From ENCODE Data to ENCODE Analyses

J. Seth Strattan, PhD

ENCODE Data Coordinating Center (DCC)

Asia Pacific Bioinformatics Conference

January, 2016

ENCODE: Metadata, Data, and Analyses

So far, you have learned

- The ENCODE Portal is the **canonical source** for ENCODE metadata and data.
- The Portal also documents **ENCODE standards** like antibody standards, data release.
- The Portal links to **documentation and tutorials**.
- How to use the Portal to **browse and search** what ENCODE has done.

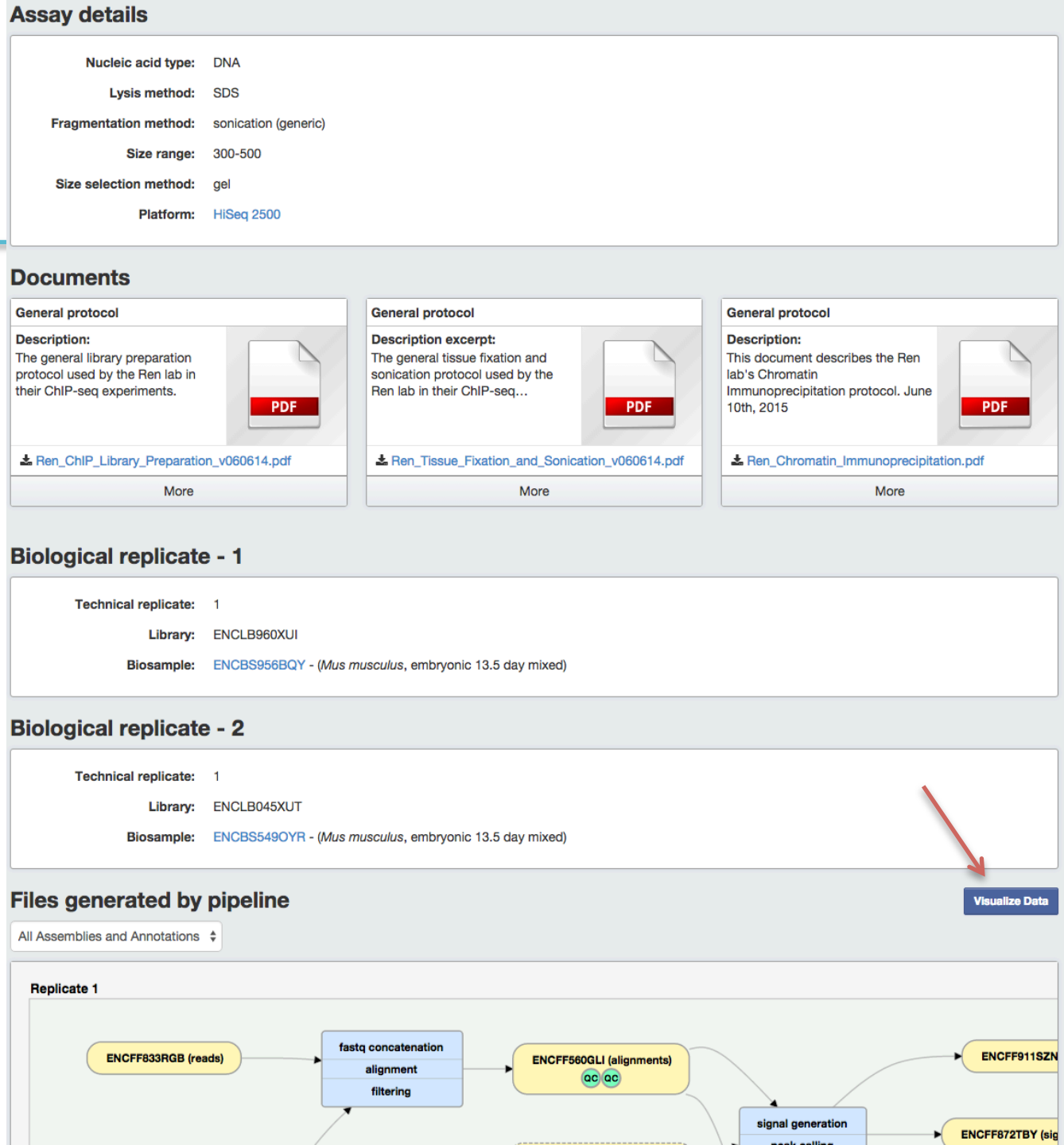
Focus for the rest of the course

- **Visualization** of ENCODE data.
- Programmatic **search and download** of ENCODE metadata and data.
- ENCODE **data analyses**, and how you can replicate them.

Find an experiment

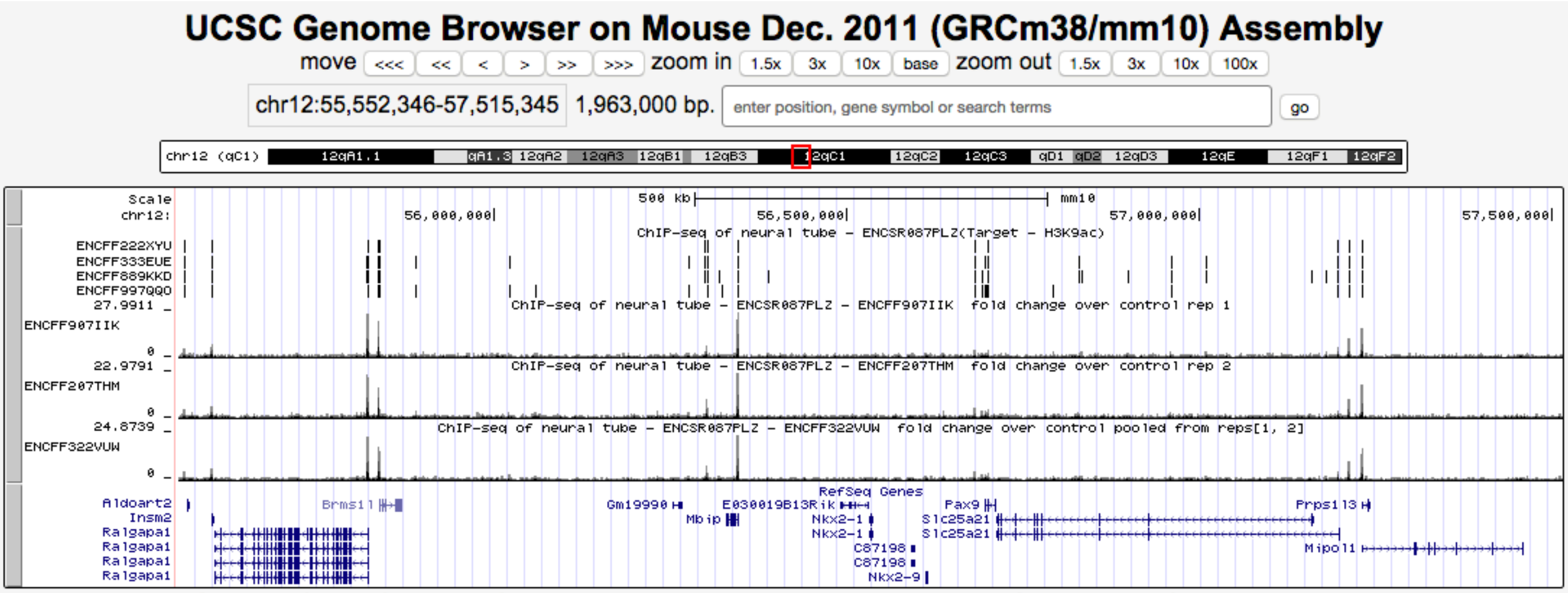
Use metadata to find data:

- Search for “H3K9ac neural tube”
- Facet on ChIP-seq; mouse; mm10 assembly
- Select an experiment, for example **https://www.encodeproject.org/experiments/ENCSR087PLZ/**
- Note metadata on protocols, replicates
- Graph: files are related by processing steps
- Download from the graph or a list
- Click on “Visualize Data” to visualize the results of this experiment.



Visualize the experiment

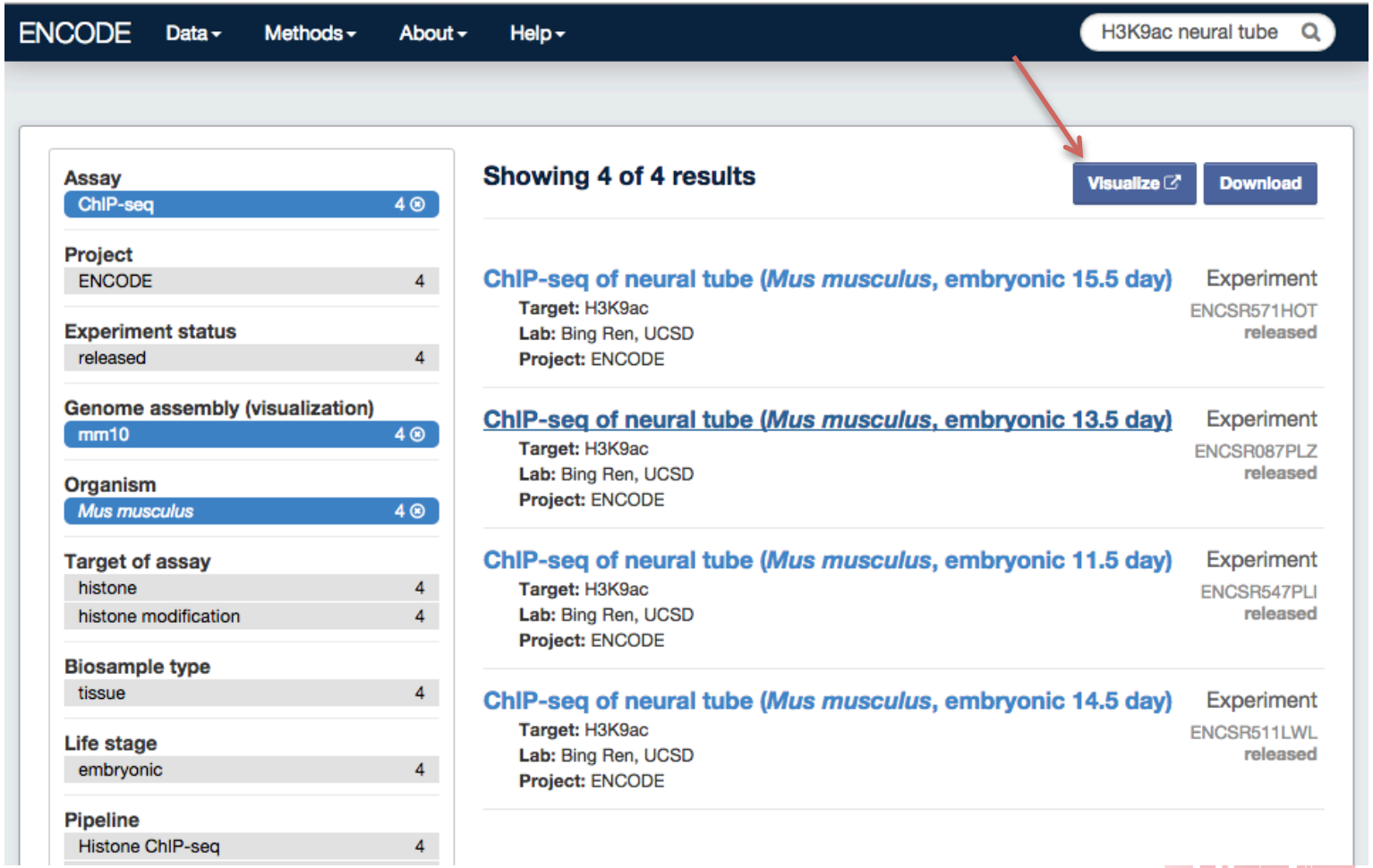
Adjust the browser settings to display fold-over-signal in ”full”



