

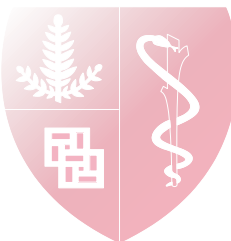
ENCODE ANALYSIS PIPELINES

J. Seth Strattan, PhD

ENCODE Data Coordinating Center (DCC)

ENCODE User's Meeting Workshop

June, 2016



Analysis Pipeline Demonstration and Workshop

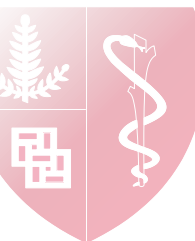
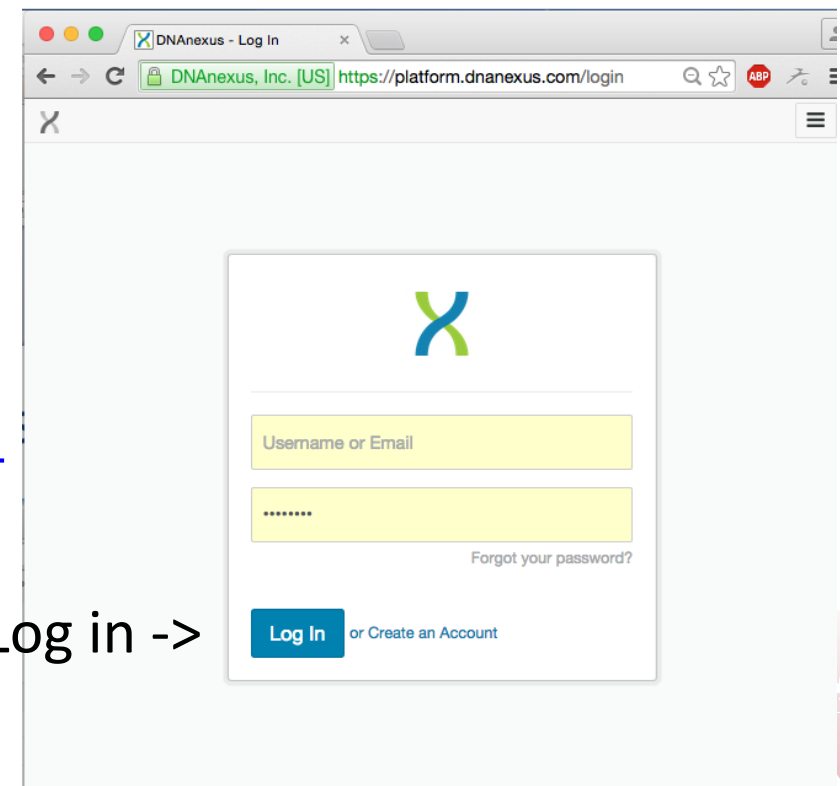
Workshop Session 4: ENCODE Uniform Processing Pipelines				
7	DCC tutorial Part II: Pipeline overview	J. Seth Strattan and Ben Hitz	Stanford University	Preparing to Run ENCODE Pipelines

To set up an account:

<https://www.encodeproject.org/tutorials/encode-meeting-2016/>

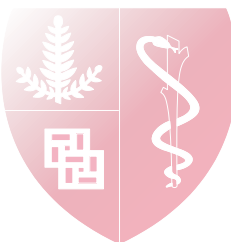
Scroll down to [“Preparing to Run ENCODE Pipelines”](#)

Log in ->



Workshop Goals

- Introduce the ENCODE Analysis Pipelines.
- Run the transcription factor ChIP-seq pipeline on a ZBED1 ChIP experiment in K562.
- Run the long RNA-seq pipeline on a total-RNA experiment from a human tissue sample.
- Understand the pipeline inputs, outputs, and QC metrics and how to navigate them.
- Visualize the results of your analyses.
- Take home the ability to replicate ENCODE analyses on your own data.



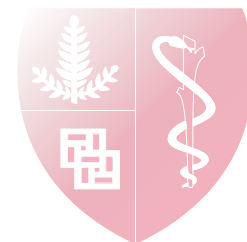
ENCODE Analysis Pipelines as Deliverables



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



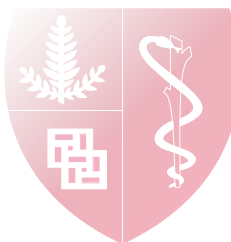
	Develop	Share	"First Bite"	Elastic	Provenance	Cost
HPC Cluster (Scripts)	Hard	Hard	Hard	Cluster-Dependent	Moderate	Obscure/Subsidized
HPC Container	Hard	Moderate	Moderate	Cluster-Dependent	Good	Obscure/Subsidized
Web/Cloud	Moderate	Easy	Easy	Highly	Excellent	Apparent but Low

Replicability – Provenance – Ease of Use – Scalability

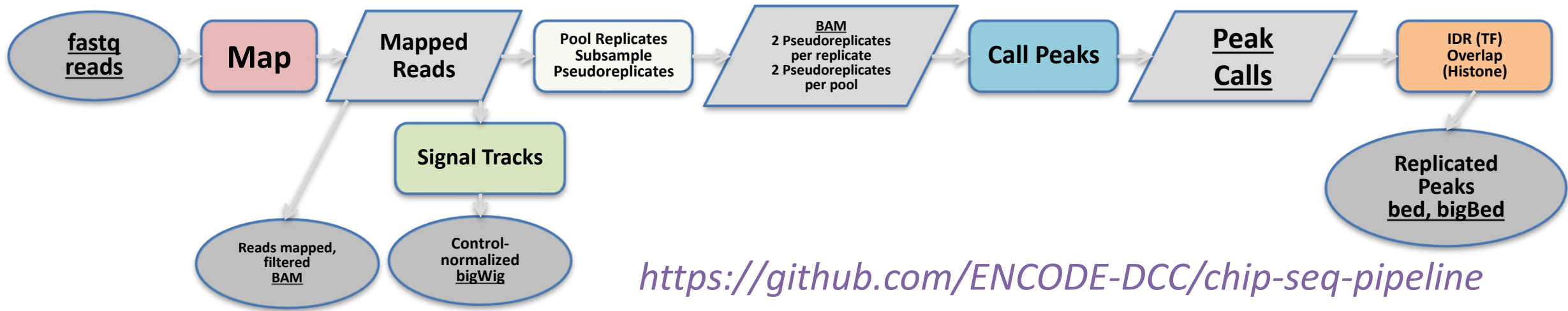
We chose to deploy first to a web/cloud-based platform, DNAnexus

