

# ENCODE CHIP PIPELINES

---

J. Seth Strattan, ENCODE DCC  
ENCODE Binding Working Group  
November, 2015









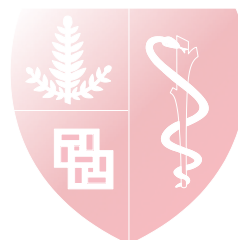
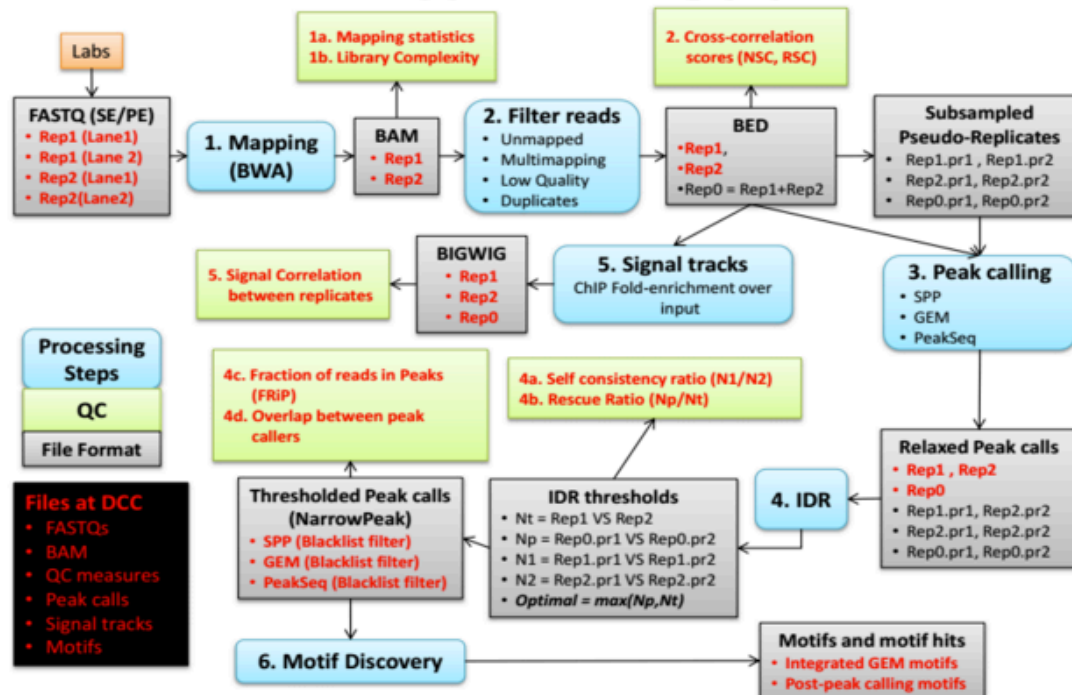
# ENCODE DAC Specifies, DCC Implements

## ENCODE3 pipeline v1 specifications

### Pipeline Overview

[https://docs.google.com/document/d/1IG\\_Rd7fnYgRpSlqrifuVIAz2dW1VaSQThzk836Db99c/edit?pli=1#heading=h.9ecc41kilcvq](https://docs.google.com/document/d/1IG_Rd7fnYgRpSlqrifuVIAz2dW1VaSQThzk836Db99c/edit?pli=1#heading=h.9ecc41kilcvq)

## TF ChIP-seq processing pipeline



# Deployment Platform Considerations



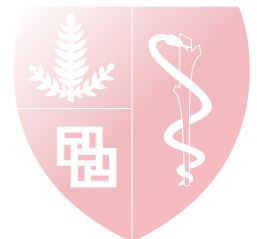
	Develop	Share	Run	Elastic	Provenance	Cost
<b>HPC Cluster (Scripts)</b>	Hard	Hard	Hard	Cluster-Dependent	Moderate	Obscure/Subsidized
<b>HPC Container</b>	Hard	Moderate	Moderate	Cluster-Dependent	Good	Obscure/Subsidized
<b>Web/Cloud</b>	Moderate	Easy	Easy	Highly	Excellent	Apparent but Low

Replicability – Provenance – Ease of Use – Scalability

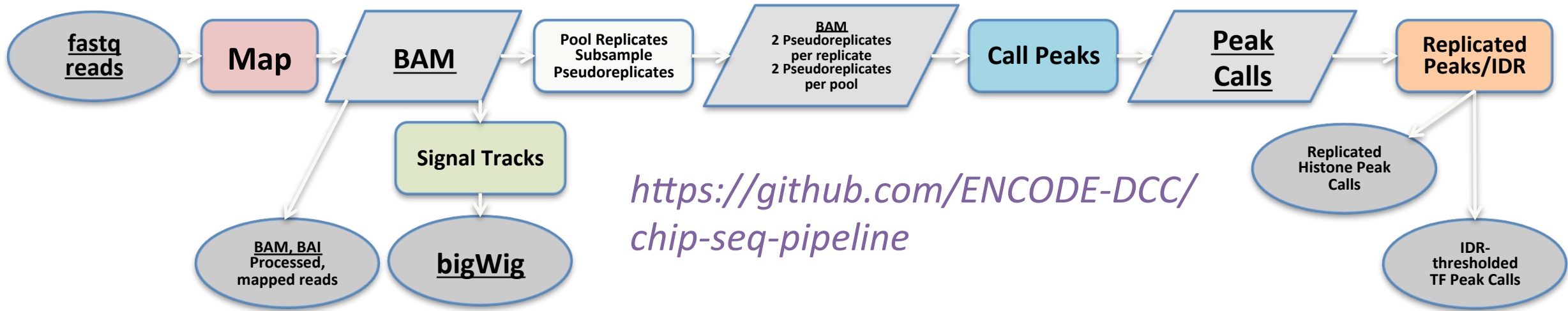
We chose to deploy first to a web/cloud-based platform, DNAnexus

