-length      <int>       [
   --fusion-min-dist           <int>       [ de
   --fusion-read-mismatches    <int>
   --fusion-multireads         <int>       [ d
   --fusion-multipairs         <int>       [ de
   --fusion-ignore-chromosomes <list>

   --fusion-do-not-resolve-conflicts

SAM Header Options (for embedding se
   --rg-id                     <string>    (read gr
   --rg-sample                 <string>    (sam
   --rg-library                <string>    (librar
   --rg-description            <string>    (de
   --rg-platform-unit          <string>    (e
   --rg-center                 <string>    (sequ
   --rg-date                   <string>    (ISO 8
   --rg-platform               <string>    (Seq
```

Tophat uses gene annotation

- Tophat maps RNA reads to genes across exons. You need a genome annotation file (GTF) with the locations of known genes and exons.
- This works faster if you pre-compute a *transcriptome index* file
- It saves compute time to skip the coverage index

```
$ tophat2 --no-coverage-search  \
--transcriptome-index=tx_index bowtie_index data1.fastq
```

GFF/GTF format

- http://cufflinks.cbcb.umd.edu/gff.html
- http://www.sequenceontology.org/gff3.shtml

```
##gff-version 3##sequence-region   ctg123 1 1497228
ctg123 . gene            1000  9000  .  +  .  ID=gene00001;Name=EDEN
ctg123 . TF_binding_site 1000  1012  .  +  .  Parent=gene00001
ctg123 . mRNA            1050  9000  .  +  .  ID=mRNA00001;Parent=gene00001
ctg123 . mRNA            1050  9000  .  +  .  ID=mRNA00002;Parent=gene00001
ctg123 . mRNA            1300  9000  .  +  .  ID=mRNA00003;Parent=gene00001
ctg123 . exon            1300  1500  .  +  .  Parent=mRNA00003
ctg123 . exon            1050  1500  .  +  .  Parent=mRNA00001,mRNA00002
ctg123 . exon            3000  3902  .  +  .  Parent=mRNA00001,mRNA00003
ctg123 . exon            5000  5500  .  +  .  Parent=mRNA00001,mRNA00002,mRNA00003
ctg123 . exon            7000  9000  .  +  .  Parent=mRNA00001,mRNA00002,mRNA00003
```

Run tophat2

```
ubuntu@ip-10-145-184-13:/mnt/data$ tophat2 --no-coverage-search
    --transcriptome-index=Transcriptomeindex/genes Bowtie2index/chr19 JZ5_R1_chr19.fastq, JZ7_R1_chr19.fastq
[2013-06-17 15:13:23] Beginning TopHat run (v2.0.6)
-----------------------------------------------
[2013-06-17 15:13:23] Checking for Bowtie:  Bowtie version:      2.0.4.0
[2013-06-17 15:13:23] Checking for Samtools:   Samtools version:       0.1.18.0
[2013-06-17 15:13:23] Checking for Bowtie index files
[2013-06-17 15:13:23] Checking for Bowtie index files
[2013-06-17 15:13:23] Checking for reference FASTA file
[2013-06-17 15:13:23] Generating SAM header for Bowtie2index/chr19
        format:                 fastq
        quality scale:          phred33 (default)
[2013-06-17 15:13:23] Reading known junctions from GTF file
[2013-06-17 15:13:29] Preparing reads
         left reads: min. length=50, max. length=50, 1790579 kept reads (1 discarded)
[2013-06-17 15:14:01] Using pre-built transcriptome index..
[2013-06-17 15:14:09] Mapping left_kept_reads to transcriptome genes with Bowtie2
[2013-06-17 15:20:53] Resuming TopHat pipeline with unmapped reads
[2013-06-17 15:20:53] Mapping left_kept_reads.m2g_um to genome chr19 with Bowtie2
[2013-06-17 15:23:05] Searching for junctions via segment mapping
[2013-06-17 15:23:27] Retrieving sequences for splices
[2013-06-17 15:23:31] Indexing splices
[2013-06-17 15:23:34] Mapping left_kept_reads.m2g_um_seg1 to genome segment_juncs with Bowtie2
[2013-06-17 15:23:42] Joining segment hits
[2013-06-17 15:23:48] Reporting output tracks
-----------------------------------------------
[2013-06-17 15:25:47] Run complete: 00:12:23 elapsed
```

Its faster to process several fastq files at once

TopHat output is a BAM file

- You can view the BAM file with IGV (or any other genomic data browser)
- Generally your goal from RNA-seq is to get gene expression values
- Cufflinks calculated gene expression values from the TopHat output
- Cufflinks can find novel splice junctions, or even completely new transcripts from unannotated regions of the genome

http://cufflinks.cbcb.umd.edu/tutorial.html

Differential expression

- Your **real** goal in an RNA-seq experiment is usually to find differentially expressed (DE) genes.