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

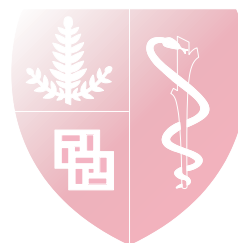
<https://github.com/ENCODE-DCC>



Schema: ENCODE ChIP-seq IDR Pipeline



Target	Key Software	Input Files	Output Files	QA Metrics
TF's	bwa	fastq's (SE or PE) Two biological replicates Matched controls	One bam per replicate	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
	Picard markDuplicates samtools		bigWig fold signal over control	
	MACS2 (Signal tracks)		bigWig p-value signal over control	
	SPP (PeakSeq, GEM future)		bed/bigBed true replicates peaks	
	IDR2		bed/bigBed pooled replicates peaks	
			bed/bigBed IDR thresholded peaks	
Histone Mods	MACS2 for peaks			
	Overlap thresholding			
	IDR2 (future)		bed/bigBed Replicated peaks	



Analysis Pipeline Demonstration and Workshop

Workshop Session 4: ENCODE Uniform Processing Pipelines				
7	DCC tutorial Part II: Pipeline overview	J. Seth Strattan and Ben Hitz	Stanford University	Preparing to Run ENCODE Pipelines

To set up an account:

<https://www.encodeproject.org/tutorials/encode-meeting-2016/>

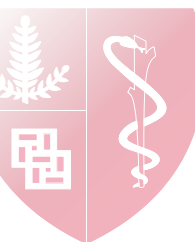
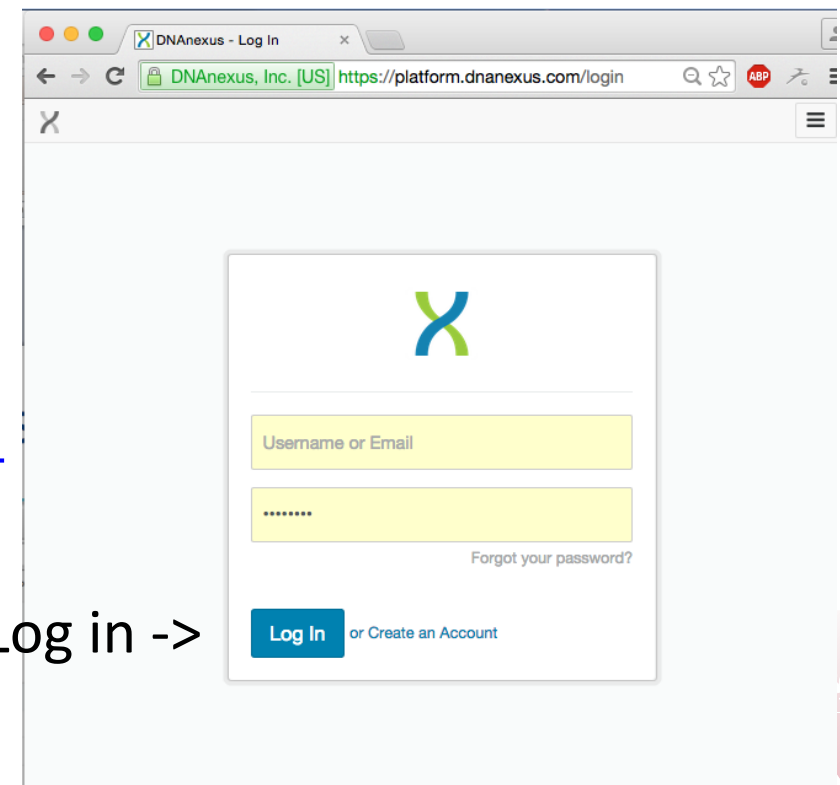
Scroll down to [“Preparing to Run ENCODE Pipelines”](#)

[Exercise 1](#)
[TF Chip-seq](#)

[Exercise 2](#)
[RNA-seq](#)

[Exercise 3](#)
[Histone ChIP-seq](#)

Log in ->



ChIP-seq Results On the ENCODE Portal

Histone ChIP-seq Example

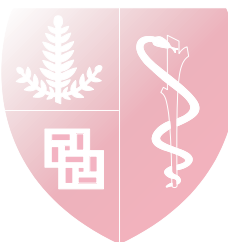
<https://www.encodeproject.org/experiments/ENCSR087PLZ/>

- 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

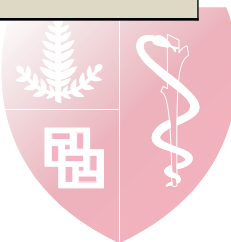
<https://www.encodeproject.org/experiments/ENCSR286PCG/>

- 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 of the pooled replicates.