Find several experiments

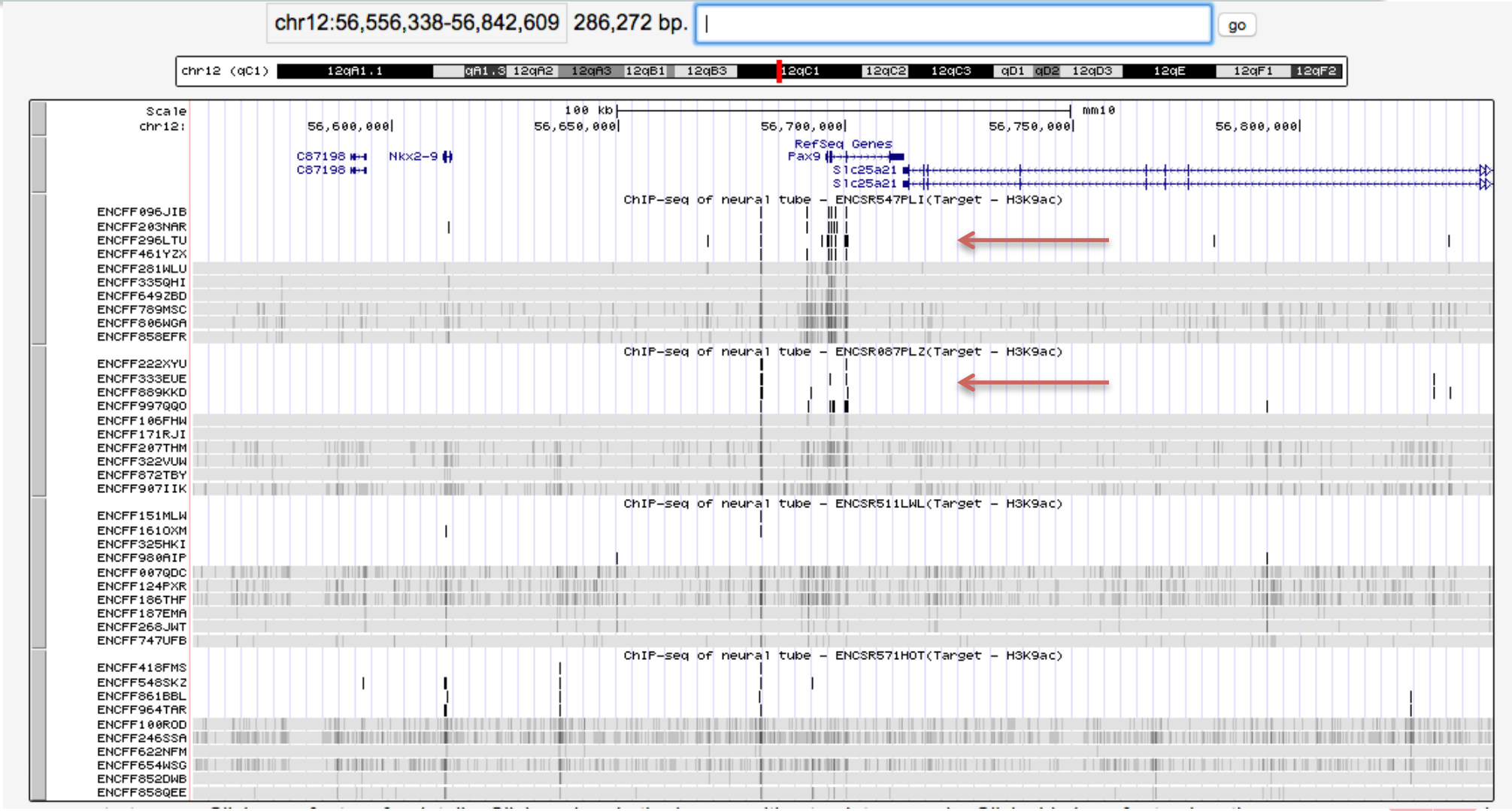
Use metadata to find data:

- Search for “H3K9ac neural tube”
- Facet on ChIP-seq; mouse; mm10 assembly
- Get a list of several experiments
- Click on “Visualize Data” to visualize all the experiments matching this search.



Visualize several experiments

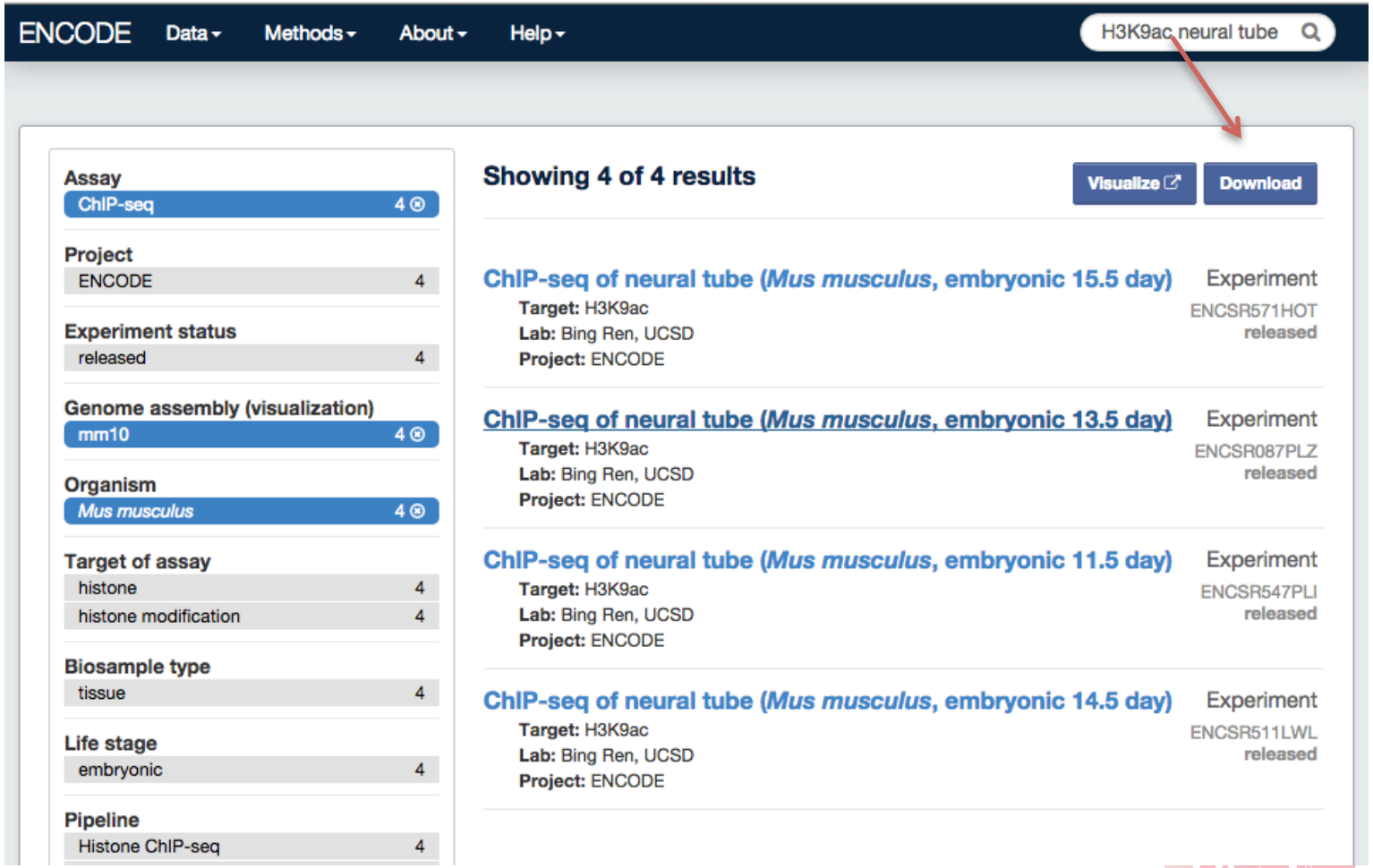
Stage-dependent H3K9ac signal present at Pax9 in neural tube at e11.5, e13.5.



Find & download several experiments

Use metadata to find data:

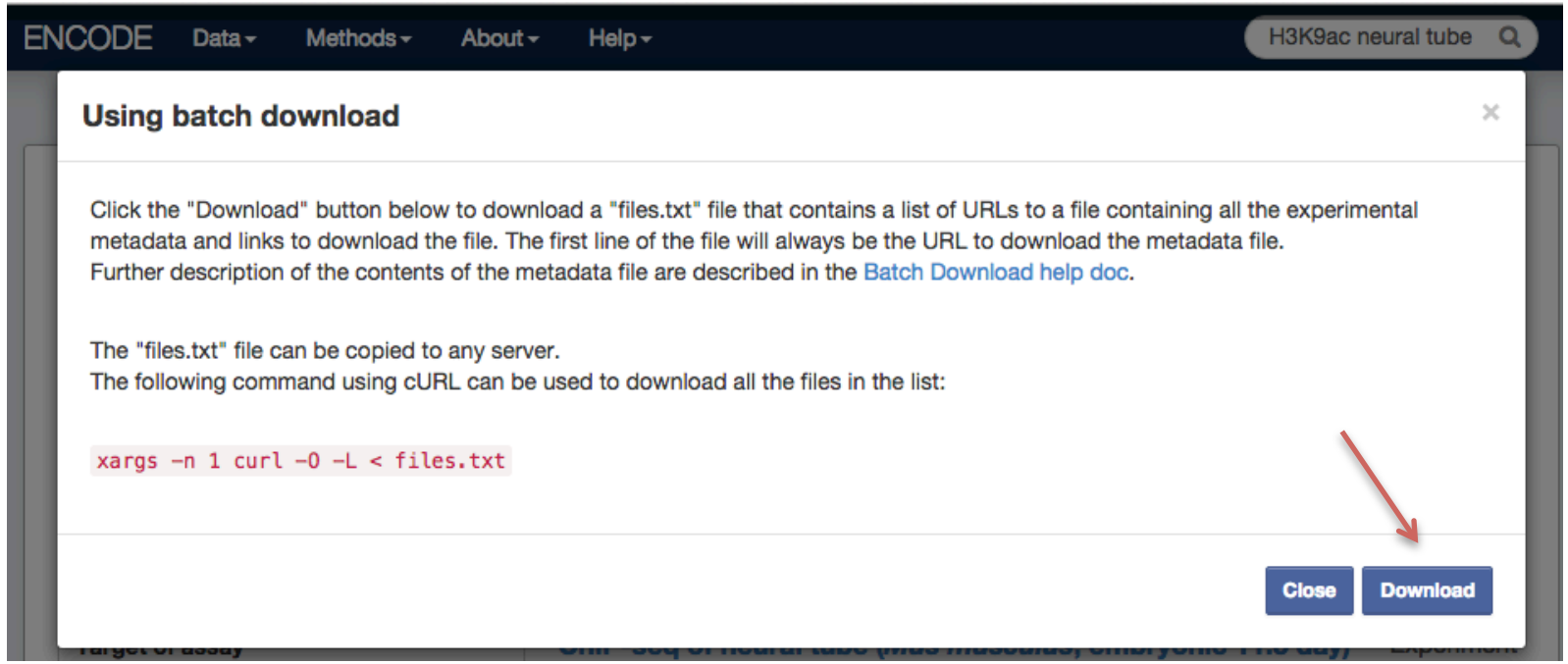
- Search for “H3K9ac neural tube”
- Facet on ChIP-seq; mouse; mm10 assembly
- Get a list of several experiments
- Click on “Download” to download selected metadata and complete links to data.



Download several experiments

Use metadata to find data:

- Search for “H3K9ac neural tube”
- Facet on ChIP-seq; mouse; mm10 assembly
- Get a list of several experiments
- Click on “Download” to download selected metadata and complete links to data.

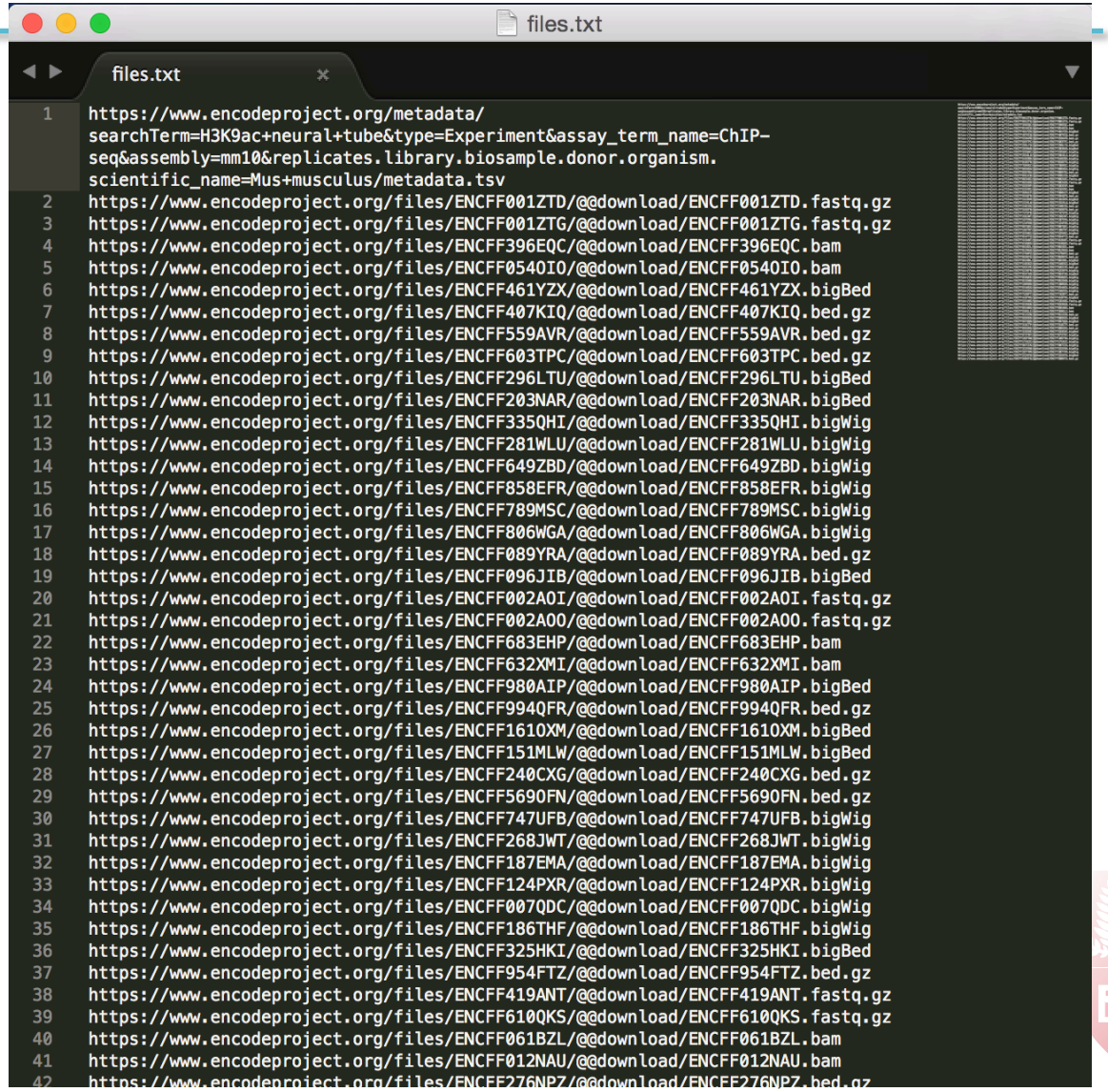


Download several experiments

- “Download” produces a file with a list of links to all the files for all the experiments in your search.
- You can iterate through the list in your own script.
- Or:

```
xargs –n 1 curl –O –L < files.txt
```

- The first link is to a file called metadata.tsv that contains metadata you need to interpret what each file is.