Code is open source now and will be deployed to additional platforms

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



# Schema: ENCODE ChIP-seq Pipeline

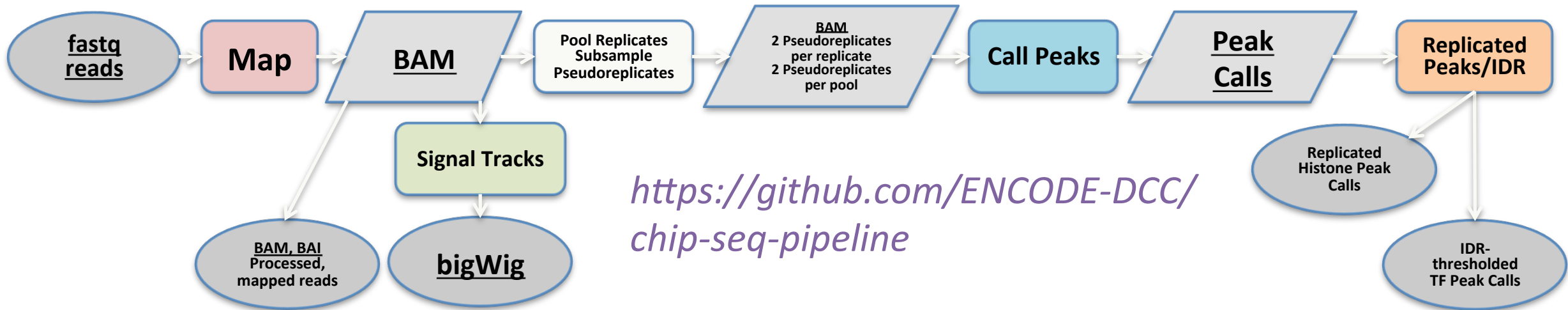


Target	Key Software	Input Files	Output Files	QA Metrics
TF's	bwa	fastq's (SE or PE) Two biological replicates Matched, replicated controls	One bam per replicate	NRF (Non-redundant fraction) PBC1 and 2 (PCR bottleneck coefficients) Number of distinct uniquely-mapping reads NSC/RSC (Strand cross-correlation)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





# ENCODE ChIP-seq Quality Metrics: Resources



Estimates	Description	References
<b>Depth</b>	Number of mappable reads Useable fragments (=SE filtered reads, PE/2)	Jung YL, et al. Nucleic Acids Research. 2014;42(9):e74
<b>Library Complexity</b>	Non-Redundant Fraction PCR Bottleneck Coefficient	Landt S, et al. Genome Res. 2012. 22: 1813-1831
<b>ChIP Quality</b>	Normalized Strand Cross-Correlation Relative Strand Cross-Correlation	
<b>Replicate Concordance</b>	IDR Rescue Ratio IDR Self-Consistency Ratio IDR Reproducibility Test	Li Q, et al. Annals Applied Statistics. 2011, Vol. 5, No. 3, 1752-1779



# Transcription Factor ChIP-seq Pipeline

## Pipeline Overview

The pipeline takes as inputs both ChIP-seq reads (from paired-end stranded or single-end unstranded libraries) and a set of reference fasta's, and outputs several products:

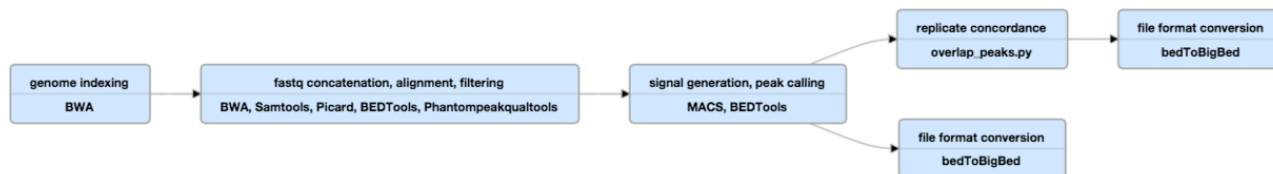
- Mapping to the appropriate genome assembly and subsequent filtration and deduplication creates an alignment file in bam file format.
- Signal tracks depicting control-normalized tag density in the bigWig file format are created for each replicate, and for both replicates' reads pooled together. The signal is expressed in two ways: as fold-over control at each position, and as a p-value to reject the null hypothesis that the signal at that location is present in the control.
- Relaxed peak calls in bed and bigBed format for each replicate and for both replicates' reads pooled together. These peak calls are thresholded in such a way as to sample enough noise in the experiment for efficient statistical comparison of replicates in subsequent steps. As such, many false positives are expected to be present in these peak sets. They are not meant to be interpreted as definitive binding events, but are rather intended to be used as input for subsequent statistical comparison of replicates.
- Final peak calls in the bed and bigBed format are the set of peak calls that pass IDR at a threshold of 2%. The conservative set are peaks derived from IDR analysis of biological replicates, whereas the optimal set are the largest set of peaks derived from IDR analysis of biological replicates and pseudoreplicates. Pseudoreplicates are peak sets called on half of the pooled reads, chosen at random without replacement.