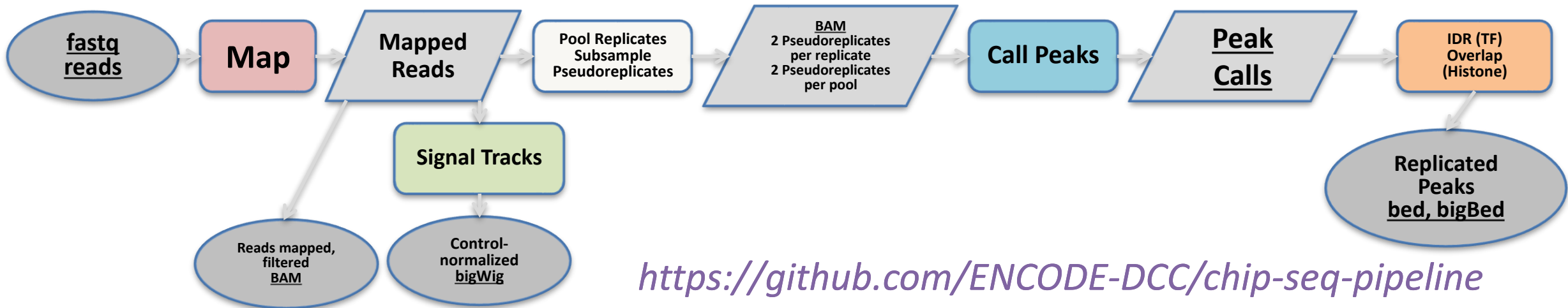


What are the Pipeline Outputs?

Analysis	Basic Files		Special-Purpose Files	
	Content	Stage : Output	Content	Stage : Output
TF ChIP	Optimal IDR Peaks	Final IDR Peak Calls : optimal_set_bb	Conservative IDR Peaks (comparing true replicates only, not pool)	Final IDR Peak Calls : conservative_set_bb
	Pooled control-normalized signal	Final IDR Peak Calls : pooled_signal	Per-replicate signals (visual check of replication)	Final IDR Peak Calls : rep[1,2]_signal
			p-value signals (vs the control)	ENCODE Peaks : rep[1,2]_pvalue_signal
Histone ChIP	Replicated narrowPeaks	Final narrowpeaks : overlapping_peaks	Per-replicate signals (visual check of replication)	Final narrowpeaks : rep[1,2]_signal
			p-value signals (vs the control)	ENCODE Peaks : rep[1,2]_pvalue_signal
	Pooled control-normalized signal	Final narrowpeaks : pooled_signal	gappedPeaks (a set of connected narrowPeaks)	Final gappedpeaks : overlapping_peaks
			broadPeaks (a region of enrichment)	Final broadpeaks : overlapping_peaks
long RNA-seq	plus/minus strand signal from uniquely mapping reads	bam to stranded signal : [plus,minus]_uniq_bw	Isoform quantitation	RSEM quantify genes : rsem_iso_results
	Per-gene quantitation	RSEM quantify genes : rsem_gene_results		



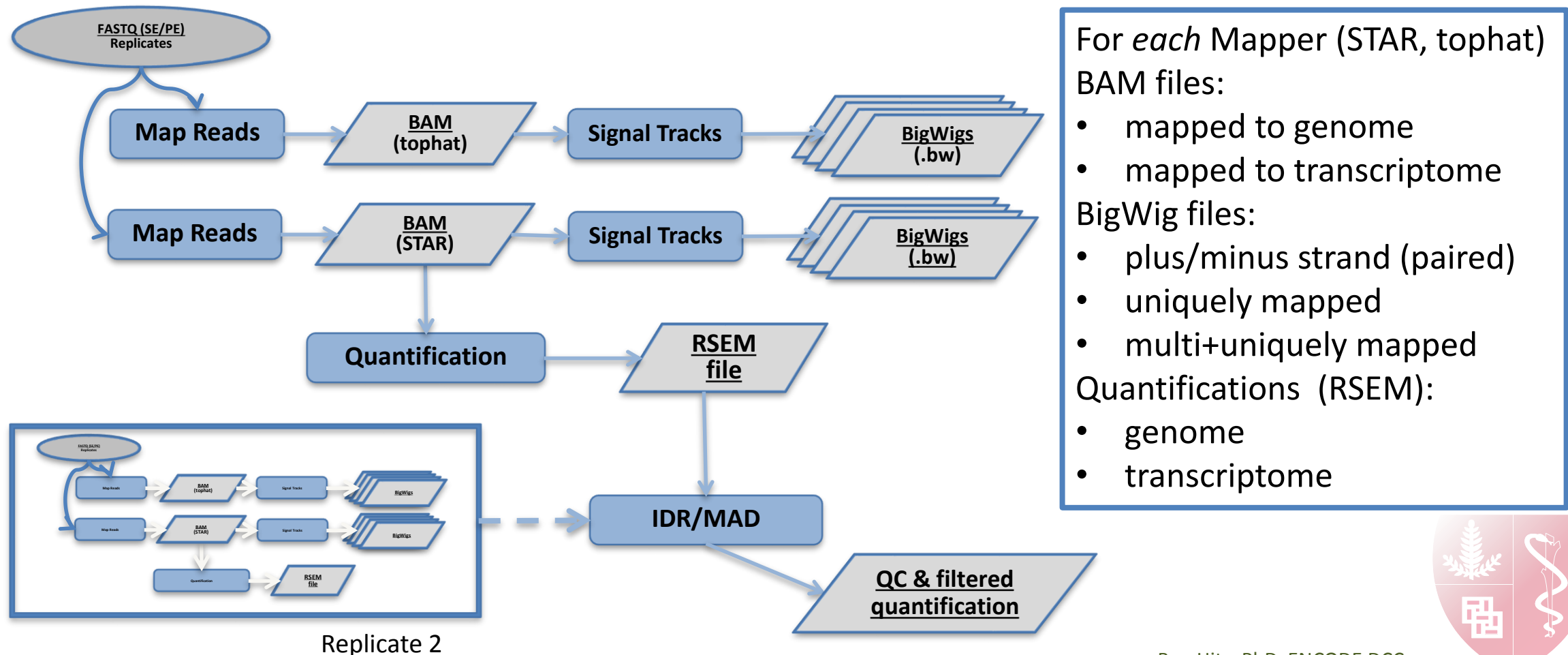
ENCODE ChIP-seq Quality Metrics: Resources