Download several experiments

- metadata.tsv: Each line contains metadata on a file from the download package.

A1 | File accession

	A	B	C	D	E	F	G	H	I	J	K
1	File accession	File format	Output type	Experiment accession	Assay	Biosample term id	Biosample term name	Biosample type	Biosample life stage	Biosample sex	Biosample organism
2	ENCFF001ZTD	fastq	reads	ENCSR547PLI	ChIP-seq	UBERON:0001049	neural tube	tissue	embryonic	mixed	Mus musculus
3	ENCFF001ZTG	fastq	reads	ENCSR547PLI	ChIP-seq	UBERON:0001049	neural tube	tissue	embryonic	mixed	Mus musculus
4	ENCFF396EQC	bam	alignments	ENCSR547PLI	ChIP-seq	UBERON:0001049	neural tube	tissue	embryonic	mixed	Mus musculus
5	ENCFF054OIO	bam	alignments	ENCSR547PLI	ChIP-seq	UBERON:0001049	neural tube	tissue	embryonic	mixed	Mus musculus
6	ENCFF461YZX	bigBed narrowPeak	peaks	ENCSR547PLI	ChIP-seq	UBERON:0001049	neural tube	tissue	embryonic	mixed	Mus musculus
7	ENCFF407KIQ	bed narrowPeak	peaks	ENCSR547PLI	ChIP-seq	UBERON:0001049	neural tube	tissue	embryonic	mixed	Mus musculus
8	ENCFF559AVR	bed narrowPeak	peaks	ENCSR547PLI	ChIP-seq	UBERON:0001049	neural tube	tissue	embryonic	mixed	Mus musculus
9	ENCFF603TPC	bed narrowPeak	peaks	ENCSR547PLI	ChIP-seq	UBERON:0001049	neural tube	tissue	embryonic	mixed	Mus musculus
10	ENCFF296LTU	bigBed narrowPeak	peaks	ENCSR547PLI	ChIP-seq	UBERON:0001049	neural tube	tissue	embryonic	mixed	Mus musculus
11	ENCFF203NAR	bigBed narrowPeak	peaks	ENCSR547PLI	ChIP-seq	UBERON:0001049	neural tube	tissue	embryonic	mixed	Mus musculus
12	ENCFF335QHI	bigWig	signal p-value	ENCSR547PLI	ChIP-seq	UBERON:0001049	neural tube	tissue	embryonic	mixed	Mus musculus
13	ENCFF281WLU	bigWig	signal p-value	ENCSR547PLI	ChIP-seq	UBERON:0001049	neural tube	tissue	embryonic	mixed	Mus musculus
14	ENCFF649ZBD	bigWig	signal p-value	ENCSR547PLI	ChIP-seq	UBERON:0001049	neural tube	tissue	embryonic	mixed	Mus musculus
15	ENCFF858EFR	bigWig	fold change over control	ENCSR547PLI	ChIP-seq	UBERON:0001049	neural tube	tissue	embryonic	mixed	Mus musculus
16	ENCFF789MSC	bigWig	fold change over control	ENCSR547PLI	ChIP-seq	UBERON:0001049	neural tube	tissue	embryonic	mixed	Mus musculus
17	ENCFF806WGA	bigWig	fold change over control	ENCSR547PLI	ChIP-seq	UBERON:0001049	neural tube	tissue	embryonic	mixed	Mus musculus
18	ENCFF089YRA	bed narrowPeak	replicated peaks	ENCSR547PLI	ChIP-seq	UBERON:0001049	neural tube	tissue	embryonic	mixed	Mus musculus
19	ENCFF096JIB	bigBed narrowPeak	replicated peaks	ENCSR547PLI	ChIP-seq	UBERON:0001049	neural tube	tissue	embryonic	mixed	Mus musculus
20	ENCFF002AOI	fastq	reads	ENCSR511LWL	ChIP-seq	UBERON:0001049	neural tube	tissue	embryonic	mixed	Mus musculus
21	ENCFF002AOO	fastq	reads	ENCSR511LWL	ChIP-seq	UBERON:0001049	neural tube	tissue	embryonic	mixed	Mus musculus
22	ENCFF683EHP	bam	alignments	ENCSR511LWL	ChIP-seq	UBERON:0001049	neural tube	tissue	embryonic	mixed	Mus musculus
23	ENCFF632XMI	bam	alignments	ENCSR511LWL	ChIP-seq	UBERON:0001049	neural tube	tissue	embryonic	mixed	Mus musculus
24	ENCFF980AIP	bigBed narrowPeak	peaks	ENCSR511LWL	ChIP-seq	UBERON:0001049	neural tube	tissue	embryonic	mixed	Mus musculus
25	ENCFF994QFR	bed narrowPeak	peaks	ENCSR511LWL	ChIP-seq	UBERON:0001049	neural tube	tissue	embryonic	mixed	Mus musculus
26	ENCFF161OXM	bigBed narrowPeak	peaks	ENCSR511LWL	ChIP-seq	UBERON:0001049	neural tube	tissue	embryonic	mixed	Mus musculus
27	ENCFF151MLW	bigBed narrowPeak	peaks	ENCSR511LWL	ChIP-seq	UBERON:0001049	neural tube	tissue	embryonic	mixed	Mus musculus

Programmatic access via the ENCODE REST API

- All Portal content is accessible via URL's; just add `?format=json`
- The database record is returned in JSON format
- JSON can be parsed in your language of choice

GET_object.py

```
1  #!/usr/bin/env python
2
3  import requests
4
5  URL = 'https://www.encodeproject.org/experiments/ENCSR236EGS/?format=json'
6
7  response = requests.get(URL)
8
9  experiment = response.json()
10
11 print experiment['accession']
12 print experiment['description']
13
```

Programmatic access via the ENCODE REST API

GET_object.py

```
1  #!/usr/bin/env python
2
3  import requests
4
5  URL = 'https://www.encodeproject.org/experiments/ENCSR236EGS/?format=json'
6
7  response = requests.get(URL)
8
9  experiment = response.json()
10
11 print experiment['accession']
12 print experiment['description']
13
```

```
jseth:Keystone Epigenomics 2015 jseth$ ./GET_object.py
ENCSR236EGS
RNA-seq on a dissected area of layer V from an 8 month old male wild type C57Bl6 mouse
jseth:Keystone Epigenomics 2015 jseth$
```

Programmatic access via the ENCODE REST API

GET_search.py

```python
#!/usr/bin/env python

import requests

URL = ('https://www.encodeproject.org/search/?'
       'type=experiment&'
       'assay_term_name=ChIP-seq&'
       'replicates.library.biosample.donor.organism.scientific_name=Homo sapiens&'
       'target.investigated_as=transcription factor&'
       'replicates.library.biosample.biosample_type=in vitro differentiated cells&'
       'format=json')

response = requests.get(URL)

search_result = response.json()['@graph']

#extract and print the target for each experiment
print '\n'.join([experiment['target']['label'] for experiment in search_result])
```

Programmatic access via the ENCODE REST API

GET_search.py

```
1  #!/usr/bin/e
2
3  import reque
4
5  URL = ('http
6          'type=
7          'assay
8          'repli
9          'targe
10         'repli
11         'forma
12
13 response = r
14
15 search_result = response.json()['@graph']
16
17 #extract and print the target for each experiment
18 print '\n'.join([experiment['target']['label'] for experiment in search_result])
19
```

```
jseth:Keystone Epigenomics 2015 jseth$ ./GET_search.py
SMC3
RAD21
MXI1
EP300
CTCF
EZH2
EZH2
CTCF
POLR2AphosphoS5
REST
TAF1
jseth:Keystone Epigenomics 2015 jseth$
```

The ENCODE Portal: Recap

- Interactive access to ENCODE metadata via faceted browsing and search
- Interactive retrieval of ENCODE data one file at a time
- Batch download of ENCODE metadata and data files
- Programmatic access using the ENCODE REST API

Next: ENCODE Data Analysis Pipelines

- What do they produce?
- How can they be run?

Pipelines Demonstration and Exercise

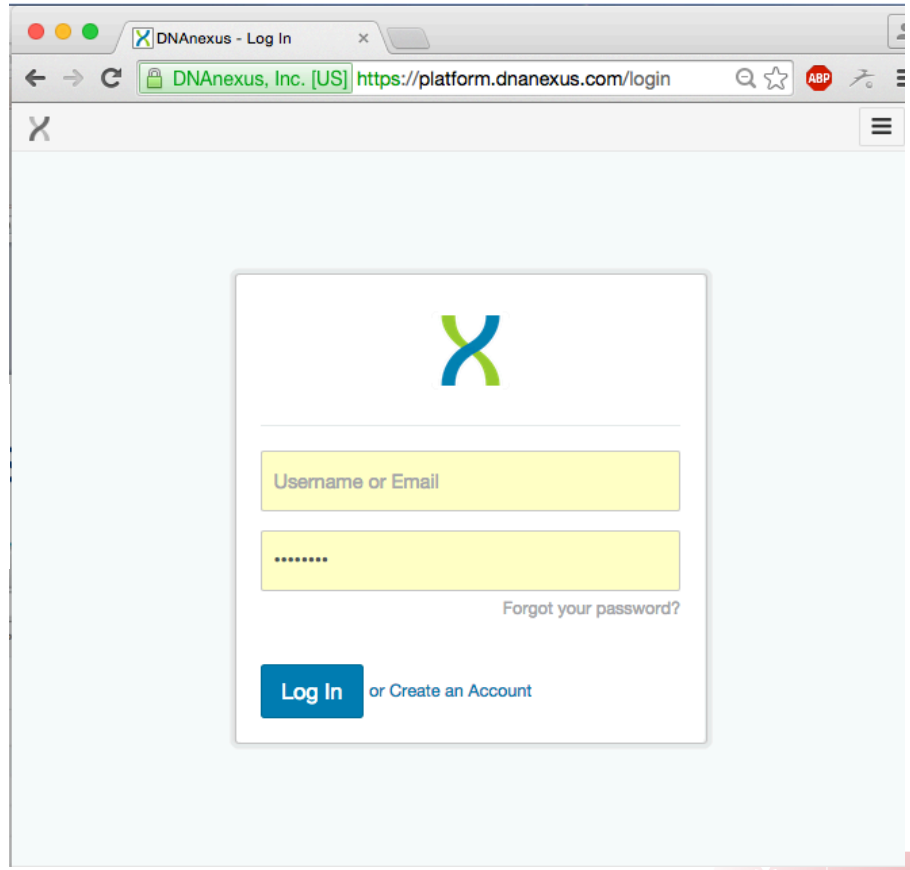
To set up an account:
https://www.encodeproject.org/tutorials/apbc-2016/

Click “Prepare to run web-based pipelines”

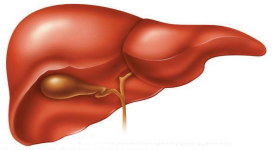
Instructions for setting up your DNAnexus environment to run ENCODE pipelines are here:

Prepare to run web-based pipelines

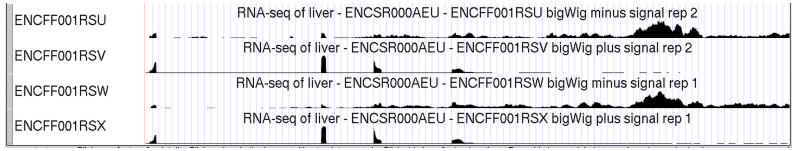
Log in ->



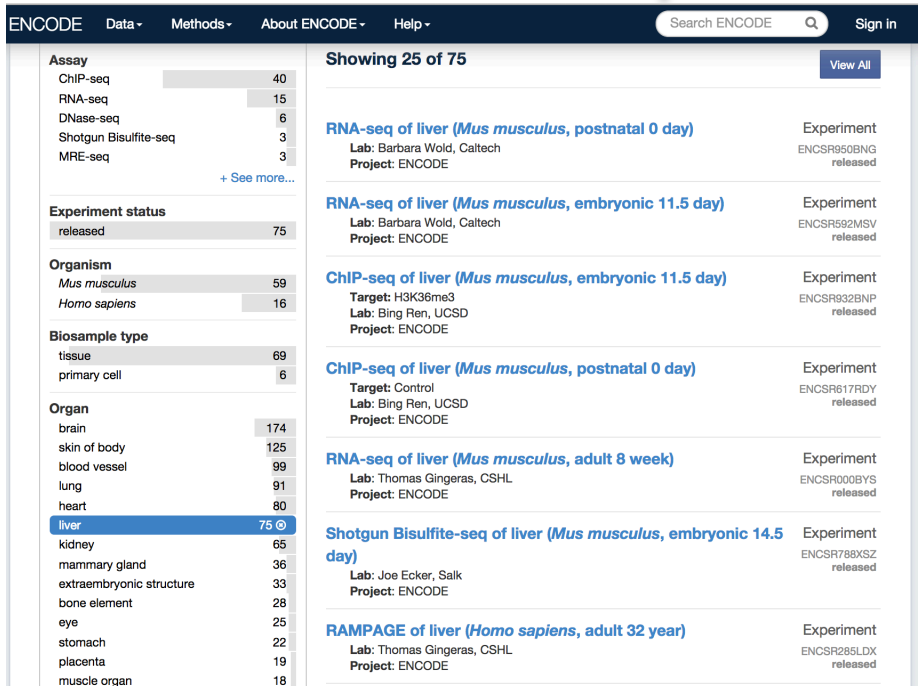
DCC Delivers ENCODE Data



```
+
CCCFFFFFHHHHHGIJJIGGHEIIEGGEGGIJJBHIG
@BI:SL-HAB:D0RRAACXX:8:2309:21201:7829 1:X:0:GCCGTCGA
CTAACCCTAACCCTAACCCTAACCCTAACCCTAACC
+
CCCFFFFFHHHHHHJJJJJJJGJJJJIIJJJJGGIGJ
@BI:SL-HAB:D0RRAACXX:8:2113:4623:40045 1:X:0:GCCGTCGA
GGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTA
+
??@ADDBDH:CDHHI+AEFHI?GGHII:EFIII?F=
@BI:SL-HAB:D0RRAACXX:8:2206:11680:21762 1:X:0:GCCGTCGA
AGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTT
+
```



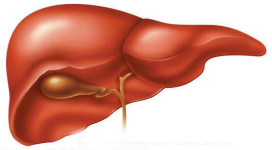
Sample → Library → Primary Data → Processed Data



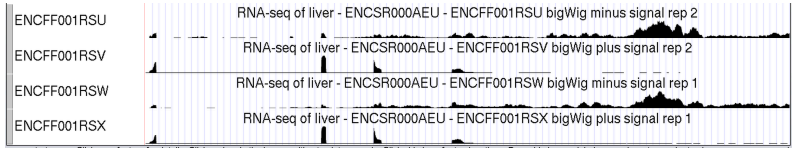
ENCFF001RTZ	fastq	reads	2	(1, 1)	Esther Chan	2013-08-15	Download pending
ENCFF001RSV	bigWig	plus signal		(2, 1)	Esther Chan	2013-07-26	Download pending
ENCFF001RSU	bigWig	minus signal		(2, 1)	Esther Chan	2013-07-26	Download pending

AWS S3 Bucket
ENCODE Files

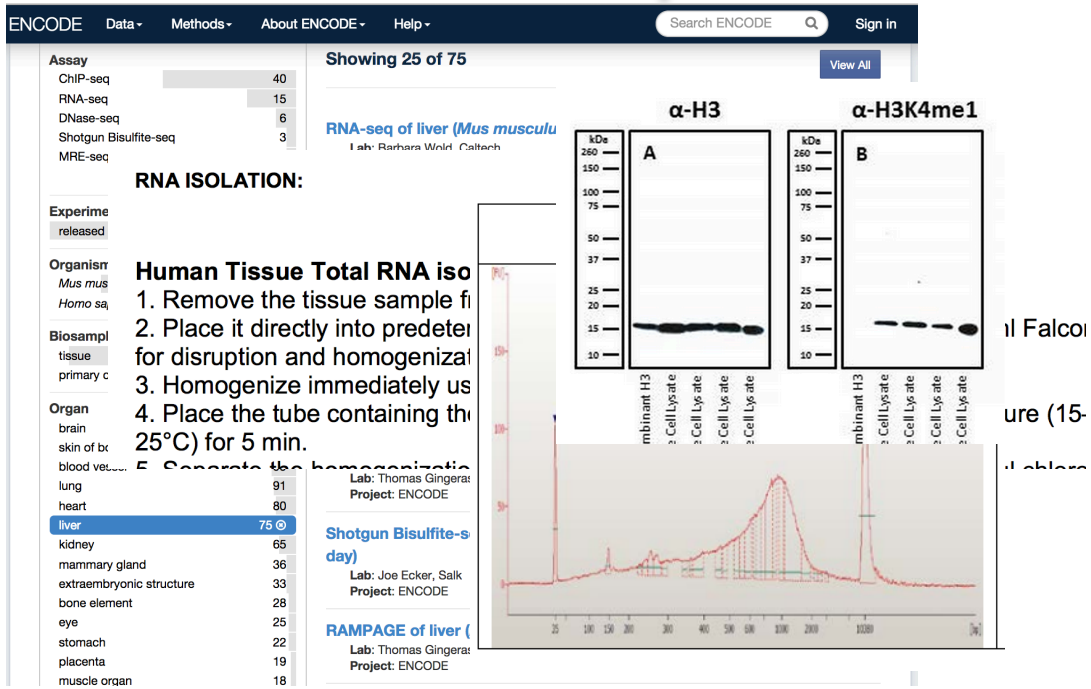
ENCODE DCC Delivers ENCODE Metadata



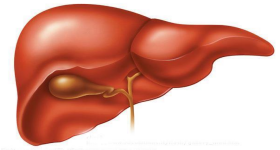
```
+
CCCFFFFFHHHHHGIJJIGGHEIIEGGEGGIJJBHIG
@BI:SL-HAB:D0RRAACXX:8:2309:21201:7829 1:X:0:GCCGTCGA
CTAACCCTAACCCTAACCCTAACCCTAACCCTAACC
+
CCCFFFFFHHHHHHJJJJJJJGJJJJIIJJJJGGIGJ
@BI:SL-HAB:D0RRAACXX:8:2113:4623:40045 1:X:0:GCCGTCGA
GGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTA
+
??@ADDBDH:CDHHI+AEFHI?GGHII:EFIII?F=
@BI:SL-HAB:D0RRAACXX:8:2206:11680:21762 1:X:0:GCCGTCGA
AGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTT
+
```

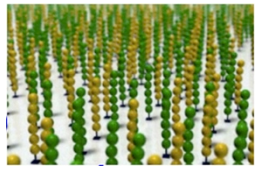


Sample → Library → Primary Data → Processed Data

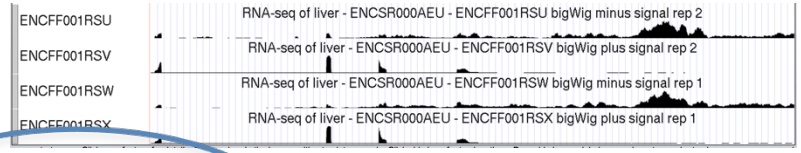


ENCODE Analysis Pipelines as Deliverables





```
+
CCCFFFFFHHHHHGIJJIGGHEIIEGGEGGIJJBHIG
@BI:SL-HAB:D0RRAACXX:8:2309:21201:7829 1:X:0:GCCGTCGA
CTAACCCTAACCCTAACCCTAACCCTAACCCTAACC
+
CCCFFFFFHHHHHHJJJJJJJGJJJJIIJJJJGGIGJ
@BI:SL-HAB:D0RRAACXX:8:2113:4623:40045 1:X:0:GCCGTCGA
GGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTA
+
??@ADDBDH:CDHHI+AEFHI?GGHII:EFIII?F=
@BI:SL-HAB:D0RRAACXX:8:2206:11680:21762 1:X:0:GCCGTCGA
AGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTT
+
```



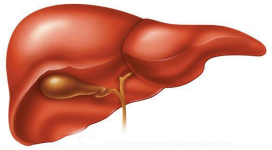
Sample → Library → Primary Data → Processed Data

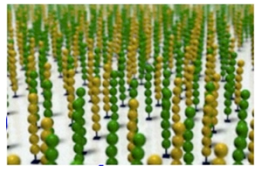
Goals:

1. Deploy ENCODE-defined pipelines for ChIP-seq, RNA-seq, DNase-seq, methylation.
2. Use those pipelines to generate the standard ENCODE peaks, quantitations, CpG.
3. Capture metadata to make clear what software, versions, parameters, inputs were used.
4. Capture, accession, and distribute the output.
5. Deliver *exactly the same* pipelines in a form that *anyone can run* on their data or with ENCODE data – one experiment or 1000.

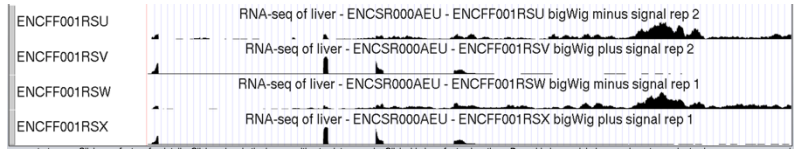
Replicability – Provenance – Ease of Use – Scalability

Deployment Platform Considerations





```
+
CCCFFFFFHHHHHGIJJIGGHEIIEGGEGGIJJBHIG
@BI:SL-HAB:D0RRAACXX:8:2309:21201:7829 1:X:0:GCCGTCGA
CTAACCCTAACCCTAACCCTAACCCTAACCCTAACC
+
CCCFFFFFHHHHHHJJJJJJJGJJJJIIJJJJGGIGJ
@BI:SL-HAB:D0RRAACXX:8:2113:4623:40045 1:X:0:GCCGTCGA
GGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTA
+
??@ADDBDH:CDHHI+AEFHI?GGHII:EFIII?F=
@BI:SL-HAB:D0RRAACXX:8:2206:11680:21762 1:X:0:GCCGTCGA
AGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTT
+
```



Sample → Library → Primary Data → Processed Data

	Develop	Share	Run	Elastic	Provenance	Cost
HPC Cluster (Scripts)	Hard	Hard	Hard	Cluster-Dependent	Moderate	Obscure/Subsidized
HPC Container	Hard	Moderate	Moderate	Cluster-Dependent	Good	Obscure/Subsidized
Web/Cloud	Moderate	Easy	Easy	Highly	Excellent	Apparent but Low

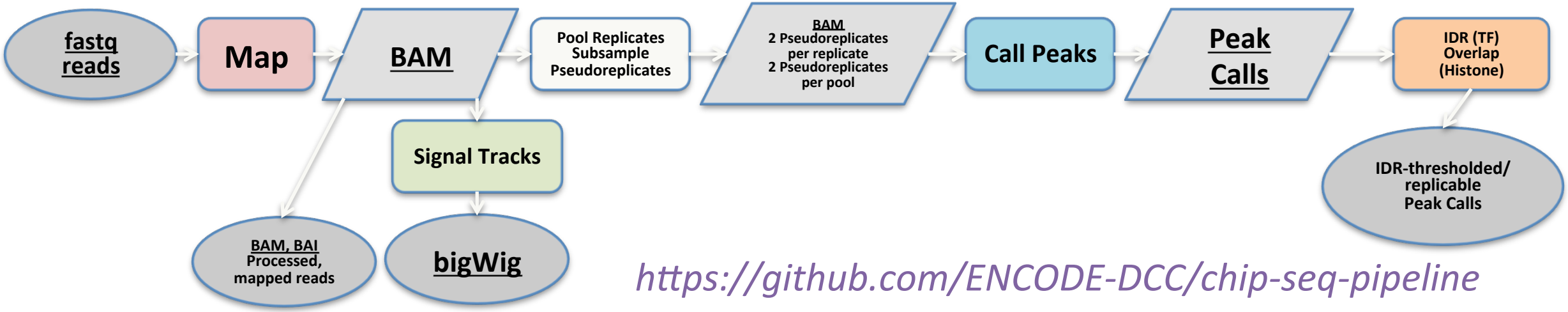
Replicability – Provenance – Ease of Use – Scalability

We chose to deploy first to a web/cloud-based platform, DNAnexus

Code is open source and adaptable for deployment to your HPC environment

https://github.com/ENCODE-DCC

Schema: ENCODE ChIP-seq IDR Pipeline



https://github.com/ENCODE-DCC/chip-seq-pipeline

Target	Key Software	Input Files	Output Files	QA Metrics
TF's	bwa Picard markDuplicates samtools MACS2 (Signal tracks) SPP (PeakSeq, GEM future) IDR2	fastq's (SE or PE) Two biological replicates Matched controls	One bam per replicate bigWig fold signal over control bigWig p-value signal over control bed/bigBed true replicates peaks bed/bigBed pooled replicates peaks bed/bigBed IDR thresholded peaks	NRF (Non-redundant fraction) PBC1 and 2 (PCR bottleneck coefficients) Number of distinct uniquely-mapping reads NSC/RSC (Strand cross-correlation) IDR Rescue Ratio IDR Self-Consistency Ratio IDR Reproducibility Test
Histone Mods	MACS2 for peaks Overlap thresholding IDR2 (future)		bed/bigBed Replicated peaks	