- Cuffdiff calculates differential expression from the Cufflinks output of different samples.
- You can also use many other tools (Deseq, edgeR, etc.) on these same expression values.
- Statistical significance of DE depends on an adequate number of replicates to control for multiple testing and false discovery – see your friendly neighborhood biostatistician for a sample size estimate

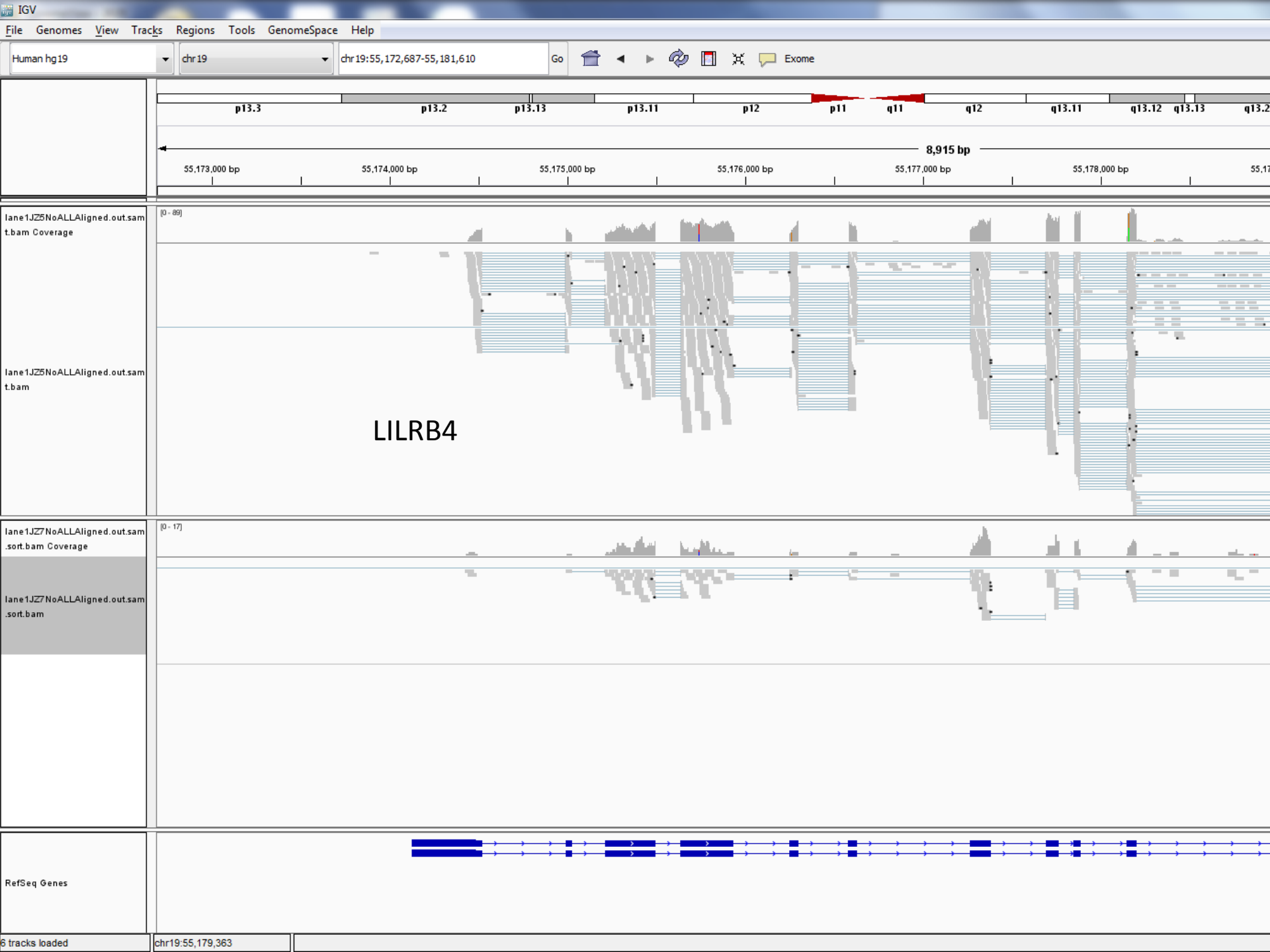
Run cuffdiff

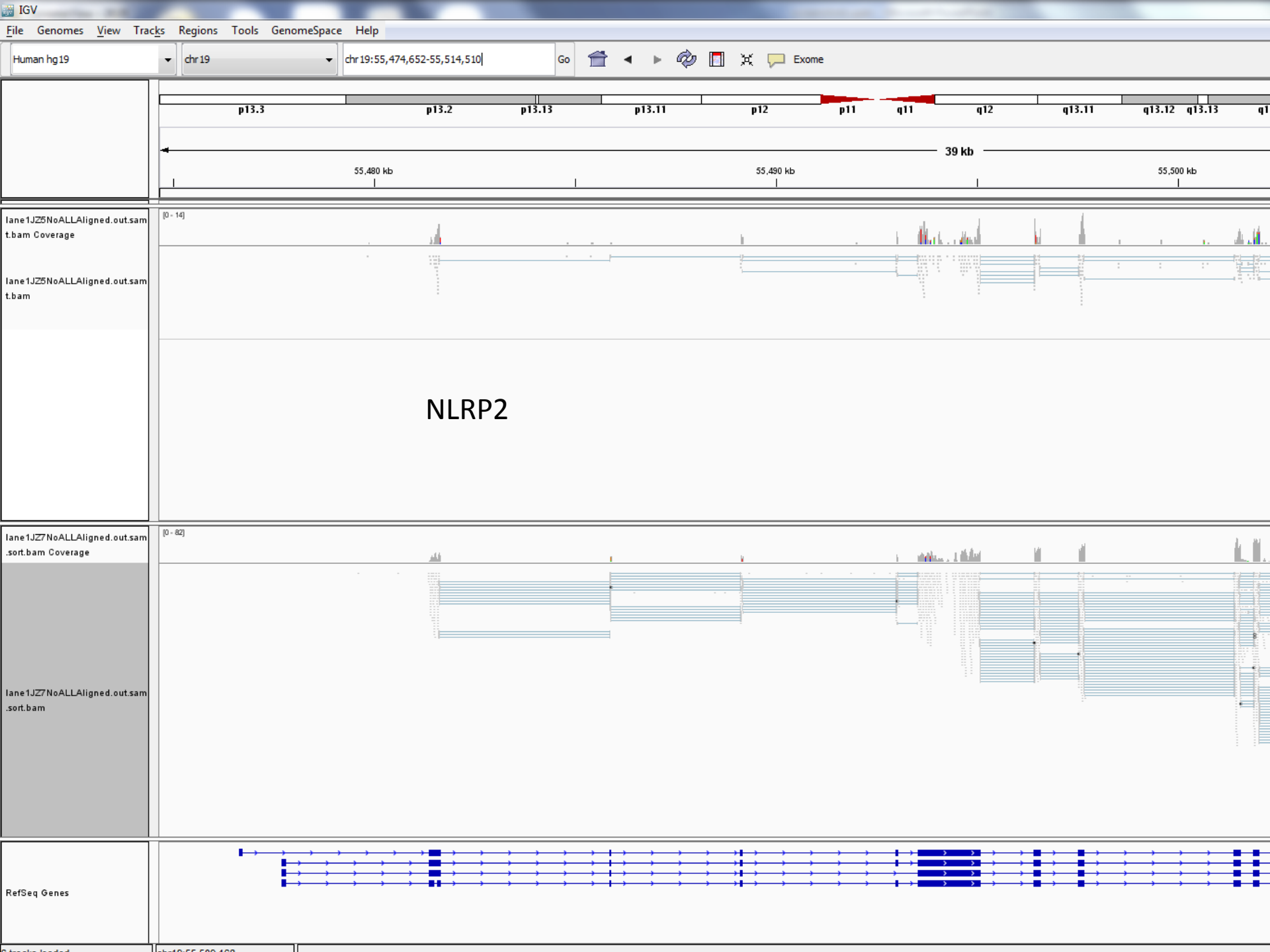
cuffdiff Transcriptomeindex/genes.gff JZ5-hits.bam JZ7-hits.bam

```
[17:42:00] Loading reference annotation.