Estimates	Description	References
Depth	Number of mapped reads Number of useable fragments after filtration	Jung YL, et al. Nucleic Acids Research. 2014;42(9):e74
Library Complexity	Non-Redundant Fraction (NRF) PCR Bottleneck Coefficient (PBC1 and PBC2)	Landt S, et al. Genome Res. 2012. 22: 1813-1831
ChIP Quality	Strand Cross-Correlation (NSC and RSC) Cross-Correlation Plot	
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 http://www.personal.psu.edu/users/q/u/qu12/IDR101.pdf

Schema: ENCODE RNA-seq Pipeline

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



For *each* Mapper (STAR, tophat) BAM files:

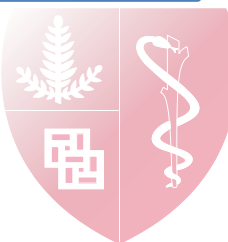
- mapped to genome
- mapped to transcriptome

BigWig files:

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

Quantifications (RSEM):

- genome
- transcriptome

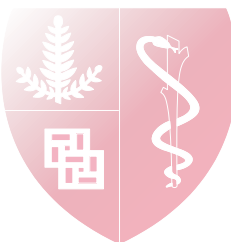


RNA-seq Results On the ENCODE Portal

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



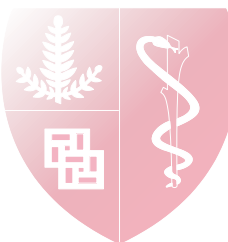
Results from the ChIP-seq exercise

ENCORE2016 Workshop Cherry Lab billed to Admin your access [Share 3 Members](#)

Settings Manage **Monitor** Visualize

Bookmarks SEARCH SCOPE: Root executions only STATE: Any (7 recent jobs) NAME: Any ID: Any CREATED: Any LAUNCHED BY: Any (1) PROPERTIES: Any EXECUTABLE: Any Filters

Status	Name	Executable	Launched by	Started	Duration	Price
Done	ENCSR000AFI RNA	ENCODE RNA-Seq (Long) Pipeline - 1 (pair...	J. Seth Strattan	06/07/2016 3:55 pm	1h 2m	\$1.46
Terminated	ENCSR000AFI RNA 2	ENCODE RNA-Seq (Long) Pipeline - 1 (pair...	J. Seth Strattan	06/07/2016 1:58 pm	31m	\$0.64
Failed	ENCSR000AFI RNA	ENCODE RNA-Seq (Long) Pipeline - 1 (pair...	J. Seth Strattan	06/07/2016 12:32 pm	1h 2m	\$0.88
Done	ENCSR464DKE hCTCF	ENCODE TF ChIP-seq (hg19)	J. Seth Strattan	06/06/2016 3:30 pm	59m	\$1.88
Done	ENCSR087PLZ mH3K9ac	ENCODE Histone ChIP-seq (mm10)	J. Seth Strattan	06/06/2016 2:25 pm	49m	\$1.50
Done	ENCSR286PCG hZBED1	ENCODE TF ChIP-seq (hg19)	J. Seth Strattan	06/06/2016 2:23 pm	58m	\$1.96
Done	ENCSR000EEB hMAFK	ENCODE TF ChIP-seq Unary Control (hg19)	J. Seth Strattan	06/06/2016 2:21 pm	56m	\$1.41



Results from the ChIP-seq exercise

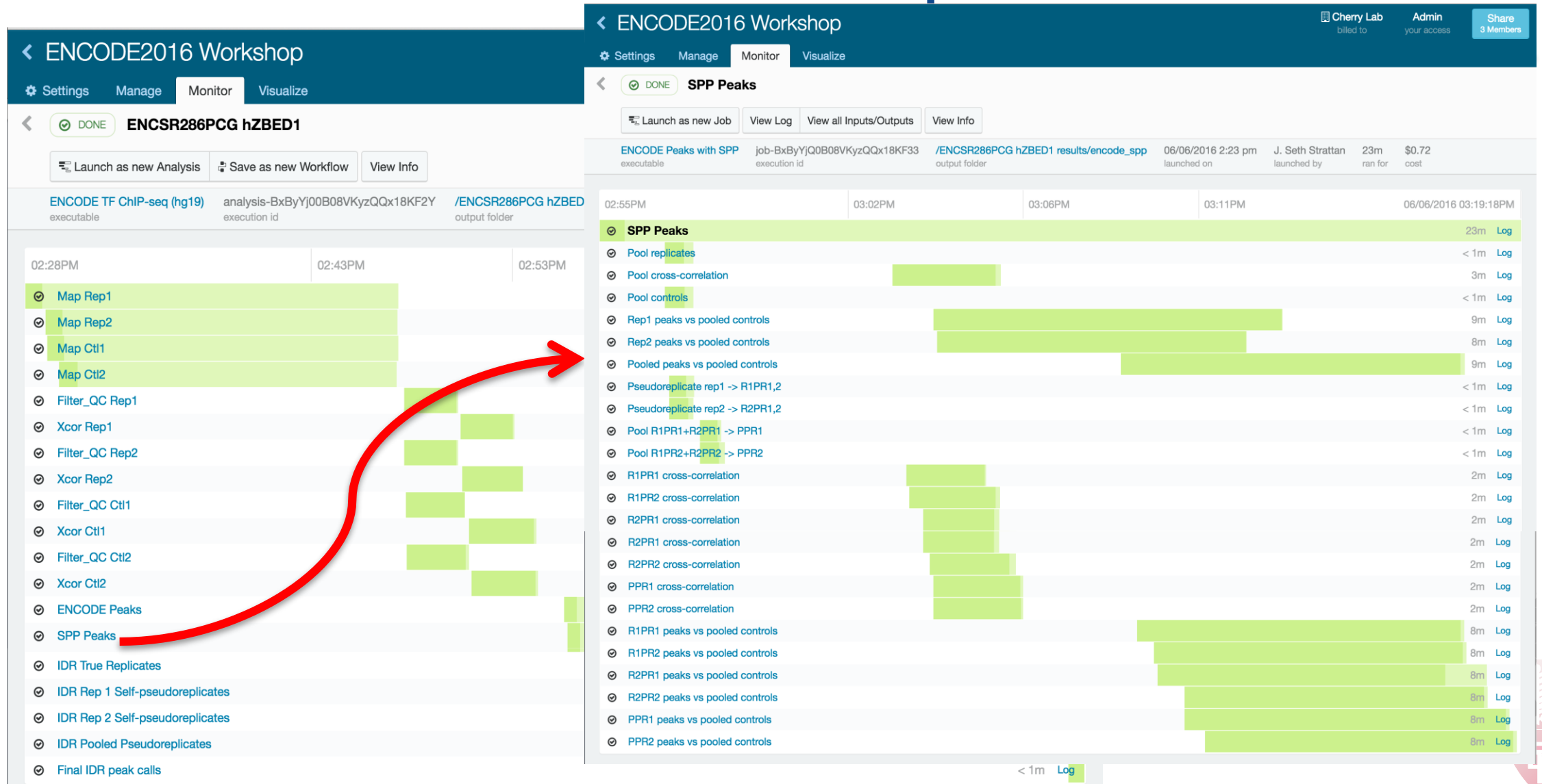
The screenshot displays the ENCODE2016 Workshop interface. On the left, a list of analyses is shown with columns for Status and Name. The analysis 'ENCSR000AFI RNA' is marked as 'Failed', and a red arrow points from this entry to the detailed view on the right.