When multiple fastqs are generated from a single biological replicate (in multiple sequencing runs, for example), they are concatenated before mapping.

To view further details of the pipeline: [View Pipeline](#)

To view experiments analyzed by this pipeline: [View Experiments](#)

## Pipeline Schematic



Pipeline overview pages on the ENCODE Portal:

<https://www.encodeproject.org/chip-seq>

[https://www.encodeproject.org/chip-seq/transcription\\_factor/](https://www.encodeproject.org/chip-seq/transcription_factor/)

<https://www.encodeproject.org/chip-seq/histone/>

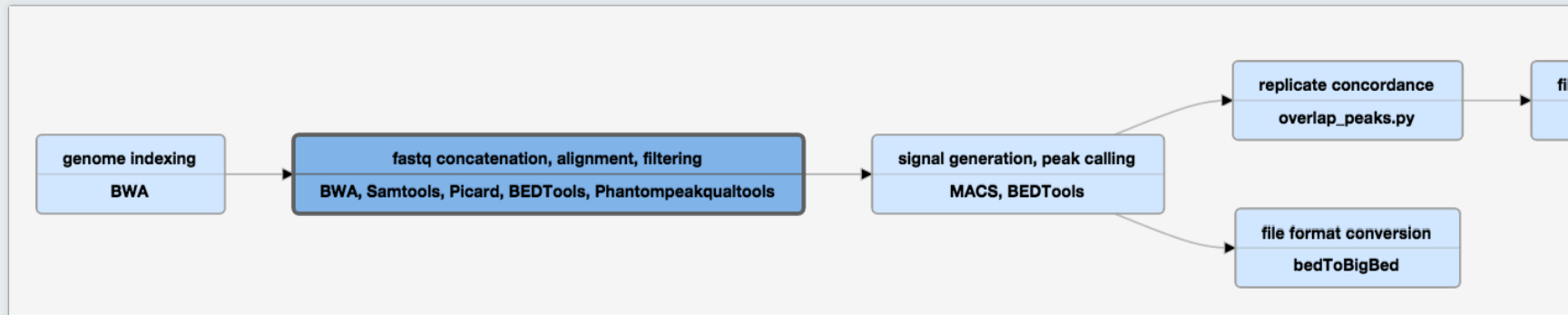


# Histone CHIP-seq

Status: active

**Title:** Histone CHIP-seq  
**Assay:** CHIP-seq  
**Description:** ENCODE defined analysis pipeline for histone CHIP-seq  
**Lab:** ENCODE Processing Pipeline  
**Award PI:** J. Michael Cherry, Stanford

## Pipeline schematic



Download Graph

**Name:** BWA alignment and filtration step - Version 1  
**Step type:** fastq concatenation, alignment, filtering  
**Step aliases:** encode:bwa-alignment-step-v-1  
**Input:** genome index, reads  
**Output:** alignments

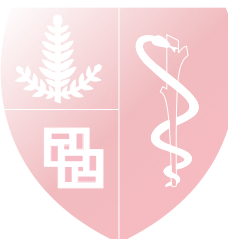
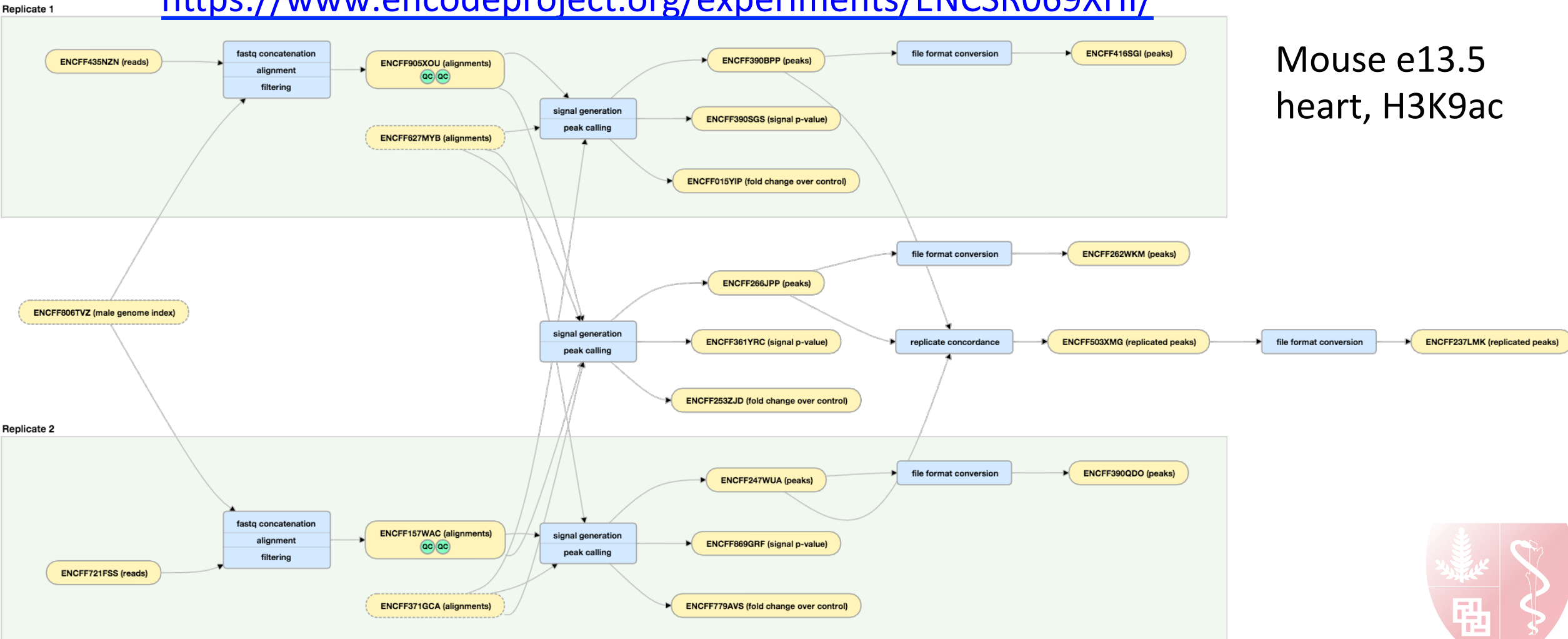
**Version 1 — software:** BWA 0.7.10 Samtools 1.0 Picard 1.92 BEDTools 2.17.0 Phantompeakqualtools 1.1