Pipelines Demonstration and Exercise

To set up an account:
https://www.encodeproject.org/tutorials/apbc-2016/

Instructions for setting up your DNAnexus environment to run ENCODE pipelines are here:

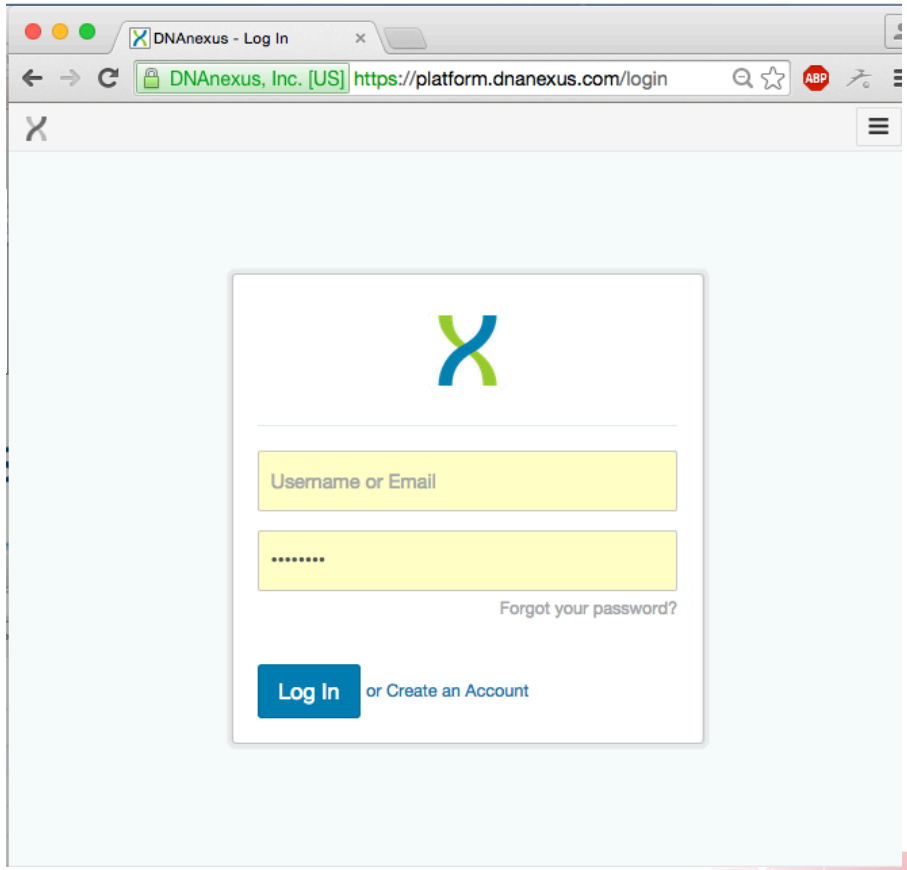
Prepare to run web-based pipelines

J. Seth Strattan	ENCODE DCC, Stanford University	Slides (coming soon) | Prepare to run web-based pipelines | ChIP-seq Pipeline How-to | RNA-seq Pipeline How-to

Exercises

Histone ChIP-seq RNA-seq

Log in ->



Uniformly Processed Data On the ENCODE Portal

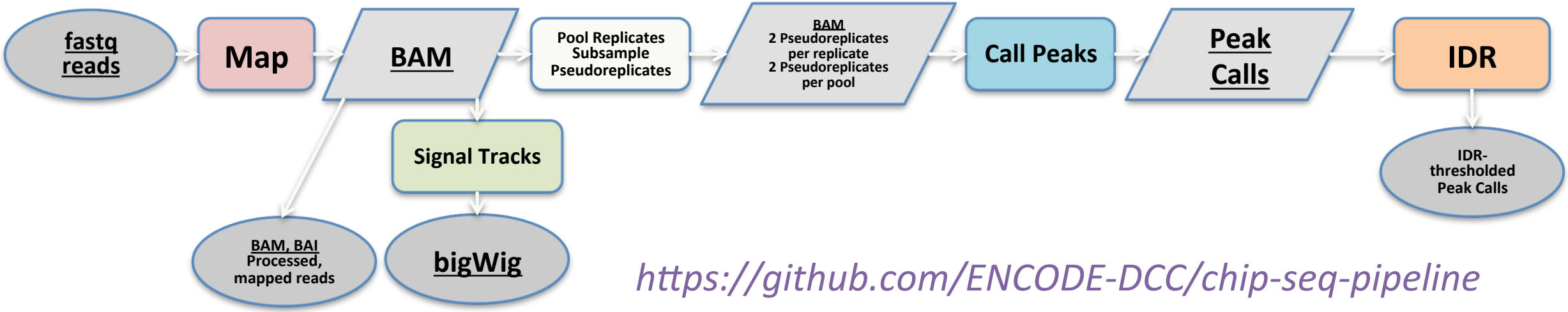
Histone ChIP-seq Example
https://www.encodeproject.org/experiments/ENCSR087PLZ/

- Pipeline graph shows relationships between files
- Click on files to see more file metadata and download links
- Click on steps to see more software metadata and download links

Transcription Factor ChIP-seq Example
https://www.encodeproject.org/experiments/ENCSR077DKV/

- Same mapping, signal tracks and peak calls
- Also have the IDR-thresholded peak calls
- "Conservative" set, based on "true" replicates; "optimal" set if peaks can be rescued by pseudo-replication.

ENCODE ChIP-seq Quality Metrics: Resources

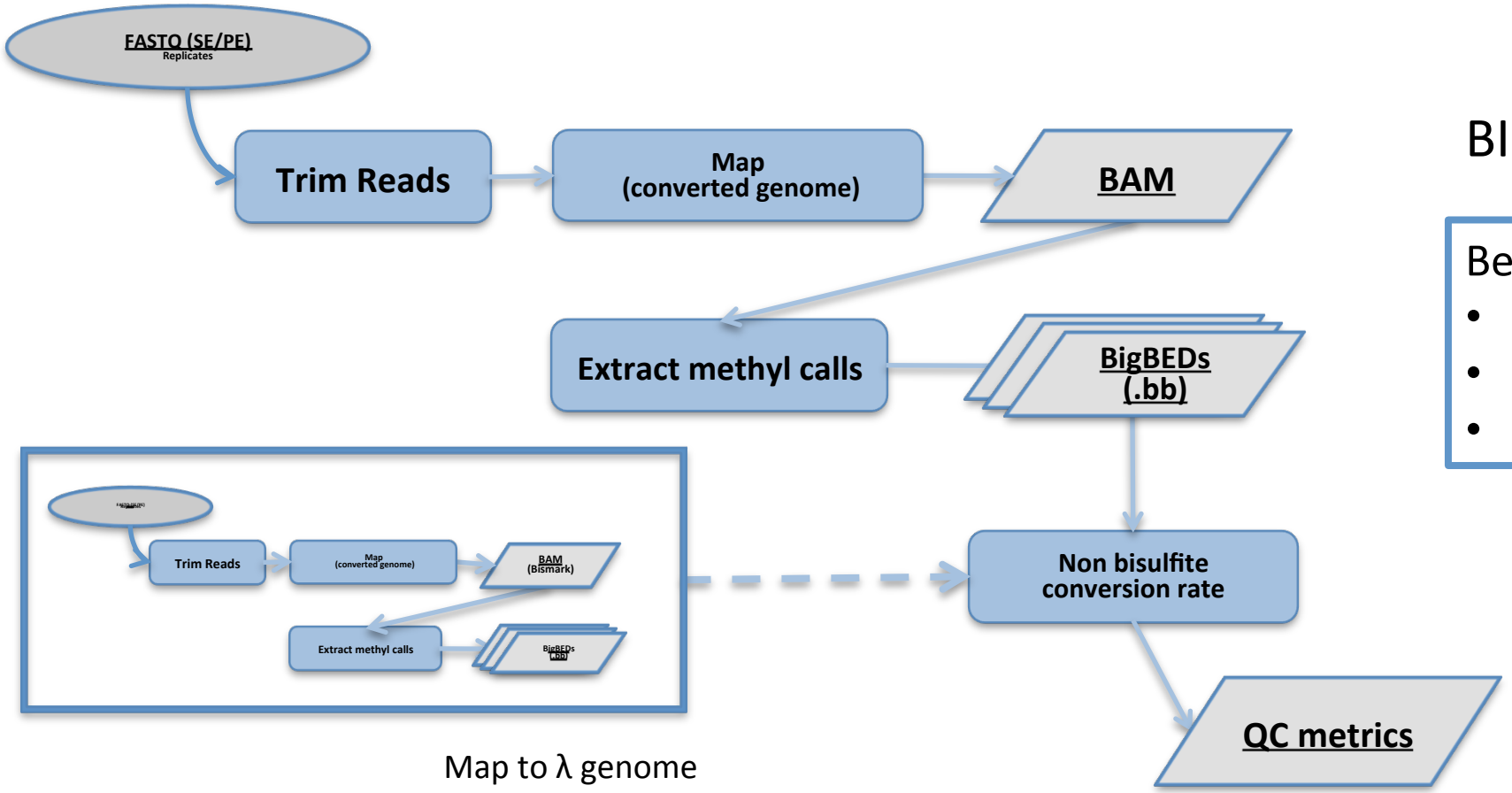


https://github.com/ENCODE-DCC/chip-seq-pipeline

Estimates	Description	References
Depth	Number of uniquely mapping reads Number of distinct uniquly mapping reads	Jung YL, et al. Nucleic Acids Research. 2014;42(9):e74
Library Complexity **ChIP Quality**	Non-Redundant Fraction PCR Bottleneck Coefficient Normalized Strand Cross-Correlation Relative Strand Cross-Correlation	Landt S, et al. Genome Res. 2012. 22: 1813-1831
Replicate Concordance	IDR Rescue Ratio IDR Self-Consistency Ratio IDR Reproducibility Test	Li Q, et al. Annals Applied Statistics. 2011, Vol. 5, No. 3, 1752–1779

Schema: ENCODE WGBS Pipeline

https://github.com/ENCODE-DCC/dna-me-pipeline



Map to λ genome

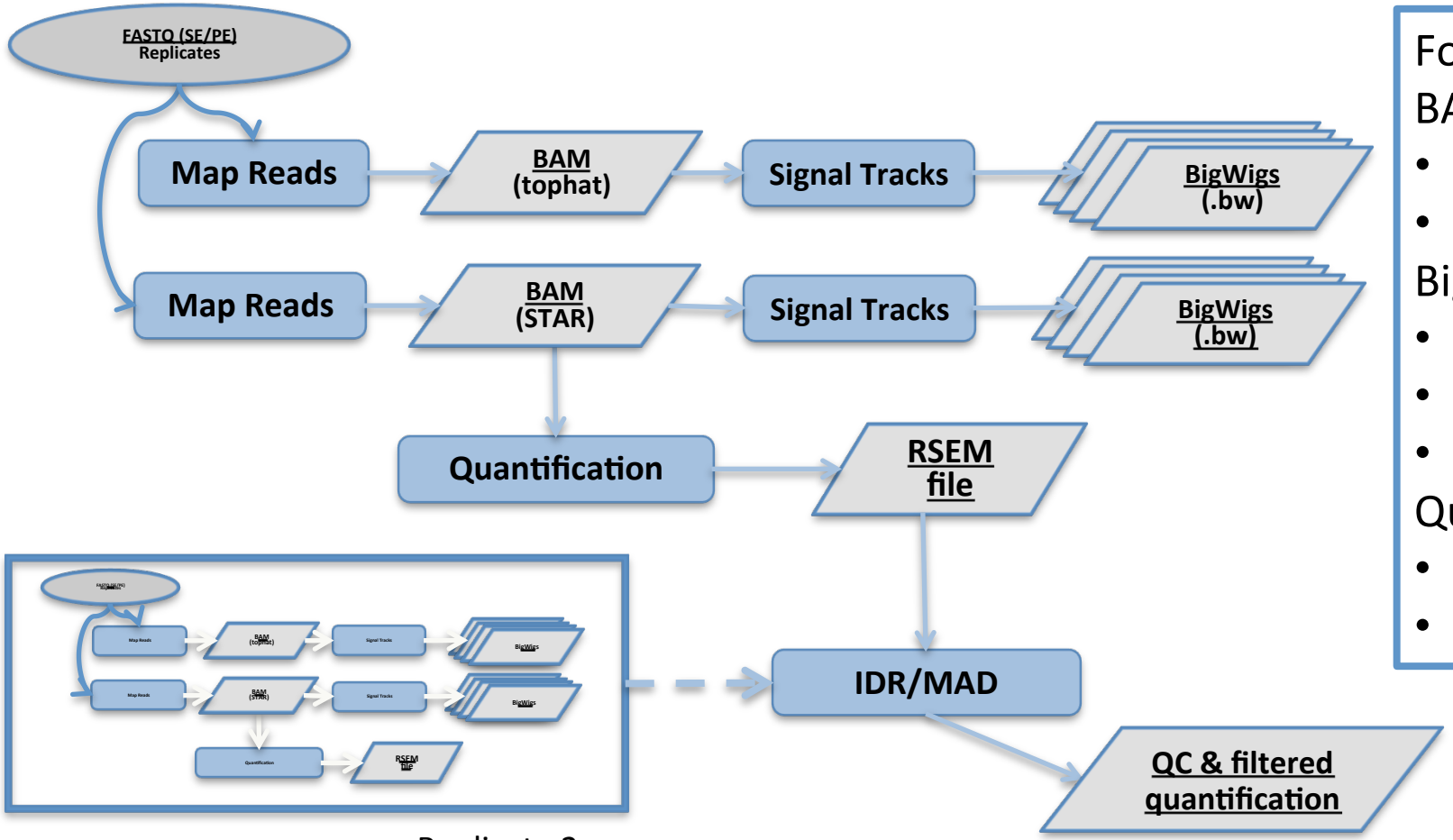
BISMARK (v 0.10)