The detailed view for 'ENCSR286PCG hZBED1' shows a Gantt chart with the following tasks and durations:

Task	Start Time	End Time	Duration
Map Rep1	02:28PM	02:43PM	19m
Map Rep2	02:28PM	02:43PM	18m
Map Ctl1	02:28PM	02:43PM	18m
Map Ctl2	02:28PM	02:43PM	17m
Filter_QC Rep1	02:43PM	02:53PM	3m
Xcor Rep1	02:43PM	02:53PM	3m
Filter_QC Rep2	02:43PM	02:53PM	3m
Xcor Rep2	02:43PM	02:53PM	3m
Filter_QC Ctl1	02:43PM	02:53PM	3m
Xcor Ctl1	02:43PM	02:53PM	3m
Filter_QC Ctl2	02:43PM	02:53PM	3m
Xcor Ctl2	02:43PM	02:53PM	3m
ENCODE Peaks	02:53PM	03:08PM	15m
SPP Peaks	02:53PM	03:26PM	23m
IDR True Replicates	03:08PM	03:09PM	1m
IDR Rep 1 Self-pseudoreplicates	03:08PM	03:09PM	1m
IDR Rep 2 Self-pseudoreplicates	03:08PM	03:09PM	1m
IDR Pooled Pseudoreplicates	03:08PM	03:09PM	1m
Final IDR peak calls	03:08PM	03:09PM	< 1m



Results from the ChIP-seq exercise



Results from the ChIP-seq exercise

Final IDR peak calls

INPUTS

Blacklist (gzipped BED file) (blacklist)
`wgEncodeDacMapabilityConsensusExcludable.bed.gz`

IDR peaks from replicate 2 self-pseudoreplicates (r2pr_peaks)
`R2-ENCFF986vR2-ENCFF986.IDRv2.IDR0.05.narrowPeak.gz`

IDR peaks from replicate 1 self-pseudoreplicates (r1pr_peaks)
`R1-ENCFF016vR1-ENCFF016.IDRv2.IDR0.05.narrowPeak.gz`

Second replicate signal track (rep2_signal)
`r2.fc_signal.bw`

Pooled replicates signal (pooled_signal)
`pool.fc_signal.bw`

IDR peaks from pooled self-pseudoreplicates (pooledpr_peaks)
`R1-ENCFF016vR1-ENCFF016.IDRv2.IDR0.05.narrowPeak.gz`

First replicate signal track (rep1_signal)
`r1.fc_signal.bw`

IDR peaks from true replicates (reps_peaks)
`R1-ENCFF016vR2-ENCFF986.IDRv2.IDR0.05.narrowPeak.gz`

chrom.sizes for bedToBigBed (chrom_sizes)
`male.hg19.chrom.sizes`

.as file for bedToBigBed (as_file)
`narrowPeak.as`

OUTPUTS

Final peak calls - optimal set bigBed (optimal_set_bb)
`IDR_final_optimal.narrowPeak.bb`

Final peak calls - conservative set (conservative_set)
`IDR_final_conservative.narrowPeak.gz`

Number of peaks in the optimal set (No)
169

Number of peaks in the conservative set (Nc)
117

Final peak calls - optimal set (optimal_set)
`IDR_final_optimal.narrowPeak.gz`

Final peak calls - conservative set bigBed (conservative_set_bb)
`IDR_final_conservative.narrowPeak.bb`

Self-consistency ratio (self_consistency_ratio)
1.2127659574468086

Pooled replicates signal (pooled_signal)
`pool.fc_signal.bw`

Second replicate signal track (rep2_signal)
`r2.fc_signal.bw`

Result of the reproducibility test (reproducibility_test)
pass

Rescue ratio (rescue_ratio)
1.4344262295081966

Number of peaks from pooled pseudoreplicates (Np)
175

Number of peaks from replicate 1 self-pseudoreplicates (Nr)
47

First replicate signal track (rep1_signal)
`r1.fc_signal.bw`

Number of peaks from true replicates (Nt)
122

Number of peaks from replicate 2 self-pseudoreplicates (Ns)
57

OUTPUTS

Final peak calls - optimal set bigBed (optimal_set_bb)

`IDR_final_optimal.narrowPeak.bb`

Final peak calls - conservative set (conservative_set)