# Uniformly Processed Data On the ENCODE Portal

<https://www.encodeproject.org/experiments/ENCSR069XHI/>

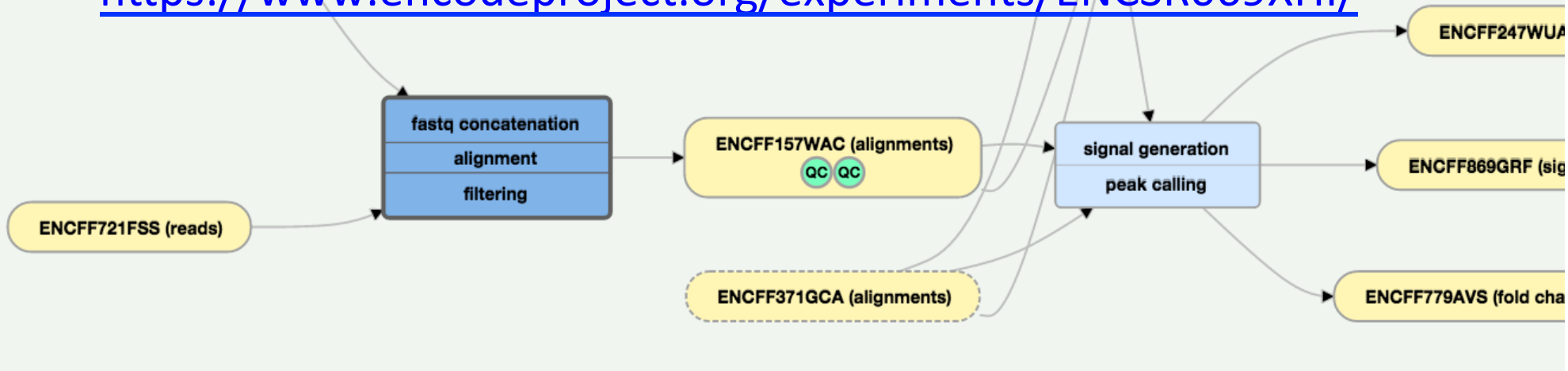
Mouse e13.5  
heart, H3K9ac



# Uniformly Processed Data On the ENCODE Portal

Replicate 2

<https://www.encodeproject.org/experiments/ENCSR069XHI/>



Download Graph

**Name:** BWA alignment and filtration step - Version 1 — Version 1

**Step type:** fastq concatenation, alignment, filtering

**Step aliases:** encode:bwa-alignment-step-v-1

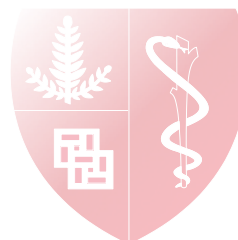
**Input:** genome index, reads

**Output:** alignments

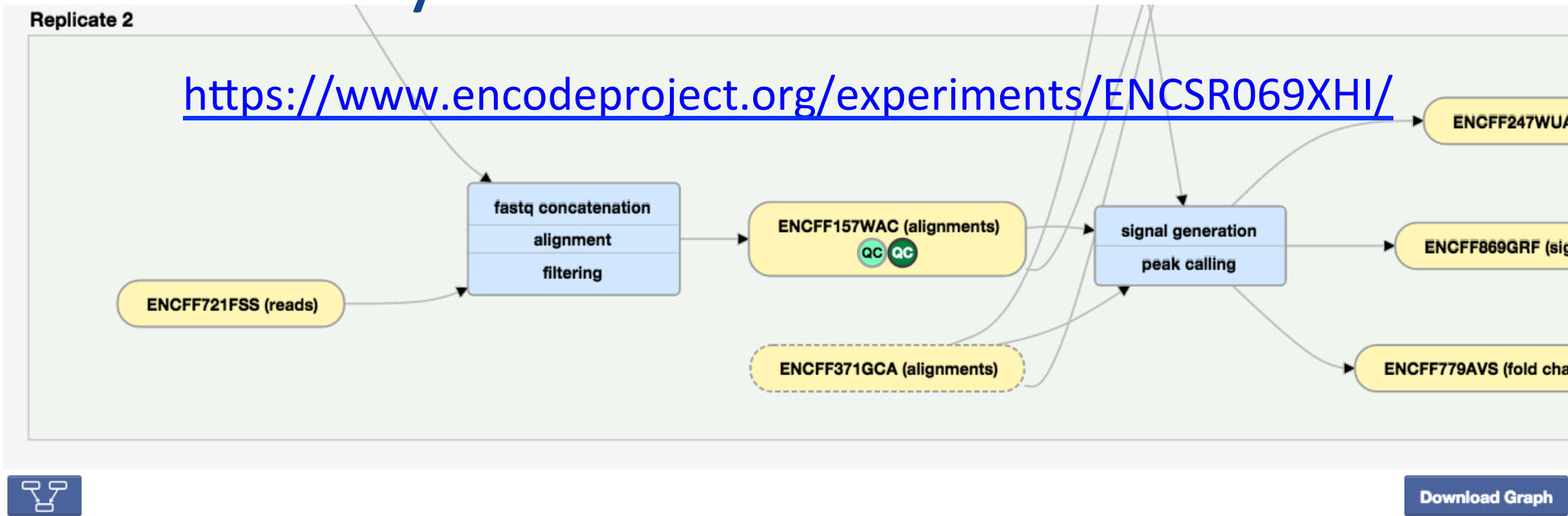
**Pipeline:** [Histone ChIP-seq](#),  
[Transcription factor ChIP-seq](#)

**Software:** BWA 0.7.10 Samtools 1.0 Picard 1.92 BEDTools 2.17.0 Phantompeakqualtools 1.1

## Analysis step metadata



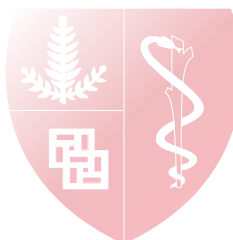