Bed/BigBed files for:

- CG context
- CHG context
- CHH context

Schema: ENCODE RNA-seq Pipeline

https://github.com/ENCODE-DCC/long-rna-seq-pipeline



Replicate 2

For *each* Mapper (STAR, tophat)
BAM files:
- mapped to genome
- mapped to transcriptome

BigWig files:
- plus/minus strand (paired)
- uniquely mapped
- multi+uniquely mapped

Quantifications (RSEM):
- genome
- transcriptome

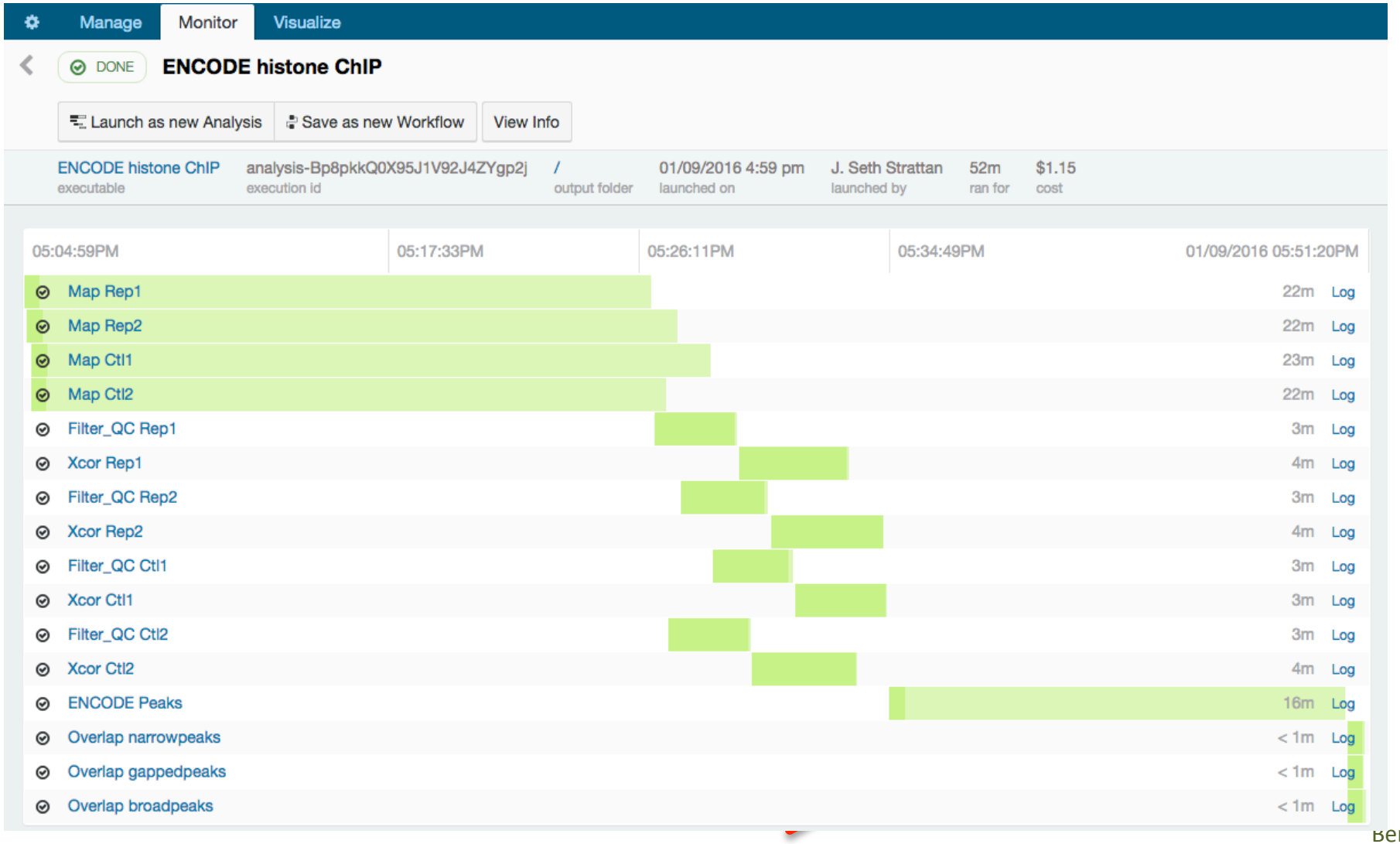
Uniformly Processed Data On the ENCODE Portal

RNA-seq Example

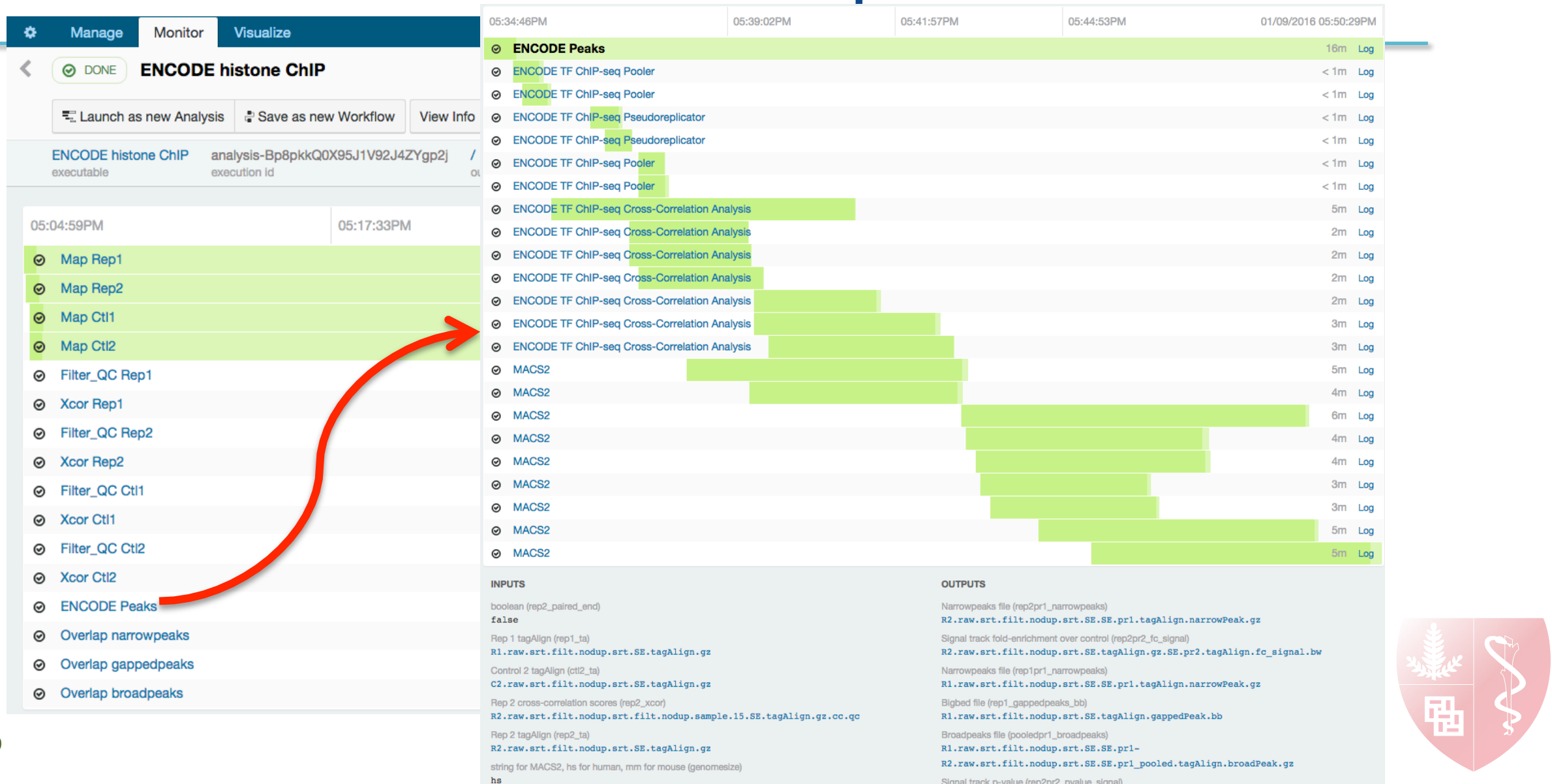
https://www.encodeproject.org/experiments/ENCSR368QPC/

- Pipeline graph shows relationships between files
- Click on files to see more file metadata and download links
- Click on steps to see more software metadata and download links

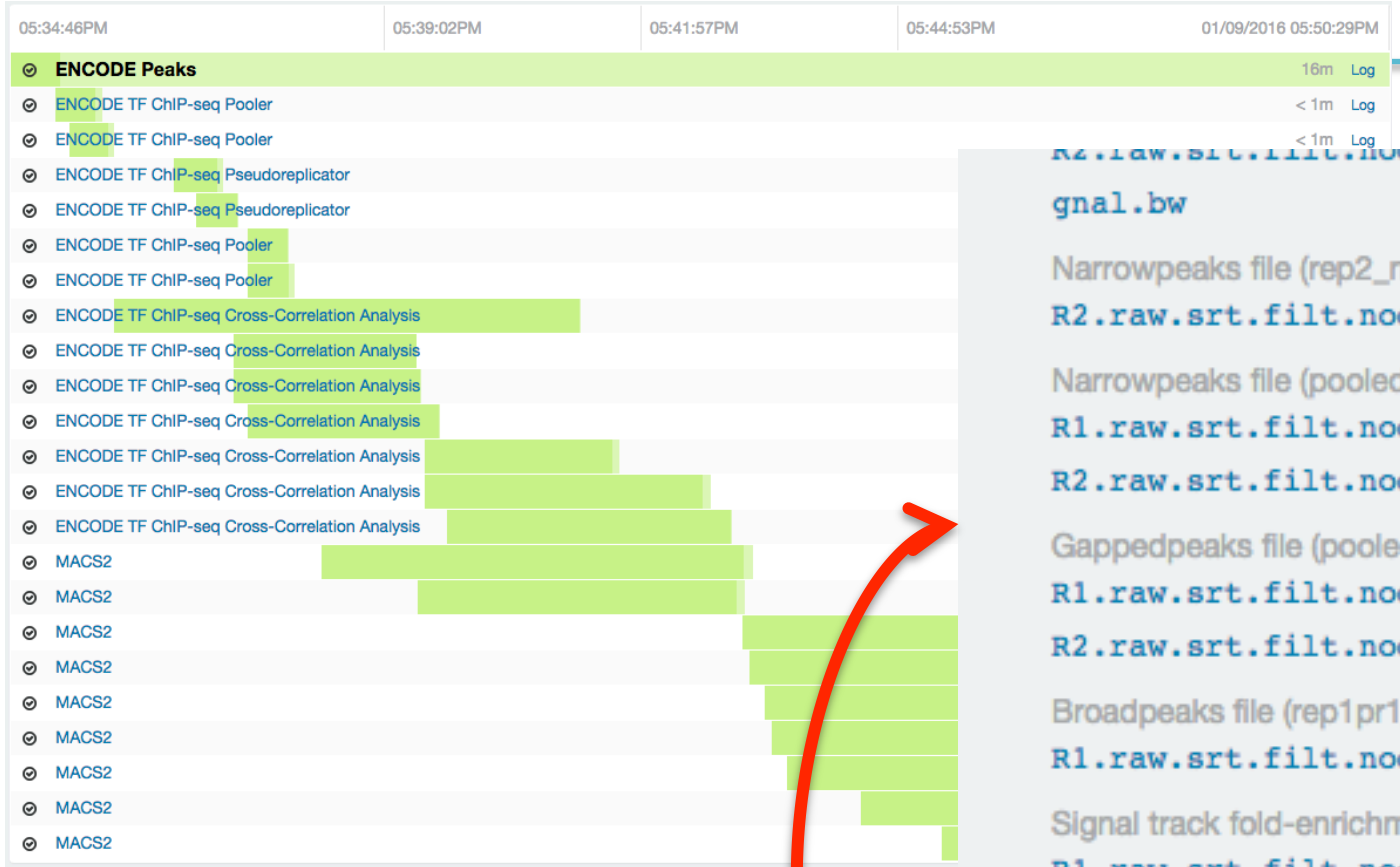
Results from the ChIP-seq exercise



Results from the ChIP-seq exercise



Results from the ChIP-seq exercise



gnal.bw

Narrowpeaks file (rep2_narrowpeaks)