`IDR_final_conservative.narrowPeak.gz`

Number of peaks in the optimal set (No)

169

Number of peaks in the conservative set (Nc)

117

Final peak calls - optimal set (optimal_set)

`IDR_final_optimal.narrowPeak.gz`

Final peak calls - conservative set bigBed (conservative_set_bb)

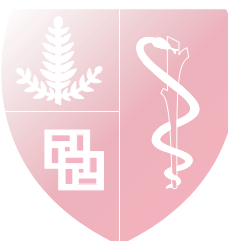
`IDR_final_conservative.narrowPeak.bb`

Self-consistency ratio (self_consistency_ratio)

1.2127659574468086

Pooled replicates signal (pooled_signal)

`pool.fc_signal.bw`



Results from the ChIP-seq exercise

OUTPUTS

Final peak calls - optimal set big

[IDR_final_optimal.narrow](#)

Final peak calls - conservative s

[IDR_final_conservative.n](#)

Number of peaks in the optimal

169

Number of peaks in the conserv

117

Final peak calls - optimal set (op

[IDR_final_optimal.narrow](#)

Final peak calls - conservative s

(conservative_set_bb)

[IDR_final_conservative.n](#)

Self-consistency ratio (self_cons

1.2127659574468086

Pooled replicates signal (pooled

[pool.fc_signal.bw](#)

The screenshot shows the ENCODE2016 Workshop interface. At the top, there is a navigation bar with a back arrow, the title 'ENCODE2016 Workshop', and tabs for 'Settings', 'Manage', 'Monitor', and 'Visualize'. Below this is a toolbar with buttons for 'Add Data', 'New Folder', 'New Workflow', and 'Start Analysis...'. To the right of the toolbar are buttons for 'Copy', 'Delete', 'Info', 'Download', and 'Open in New'. Below the toolbar is a search and filter section with a 'Bookmarks' dropdown and filters for 'SEARCH SCOPE' (Entire project), 'NAME' (Any), 'ID' (Any), 'CLASS' (Any 5), 'MODIFIED' (Any), and 'TAGS' (Any). The main content area shows a file tree on the left under 'ENCODE2016 Workshop' with folders for 'ChIP-seq', 'ENCSR000AFI results', 'ENCSR000EEB hMAFK results', 'ENCSR087PLZ mH3K9ac results', and 'ENCSR286PCG hZBED1 results'. The 'idr2' folder is expanded. On the right, a list of files is shown with checkboxes. The file 'IDR_final_optimal.narrowPeak.bb' is selected, indicated by a blue checkmark and a red arrow pointing from the 'IDR_final_optimal.narrow' link in the 'OUTPUTS' section on the left.

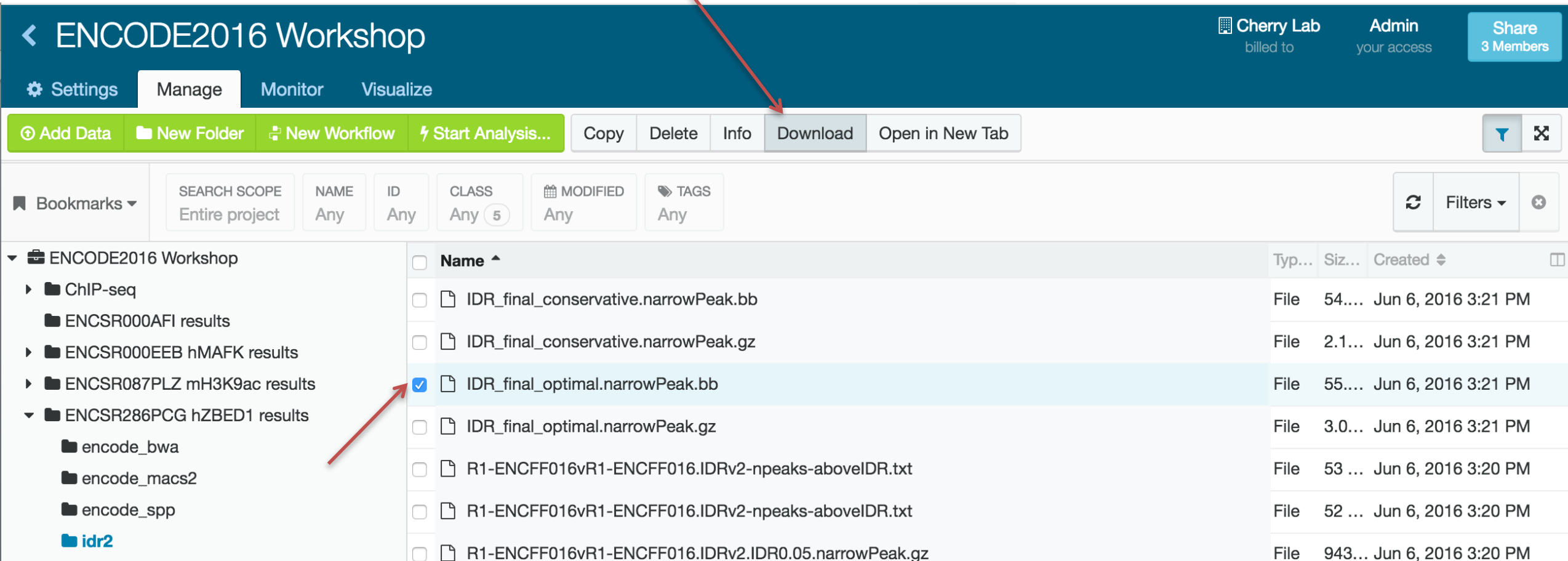
SEARCH SCOPE	NAME	ID	CLASS	MODIFIED	TAGS
Entire project	Any	Any	Any (5)	Any	Any

- ENCODE2016 Workshop
 - ChIP-seq
 - ENCSR000AFI results
 - ENCSR000EEB hMAFK results
 - ENCSR087PLZ mH3K9ac results
 - ENCSR286PCG hZBED1 results
 - encode_bwa
 - encode_mac2
 - encode_spp
 - idr2