# Uniformly Processed Data On the ENCODE Portal



## Quality metrics of ENCFF157WAC

duplicates:	0
duplicates_qc_failed:	0
mapped:	53623474
mapped_pct:	100.00%
mapped_qc_failed:	0
total:	53623474
total <sup>13</sup> _qc_failed:	0

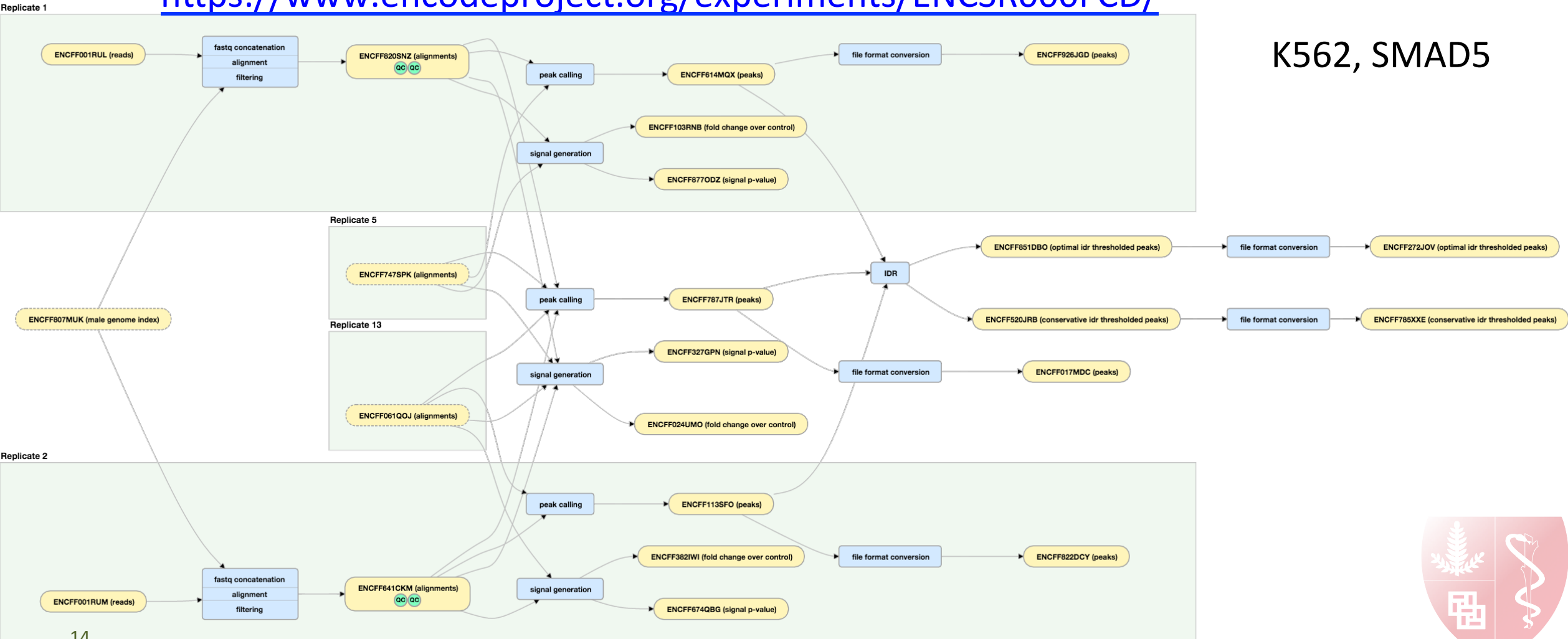
Analysis step metadata  
QC metrics associated with files



# Uniformly Processed Data On the ENCODE Portal

<https://www.encodeproject.org/experiments/ENCSR000FCD/>

K562, SMAD5



# ENCODE ChIP-seq on the Cloud

**E3 ChIP-seq** Cherry Lab billed to Admin your access Share 6 Members

Manage Monitor Visualize

Map ENCSR000FCD rep1 to hg19 and filter: haib\_102715 ... 4 apps configured Workflow Actions Readme Autosaved Start Analysis...