[17:42:05] Inspecting maps and determining fragment length distributions.
[17:42:34] Modeling fragment count overdispersion.
> Map Properties:
>       Normalized Map Mass: 2314081.06
>       Raw Map Mass: 1560963.69
>       Fragment Length Distribution: Truncated Gaussian (default)
>                 Default Mean: 200
>               Default Std Dev: 80
> Map Properties:
>       Normalized Map Mass: 2314081.06
>       Raw Map Mass: 2725420.43
>       Fragment Length Distribution: Truncated Gaussian (default)
>                 Default Mean: 200
>               Default Std Dev: 80
[17:42:36] Calculating preliminary abundance estimates
[17:42:36] Testing for differential expression and regulation in locus.
> Processing Locus chr19:19649073-19657468     [                    ]   2%
```

View DE report

CD70	CD70	CD70	chr19:6585849-6591163	q1	q2	OK	40.7938	0.441674	-6.52922	3.2505	0.00115204	0.0430794	yes
CD79A	CD79A	CD79A	chr19:42381189-42385439	q1	q2	OK	6881.15	3.6651	-10.8746	5.82704	5.64181e-09	3.48099e-06	yes
CLEC17A	CLEC17A	CLEC17A	chr19:14693895-14721956	q1	q2	OK	120.887	0.338903	-8.47857	3.96327	7.39299e-05	0.00608197	yes
CNTD2	CNTD2	CNTD2	chr19:40728114-40732597	q1	q2	OK	4.39132	97.8472	4.4778	-3.18834	0.00143090.0476079		yes
EBI3	EBI3	EBI3	chr19:4229539-4237524	q1	q2	OK	245.469	11.076	-4.47003	3.27053	0.00107340.0413959		yes
FAM129C	FAM129C	FAM129C	chr19:17634109-17664648	q1	q2	OK	237.18	0.543665	-8.76905	5.26643	1.39099e-07	4.2912e-05	yes
FCER2	FCER2	FCER2	chr19:7753642-7767032	q1	q2	OK	647.367	0.443233	-10.5123	4.64155	3.4581e-06	0.000609613	yes
FCGBP	FCGBP	FCGBP	chr19:40353962-40440533	q1	q2	OK	51.894	735.228	3.82455	-4.09755	4.17539e-05	0.00380521	yes
FLJ22184	FLJ22184	FLJ22184	chr19:7933604-7939326	q1	q2	OK	4.37947	93.7349	4.41976	-3.352	0.000802312	0.0330018	yes


IGV
File Genomes View Tracks Regions Tools GenomeSpace Help
Human hg19
chr19
chr19:55,172,687-55,181,610
Go
Exome
p13.3 p13.2 p13.13 p13.11 p12 p11 q11 q12 q13.11 q13.12 q13.13
8,915 bp
55,173,000 bp
55,174,000 bp
55,175,000 bp
55,176,000 bp
55,177,000 bp
55,178,000 bp
lane1JZ5NoALLAligned.out.sam t.bam Coverage
[0 - 89]
lane1JZ5NoALLAligned.out.sam t.bam
LILRB4
lane1JZ7NoALLAligned.out.sam .sort.bam Coverage
[0 - 17]
lane1JZ7NoALLAligned.out.sam .sort.bam
RefSeq Genes
6 tracks loaded
chr19:55,179,363


IGV
File Genomes View Tracks Regions Tools GenomeSpace Help
Human hg19
chr19
chr19:55,474,652-55,514,510
Go
Exome
p13.3 p13.2 p13.13 p13.11 p12 p11 q11 q12 q13.11 q13.12 q13.13
39 kb
55,480 kb
55,490 kb
55,500 kb
lane1JZ5NoALLAligned.out.sam t.bam Coverage
lane1JZ5NoALLAligned.out.sam t.bam
[0 - 14]
NLRP2
lane1JZ7NoALLAligned.out.sam .sort.bam Coverage
lane1JZ7NoALLAligned.out.sam .sort.bam
[0 - 82]
RefSeq Genes