Name
<input checked="" type="checkbox"/> IDR_final_optimal.narrowPeak.bb
<input type="checkbox"/> IDR_final_optimal.narrowPeak.gz
<input type="checkbox"/> R1-ENCFF016vR1-ENCFF016.IDRv2-npeaks-aboveIDR.txt
<input type="checkbox"/> R1-ENCFF016vR1-ENCFF016.IDRv2-npeaks-aboveIDR.txt
<input type="checkbox"/> R1-ENCFF016vR1-ENCFF016.IDRv2.IDR0.05.narrowPeak.gz
<input type="checkbox"/> R1-ENCFF016vR1-ENCFF016.IDRv2.IDR0.05.narrowPeak.gz

Results from the ChIP-seq exercise

“Download” to generate temporary URL’s to the selected files




The screenshot shows the ENCODE2016 Workshop interface. The top navigation bar includes 'Settings', 'Manage', 'Monitor', and 'Visualize'. The 'Manage' tab is active, showing a toolbar with 'Add Data', 'New Folder', 'New Workflow', 'Start Analysis...', 'Copy', 'Delete', 'Info', 'Download', and 'Open in New Tab'. The 'Download' button is highlighted with a red arrow. Below the toolbar, there are search filters for 'SEARCH SCOPE', 'NAME', 'ID', 'CLASS', 'MODIFIED', and 'TAGS'. The main content area displays a file list under the 'ENCODE2016 Workshop' folder. The file 'IDR_final_optimal.narrowPeak.bb' is selected, indicated by a blue checkmark in the selection column and a red arrow pointing to it. The file list includes columns for 'Name', 'Type', 'Size', and 'Created'.

Name	Typ...	Siz...	Created
<input type="checkbox"/> IDR_final_conservative.narrowPeak.bb	File	54...	Jun 6, 2016 3:21 PM
<input type="checkbox"/> IDR_final_conservative.narrowPeak.gz	File	2.1...	Jun 6, 2016 3:21 PM
<input checked="" type="checkbox"/> IDR_final_optimal.narrowPeak.bb	File	55...	Jun 6, 2016 3:21 PM
<input type="checkbox"/> IDR_final_optimal.narrowPeak.gz	File	3.0...	Jun 6, 2016 3:21 PM
<input type="checkbox"/> R1-ENCFF016vR1-ENCFF016.IDRv2-npeaks-aboveIDR.txt	File	53 ...	Jun 6, 2016 3:20 PM
<input type="checkbox"/> R1-ENCFF016vR1-ENCFF016.IDRv2-npeaks-aboveIDR.txt	File	52 ...	Jun 6, 2016 3:20 PM
<input type="checkbox"/> R1-ENCFF016vR1-ENCFF016.IDRv2.IDR0.05.narrowPeak.gz	File	943...	Jun 6, 2016 3:20 PM

Visualize on the UCSC Genome Browser

Get Your Data ? ×

⌚ For security reasons, these download links will expire in 1m 2s


Download file

Copy the link below to download from the command line or with a download manager. This link will work for the next 24 hours.

https://dl.dnanex.us/F/D/K2fYG9kyz0k.../IDR_final_optimal.narrowPeak.bb

Warning: Anyone with this link can access this data.

Name
IDR_final_optimal.narrowPeak.bb

[Get URL](#) [Download](#)

- Open Link in New Tab
- Open Link in New Window
- Open Link in Incognito Window
- Save Link As...
- Copy Link Address
- Copy
- Print...

genome.ucsc.edu

UCSC Genome Bioinformatics

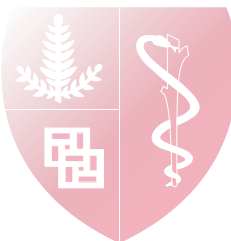
Genomes Genome Browser Tools Mirrors Downloads My Data Help

Genome About the UCSC Genome Bioinformatics Site

- Sessions
- Track Hubs
- Custom Tracks

Paste URLs or data: [Or upload:](#) No file chosen

`https://dl.dnanex.us/F/D/pP8ZGFGX3gfK0PqjxkpYJy8JykjJV203x0Kb0gj7/IDR_final_optimal.narrowPeak.bb`



Visualize on the UCSC Genome Browser

The image shows two screenshots of the UCSC Genome Browser. The top screenshot is the 'Manage Custom Tracks' interface. It features a navigation bar with 'Genomes', 'Genome Browser', 'Tools', 'Mirrors', 'Downloads', 'My Data', 'Help', and 'About Us'. Below the navigation bar, there are dropdown menus for 'genome' (set to 'Human') and 'assembly' (set to 'Feb. 2009 (GRCh37/hg19) [hg19]'). A table lists custom tracks with columns for 'Name', 'Description', 'Type', 'Doc', and 'delete'. One track is listed: 'IDR_final_optimal.narrowPeak' with description 'IDR_final_optimal.narrowPeak' and type 'bigBed'. To the right of the table is a 'view in' dropdown menu set to 'Genome Browser' and a 'go' button. Below the table is an 'add custom tracks' button. A red arrow points from the 'add custom tracks' button to the 'UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly' page shown in the bottom screenshot.