Inputs	App	Outputs
	<b>Gather Inputs ENCSR000...</b> runnable	<input type="radio"/> Output as JSON reads1 reads2
reads1 reads2 E.g. male or female hg19; m...	<b>Map ENCSR000FCD rep1 ...</b> runnable	Mapped reads Mapping statistics <input type="radio"/> paired_end <input type="radio"/> Output as JSON
via <b>Map ENCSR000F...</b> Mapped reads	<b>Filter and QC ENCSR000F...</b> runnable	Mapped reads surviving th... Indexed reads surviving th... Post-filtering mapping stati... Duplication metrics from M...
via <b>Filter and QC EN...</b> Mapped reads surviving th...	<b>Calculate cross-correlatio...</b> runnable	Final tagAlign file. BEDPE file (read pairs on e... Tab-delimited NSC/RSC re... Cross-correlation plot

Workflows map inputs to outputs





# ENCODE ChIP-seq on the Cloud

**E3 ChIP-seq** Cherry Lab billed to Admin your access Share 6 Members

Manage Monitor Visualize

**ENC SR000FCD Peaks haib\_102915** DONE

Launch as new Analysis Save as new Workflow View Info

tf\_chip\_seq executable analysis-BjkVjf00bpZg1yX0b4KFkXPK / output folder 10/30/2015 12:29 am launched on J. Seth Strattan launched by 5h 59m ran for \$28.25 cost

12:34AM 01:54AM 02:51AM 03:48AM 04:44AM 10/30/2015 06:28:49AM

Rep1 cross-correlation	16m	Log
Rep2 cross-correlation	16m	Log
SPP Peaks	5h 33m	Log
ENCODE Peaks	3h 40m	Log
IDR True Replicates	4m	Log
IDR Rep 1 Self-pseudoreplicates	3m	Log
IDR Rep 2 Self-pseudoreplicates	5m	Log
IDR Pooled Pseudoreplicates	4m	Log
Final IDR peak calls	< 1m	Log

Early stages feed subsequent stages



# ENCODE ChIP-seq on the Cloud

06:28:01AM	06:28:10AM	06:28:16AM	06:28:23AM	06:28:29AM	10/30/2015 06:28:41AM
<b>Final IDR peak calls</b>					< 1m Log
<b>INPUTS</b>			<b>OUTPUTS</b>		
IDR peaks from replicate 2 self-pseudoreplicates (r2pr_peaks) <a href="#">ENCF001RUMvENCF001RUM.IDRv2.IDR0.02.narrowPeak.gz</a>			Final peak calls - conservative set (conservative_set) <a href="#">IDR_final_conservative.narrowPeak.gz</a>		
IDR peaks from replicate 1 self-pseudoreplicates (r1pr_peaks) <a href="#">ENCF001RULvENCF001RUL.IDRv2.IDR0.02.narrowPeak.gz</a>			Final peak calls - optimal set (optimal_set) <a href="#">IDR_final_optimal.narrowPeak.gz</a>		
IDR peaks from pooled self-pseudoreplicates (pooledpr_peaks) <a href="#">ENCF001RULvENCF001RUL.IDRv2.IDR0.02.narrowPeak.gz</a>			Final peak calls - conservative set bigBed (conservative_set_bb) <a href="#">IDR_final_conservative.narrowPeak.bb</a>		
chrom.sizes for bedToBigBed (chrom_sizes) <a href="#">ENCODE - Production runs ▶ male.hg19.chrom.sizes</a>			Final peak calls - optimal set bigBed (optimal_set_bb) <a href="#">IDR_final_optimal.narrowPeak.bb</a>		
Blacklist (gzipped BED file) (blacklist) <a href="#">ENCODE Reference Files ▶ wgEncodeDacMapabilityConsensusExcludable.bed.gz</a>			Result of the reproducibility test (reproducibility_test) <b>pass</b>		
IDR peaks from true replicates (reps_peaks) <a href="#">ENCF001RULvENCF001RUM.IDRv2.IDR0.02.narrowPeak.gz</a>			Rescue ratio (rescue_ratio) <b>1.4036567105366853</b>		
.as file for bedToBigBed (as_file) <a href="#">ENCODE Reference Files ▶ narrowPeak.as</a>			Number of peaks from pooled pseudoreplicates (Np) <b>17811</b>		
			Number of peaks from replicate 1 self-pseudoreplicates (N1) <b>9167</b>		
			Number of peaks from replicate 2 self-pseudoreplicates (N2) <b>17815</b>		
			Number of peaks from true replicates (Nt) <b>12689</b>		
			Self-consistency ratio (self_consistency_ratio) <b>1.9433838769499292</b>		