R2.raw.srt.filt.nodup.srt.SE.tagAlign.narrowPeak.gz

Narrowpeaks file (pooled_narrowpeaks)

R1.raw.srt.filt.nodup.srt.SE-

R2.raw.srt.filt.nodup.srt.SE_pooled.tagAlign.narrowPeak.gz

Gappedpeaks file (pooled_gappedpeaks)

R1.raw.srt.filt.nodup.srt.SE-

R2.raw.srt.filt.nodup.srt.SE_pooled.tagAlign.gappedPeak.gz

Broadpeaks file (rep1pr1_broadpeaks)

R1.raw.srt.filt.nodup.srt.SE.SE.pr1.tagAlign.broadPeak.gz

Signal track fold-enrichment over control (rep1pr2_fc_signal)

R1.raw.srt.filt.nodup.srt.SE.tagAlign.gz.SE.pr2.tagAlign.fc_signal.bw

INPUTS

boolean (rep2_paired_end)
`false`

Rep 1 tagAlign (rep1_ta)
`R1.raw.srt.filt.nodup.srt.SE.tagAlign.gz`

Control 2 tagAlign (ctl2_ta)
`C2.raw.srt.filt.nodup.srt.SE.tagAlign.gz`

Rep 2 cross-correlation scores (rep2_xcor)
`R2.raw.srt.filt.nodup.srt.filt.nodup.sample.15.SE.tagAlign.gz.cc.qc`

Rep 2 tagAlign (rep2_ta)
`R2.raw.srt.filt.nodup.srt.SE.tagAlign.gz`

string for MACS2, hs for human, mm for mouse (genomesize)
`hs`

OUTPUTS

Narrowpeaks file (rep2pr1_narrowpeaks)
`R2.raw.srt.filt.nodup.srt.SE.SE.pr1.tagAlign.narrowPeak.gz`

Signal track fold-enrichment over control (rep2pr2_fc_signal)
`R2.raw.srt.filt.nodup.srt.SE.tagAlign.gz.SE.pr2.tagAlign.fc_signal.bw`

Narrowpeaks file (rep1pr1_narrowpeaks)
`R1.raw.srt.filt.nodup.srt.SE.SE.pr1.tagAlign.narrowPeak.gz`

Bigbed file (rep1_gappedpeaks_bb)
`R1.raw.srt.filt.nodup.srt.SE.tagAlign.gappedPeak.bb`

Broadpeaks file (pooledpr1_broadpeaks)
`R1.raw.srt.filt.nodup.srt.SE.SE.pr1-`
`R2.raw.srt.filt.nodup.srt.SE.SE.pr1_pooled.tagAlign.broadPeak.gz`

Signal track p-value (rep2pr2_pvalue_signal)

J. Seth Strattan, PhD ENCODE DCC

Results from the ChIP-seq exercise

Narrowpeaks file (rep1_narrowpeaks)

R1.raw.srt.filt.nodup.srt.SE.tagAlign.narrowPeak.gz

Gappedpeaks file (rep1_gappedpeaks)

R1.raw.srt.filt.nodup.srt.SE.

Signal track fold-enrichment over contr

R1.raw.srt.filt.nodup.srt.SE-

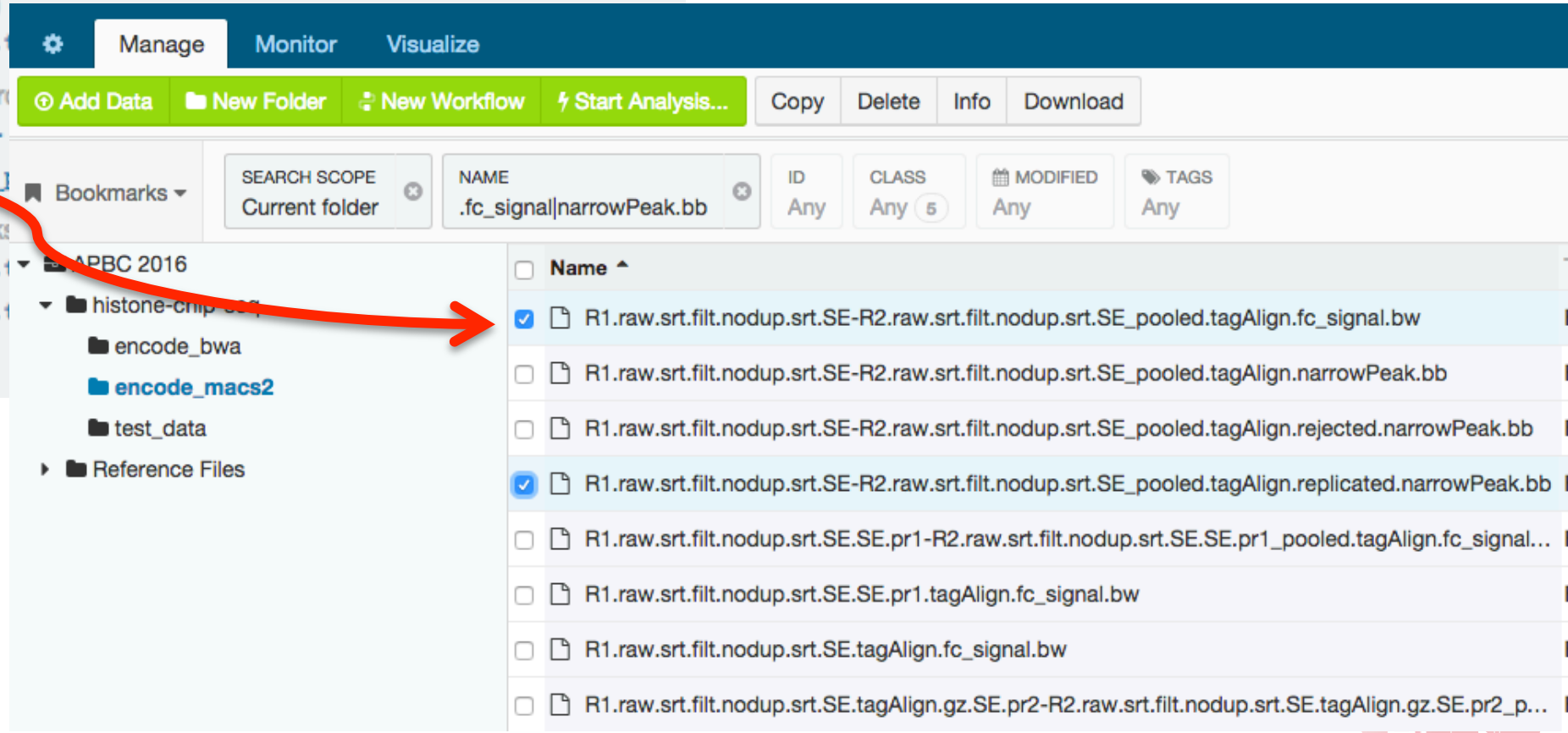
R2.raw.srt.filt.nodup.

Broadpeaks file (pooledpr2_broadpeaks

R1.raw.srt.filt.nodup.srt.SE.

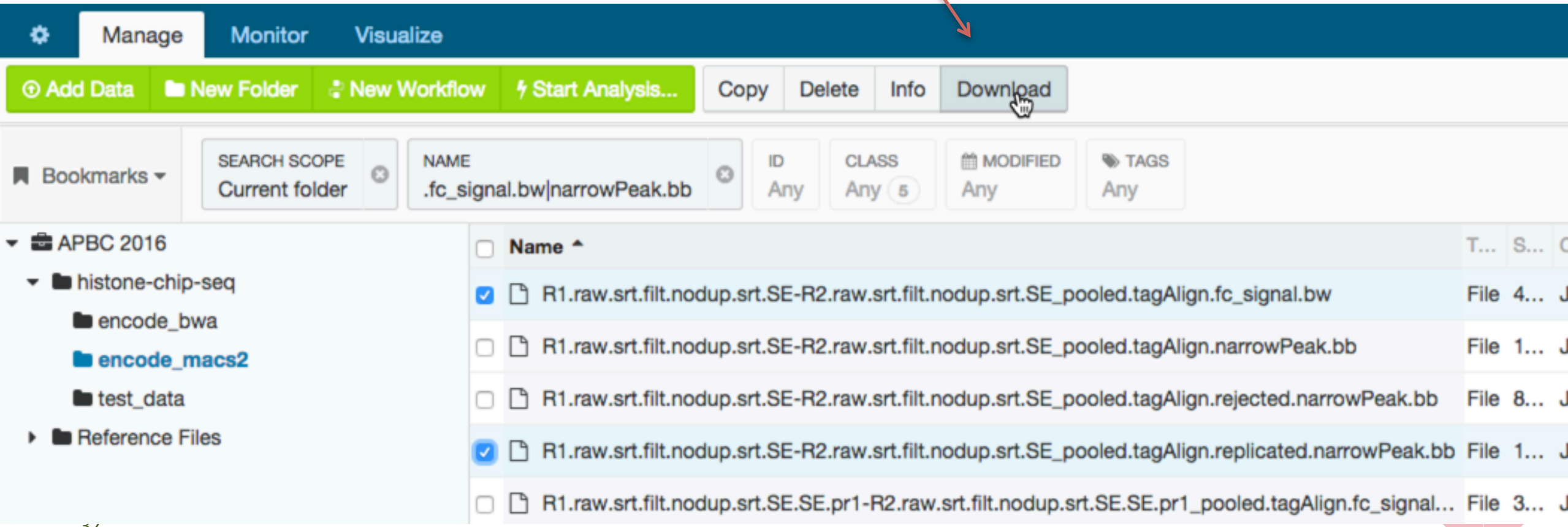
R2.raw.srt.filt.nodup.srt.SE.