The bottom screenshot shows the main UCSC Genome Browser interface. The navigation bar includes 'Genomes', 'Genome Browser', 'Tools', 'Mirrors', 'Downloads', 'My Data', 'View', 'Help', and 'About Us'. The main heading is 'UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly'. Below the heading are navigation controls: 'move' (with arrows), 'zoom in' (1.5x, 3x, 10x, base), and 'zoom out' (1.5x, 3x, 10x, 100x). A search bar contains 'chr21:33,001,785-33,064,574 62,790 bp.' and a 'go' button. A chromosome ideogram shows chromosome 21 with a red vertical line indicating the current position. Below the ideogram is a detailed genomic track view for 'chr21 (q22.11)'. The track includes a scale bar from 33,010,000 to 33,060,000 bp. The main track shows 'Basic Gene Annotation Set from GENCODE Version 19' with genes 'SOD1' and 'SNORA81'. Other tracks include 'IDR_final_optimal.narrowPeak', 'Microsatellites - Di-nucleotide and Tri-nucleotide Repeats', 'Duplications of >1000 Bases of Non-RepeatMasked Sequence', 'Simple Nucleotide Polymorphisms (dbSNP 146) Found in >= 1% of Samples', and 'Repeating Elements by RepeatMasker'. At the bottom, there are 'move start' and 'move end' buttons, a '2.0' zoom level, and a 'Click on a feature for details...' instruction.

Pipeline Workshop Summary

DCC Goals:

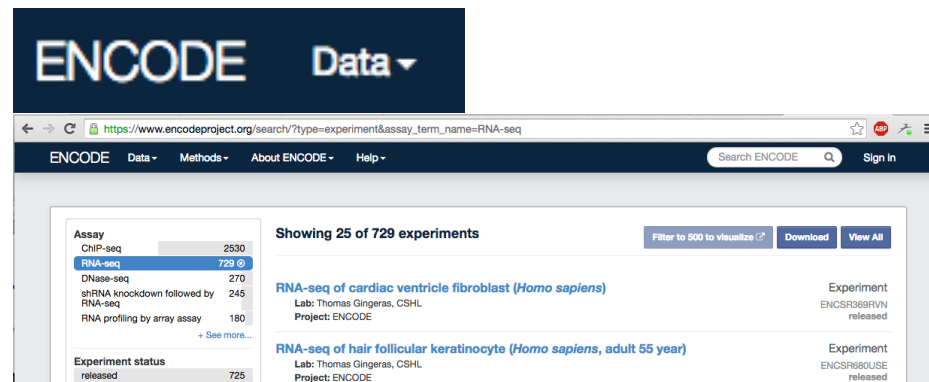
1. Deploy ENCODE-defined pipelines for ChIP-seq, RNA-seq, DNase-seq, methylation.
2. Use those pipelines to generate the standard ENCODE peaks, quantitations, CpG.
3. Capture metadata to make clear what software, versions, parameters, inputs were used.
4. Capture, accession, and distribute the output.
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

DNAneXus

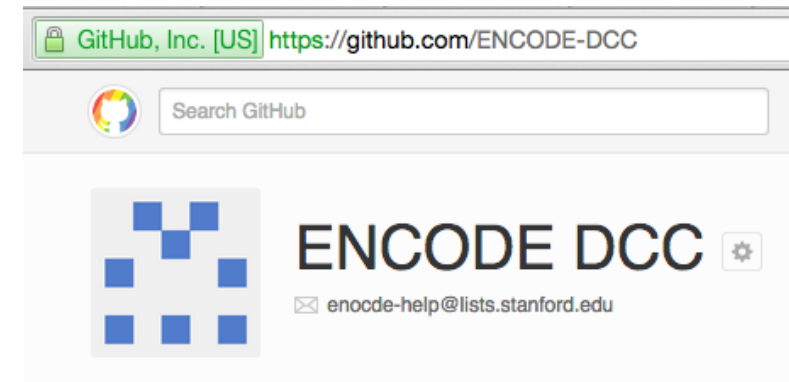
Featured projects

- ENCODE Uniform Processing Pipelines
- Parliament



The screenshot shows the ENCODE Data search results page. The URL is https://www.encodeproject.org/search/?type=experiment&assay_term_name=RNA-seq. The page displays a list of experiments, with 25 of 729 shown. The top experiment is "RNA-seq of cardiac ventricle fibroblast (Homo sapiens)" by Thomas Gingeras, CSHL. The page includes a search bar, navigation tabs, and a table of results.

Assay	Count
ChIP-seq	2530
RNA-seq	729
DNase-seq	270
shRNA knockdown followed by RNA-seq	245
RNA profiling by array assay	180
Experiment status released	725



The screenshot shows the GitHub repository page for ENCODE DCC. The URL is <https://github.com/ENCODE-DCC>. The page features the GitHub logo, a search bar, and the repository name "ENCODE DCC" with a settings icon. The contact email is encode-help@lists.stanford.edu.

Contributors

ENCODE Data Coordinating Center

Mike Cherry, PI, Stanford
Ben Hitz, Project Manager

Pipeline Developers

Ben Hitz, WGBS, Software Lead
Tim Dreszer, RNA-seq, DNase-seq
J. Seth Strattan, CHIP-seq

Portal Developers

Ulugbek Baymuradov
Laurence Rowe
Forrest Tanaka
Chris Thomas

Data Wranglers

Esther Chan
Jean Davidson
Idan Gabdank
Jason Hilton
Cricket Sloan

QA & Biocuration Assistance

Marcus Ho
Aditi Narayanan
Kathrina Oate

Support Staff

Stuart Miyasato
Matt Simison



@encodedcc



encode-help@lists.stanford.edu

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 Irizarry, Harvard
Xintao Wei, UConn
Brent Gravely, UConn
Colin Dewey, U Wisconsin
Roderic Guigó, CRG
Sarah Djebali, CRG

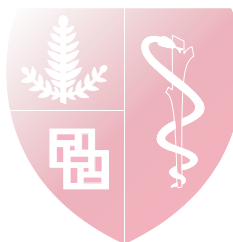
ChIP-seq

Anshul Kundaje, Stanford
Nathan Boley, Stanford
Jin Lee, Stanford

DNAnexus

Mike Lin
Andey Kislyuk
Singer Ma
Brett Hannigan
Ohad Rodeh
Joe Dale
George Asimenos
Alpha Diallo
Chai Fungtammasan
Naina Thangaraj

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



Schema: ENCODE WGBS Pipeline

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