Outputs that are accessioned at the ENCODE Portal



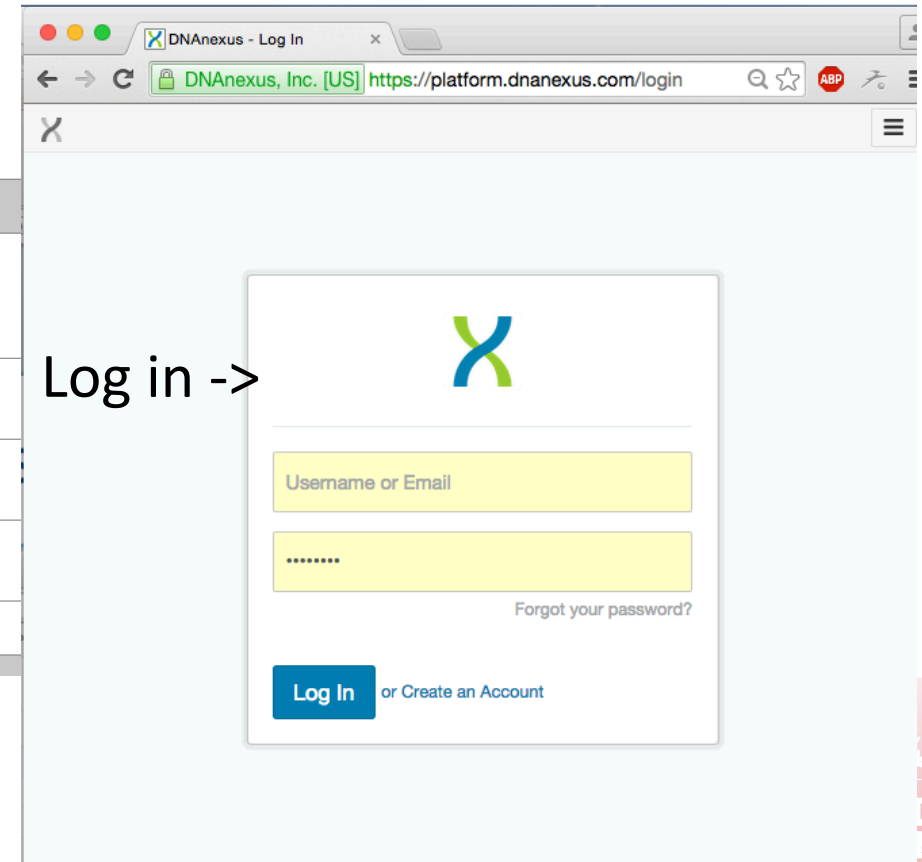
# Running Pipelines on DNAnexus

To set up an account:

<https://www.encodeproject.org/tutorials/encode-users-meeting-2015/>

Click “Prepare to run web-based pipelines”

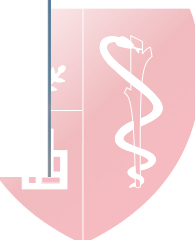
Workshop Session 3: ENCODE Uniform Processing Pipelines				
18	<a href="#">ENCODE Uniform Data Processing Pipelines: Introduction</a>	J. Seth Strattan	Stanford University	<a href="#">Slides</a>   <a href="#">Prepare to run web-based pipelines</a>
19	<a href="#">ENCODE Uniform Data Processing Pipelines: Demo and Hands-on Tutorial</a>	Ben Hitz	Stanford University	<a href="#">Hands-on Tutorial</a>
20	<a href="#">ENCODE Uniform Data Processing Pipelines: ChIP-seq IDR Architecture</a>	J. Seth Strattan	Stanford University	
21	<a href="#">ENCODE Uniform Data Processing Pipelines: Demo and Hands-on Tutorial (continued)</a>	Ben Hitz	Stanford University	
22	<a href="#">ENCODE Uniform Data Processing Pipelines: Wrap-up</a>	J. Seth Strattan	Stanford University	



# Publicly Accessible Pipelines

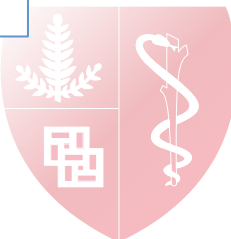
The screenshot shows the DNAnexus web interface. At the top, the browser address bar displays 'https://platform.dnanexus.com/projects/featured'. The navigation bar includes a search box, 'Projects', 'Apps', 'Help', a user profile for 'J. Seth Strattan', and a balance of '\$8,141 remaining'. Below the navigation bar, there are tabs for 'All Projects', 'Resources', and 'Featured'. The main content area is titled 'Featured Projects' and displays seven project cards in a grid. Each card features a briefcase icon, the project name, its size, and a short description.