Peak.gz



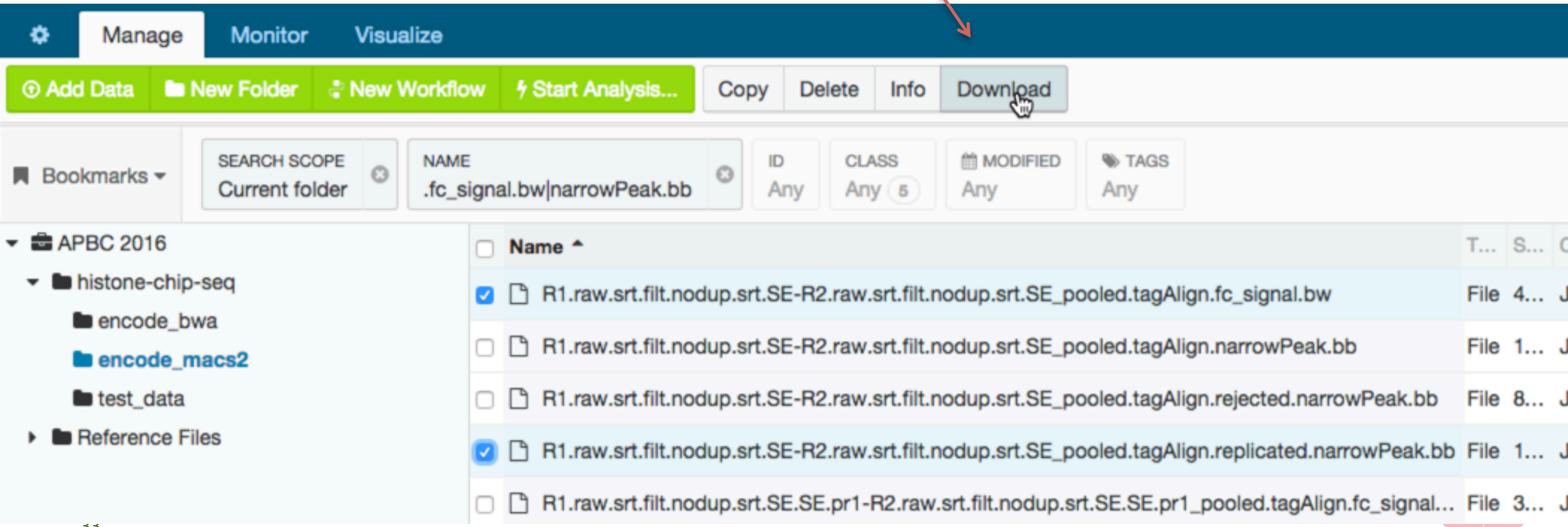
Results from the ChIP-seq exercise

"Download" to generate temporary URL's to the selected files

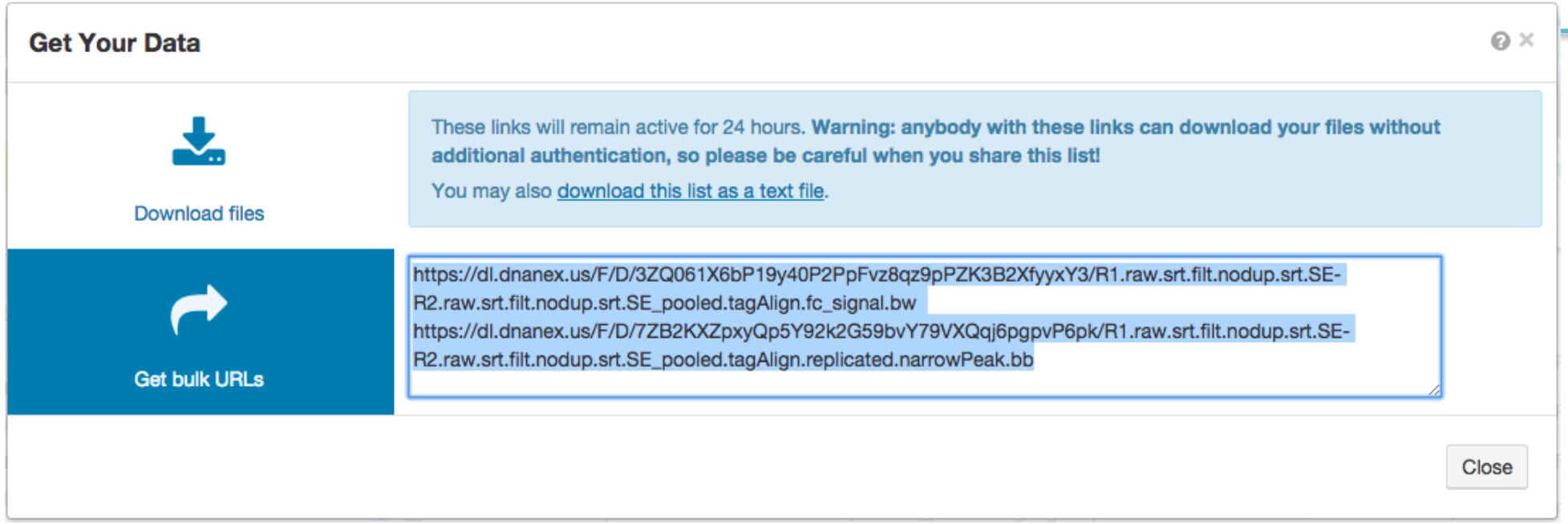


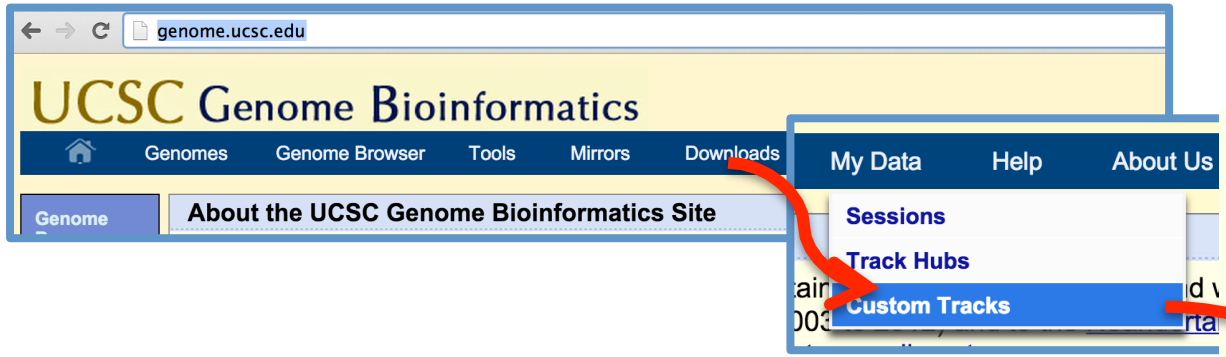
Results from the ChIP-seq exercise

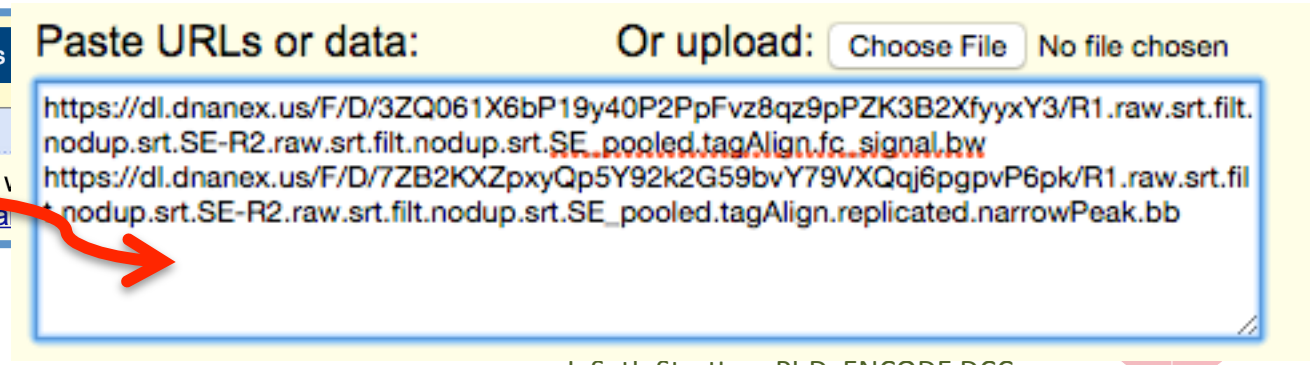
"Download" to generate temporary URL's to the selected files



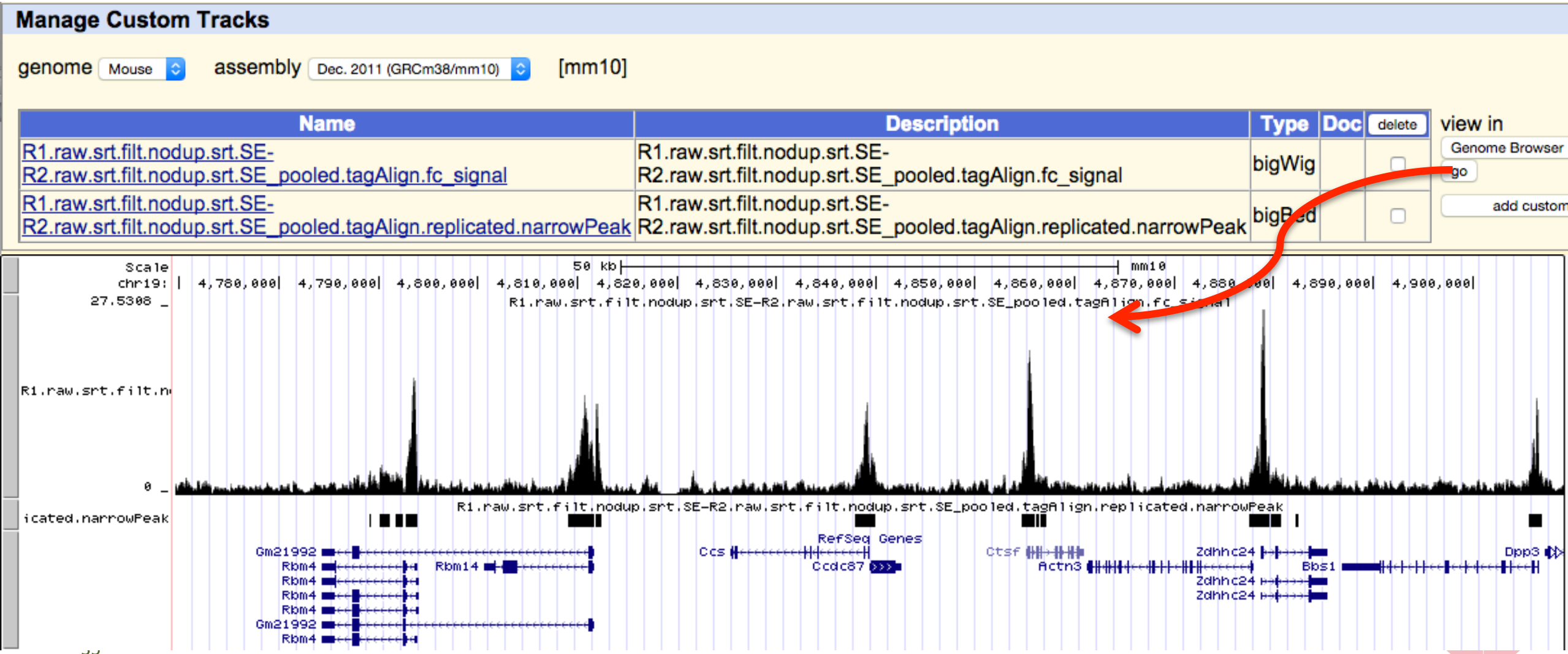
Visualize on the UCSC Genome Browser







Visualize on the UCSC Genome Browser

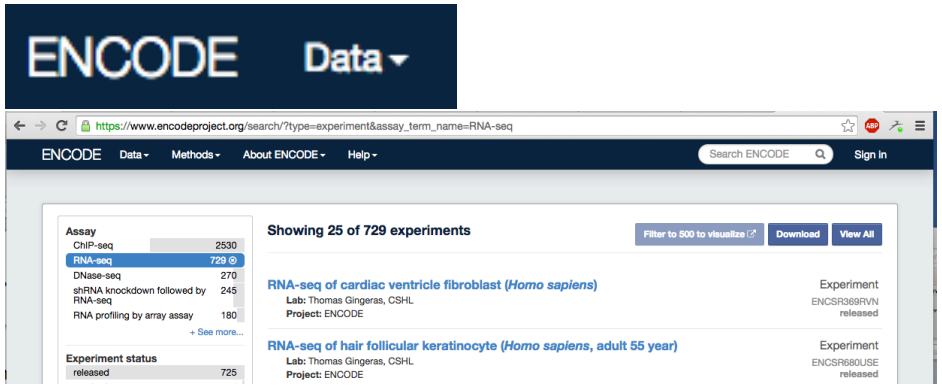


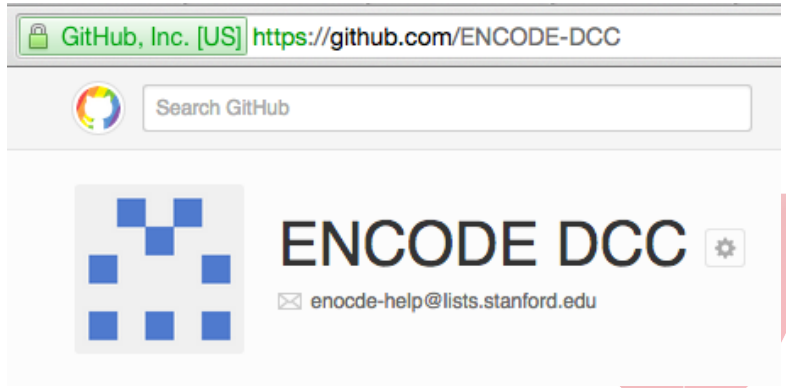
Pipeline Workshop Summary

DCC Goals:

1. Deploy ENCODE-defined pipelines for ChIP-seq, RNA-seq, DNase-seq, methylation.
2. Use those pipelines to generate the standard ENCODE peaks, quantitations, CpG.
3. Capture metadata to make clear what software, versions, parameters, inputs were used.
4. Capture, accession, and distribute the output.
5. Deliver *exactly the same* pipelines in a form that *anyone can run* on their data or with ENCODE data – one experiment or 1000.

Replicability – Provenance – Ease of Use – Scalability





Contributors

ENCODE Data Coordinating Center
Mike Cherry, PI, Stanford
Jim Kent, co-PI, UCSC
Eurie Hong, Project Manager

Pipeline Developers
Ben Hitz, WGBS, Software Lead
Tim Dreszer, RNA-seq, DNAse-seq
J. Seth Strattan, ChIP-seq

Portal Developers
Laurence Rowe
Nikhil Podduturi
Forrest Tanaka

Data Wranglers
Esther Chan
Jean Davidson
Venkat Malladi
Cricket Sloan
J. Seth Strattan

QA & Biocuration Assistance
Brian Lee
Marcus Ho
Aditi Narayanan

Support Staff
Stuart Miyasato
Matt Simison
Zhenhua Wang

ENCODE Data Analysis Center
Zhiping Weng, PI, University of Massachusetts
Mark Gerstein, co-PI, Yale

Methylation
Junko Tsuji, U Mass
Eric Mendenhall, U Alabama, HAIB

RNA-seq
Alex Dobin, CSHL
Carrie Davis, CSHL
Rafael Irizarryt, Harvard
Xintao Wei, UConn
Brent Gravely, UConn
Colin Dewey, U Wisconsin
Roderic Guigó, CRG
Sarah Djebali, CRG

ChIP-seq
Anshul Kundaje, Stanford
Nathan Boley, Stanford
Jin Lee, Stanford

DNAnexus
Mike Lin
Andey Kislyuk
Singer Ma
Brett Hannigan
Ohad Rodeh
Joe Dale
George Asimenos


@encodedcc


encode-help@lists.stanford.edu

https://github.com/ENCODE-DCC/