Project Name	Size	Description
BioNano Genomics	0.06 GB	Tools for de novo assembly, structural variation discovery, and hybrid scaffolding of NGS assemblies using long-range genomic information from BioNano physical genome maps.
ENCODE Uniform Processing Pipelines	349.46 GB	The Encyclopedia of DNA Elements (ENCODE) Consortium catalogs functional elements in the human genome. ENCODE's Data Coordinating Center, provided by the Cherry lab at Stanford University, runs the data analysis on DNAnexus, allowing for uniform processing of data. These pipelines have been defined, in collaboration with the Data Analysis Centers, and deployed by the...
Parliament	341.12 GB	Parliament is a consensus structural variant calling infrastructure that merges multiple data types and structural variant detection methods. The pipeline uses various structural variant detectors, including Breakdancer, CNVnator, Delly, and Pindel to make putative structural variant calls, and then assembles across the predicted breakpoints. Parliament's framework is flexible to take in...
Broad Inst Viral NGS	0.11 GB	Sabeti Lab pipeline for assembly and analysis of Ebola, Lassa, and other viral genomes from NGS reads.
PacBio FALCON Assembler	7,923.51 GB	
HiSeq X-Ten Data	421.93 GB	
HGSC_Mercury	11.99 GB	



# Copy Pipeline to Your Project

The screenshot shows the DNAnexus web interface. The browser address bar displays the URL: <https://platform.dnanexus.com/projects/BKpvFg00VBPV975PgJ6Q03v6/data/>. The page title is "ENCODE Uniform Processing Pipelines". The navigation bar includes "Manage", "Monitor", and "Visualize" tabs. Below the navigation bar, there are buttons for "Add Data", "New Folder", "New Workflow", "Start Analysis...", "Copy", and "Delete". The "Copy" button is highlighted with a red arrow. A dialog box titled "Copy Data to Another Project" is open, showing a text input field for "Enter a project name" and a "Copy to this folder" button. The "Copy" button in the dialog is also highlighted with a red arrow. The main content area shows a list of folders under "ENCODE Uniform Processing Pipelines": ".Applet\_archive", "histone-chip", "long-RNA-seq", "Reference Files", and "WG Bisulfite (Methylation)". The "histone-chip" folder is selected.



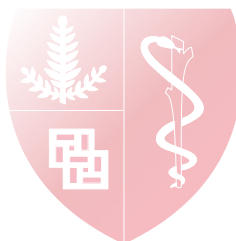
# Select a Workflow to Run

Start Analysis... Copy Delete Info Edit Run Analysis...

MODIFIED Any TAGS Any 1

<input type="checkbox"/>	Name ^	...	Si
<input type="checkbox"/>	test_data	F...	
<input type="checkbox"/>	encode_bwa	A... -	
<input type="checkbox"/>	encode_mac2	A... -	
<input type="checkbox"/>	filter_qc	A... -	
<input checked="" type="checkbox"/>	Histone ChIP-Seq	W... -	
<input type="checkbox"/>	macs2	A... -	
<input type="checkbox"/>	overlap_peaks	A... -	
<input type="checkbox"/>	pool	A... -	
<input type="checkbox"/>	pseudoreplicator	A... -	
<input type="checkbox"/>	xcor	A... -	
<input type="checkbox"/>	xcor_only	A... -	

Select the workflow  
"Run Analysis ..."



# A Workflow

Run "histone\_chip\_seq" as Analysis

View job progress in your project's [Monitor](#) tab. Modifications to an existing workflow will not be saved.

8 apps unconfigured | 8 apps configured | Workflow Actions | Set output folder... | Readme | Run as Analysis... | Refresh

Inputs	App	Outputs
reads1, reads2, E.g. male or female hg19; m...	Map Rep1 (applet) set inputs	Mapped reads, Mapping statistics, paired_end, Output as JSON
reads1, reads2, E.g. male or female hg19; m...	Map Rep2 (applet) set inputs	Mapped reads, Mapping statistics, paired_end, Output as JSON
reads1, reads2, E.g. male or female hg19; m...	Map Ctl1 (applet) set inputs	Mapped reads, Mapping statistics, paired_end, Output as JSON
reads1, reads2, E.g. male or female hg19; m...	Map Ctl2 (applet) set inputs	Mapped reads, Mapping statistics, paired_end, Output as JSON
via Map Rep1 Mapped reads	Filter_QC Rep1 (applet) runnable	Mapped reads surviving th..., Indexed reads surviving th..., Post-filtering mapping stati..., Duplication metrics from M...
via Filter_QC Rep1 Mapped reads surviving th...	Xcor Rep1 (applet) runnable	Final tagAlign file., BEDPE file (read pairs on e..., Tab-delimited NSC/RSC re..., Cross-correlation plot

Once input requirements are met ... "Run as Analysis"

INPUTS





# The “Monitor” Tab

**< E3 ChIP-seq** Cherry Lab billed to Admin your access Share 6 Members

Manage **Monitor** Visualize

**< DONE ENCSR000FCD Peaks haib\_102915**

Launch as new Analysis Save as new Workflow View Info

tf_chip_seq executable	analysis-BjkVjf00bpZg1yX0b4KfKXPK execution id	/ output folder	10/30/2015 12:29 am launched on	J. Seth Strattan launched by	5h 59m ran for	\$28.25 cost
------------------------	------------------------------------------------	-----------------	---------------------------------	------------------------------	----------------	--------------

12:34AM	01:54AM	02:51AM	03:48AM	04:44AM	10/30/2015 06:28:49AM
☑ Rep1 cross-correlation					16m Log
☑ Rep2 cross-correlation					16m Log
☑ SPP Peaks					5h 33m Log
☑ ENCODE Peaks					3h 40m Log
☑ IDR True Replicates					4m Log
☑ IDR Rep 1 Self-pseudoreplicates					3m Log
☑ IDR Rep 2 Self-pseudoreplicates					5m Log
☑ IDR Pooled Pseudoreplicates					4m Log
☑ Final IDR peak calls					< 1m Log



# Gather Results

06:28:01AM	06:28:10AM	06:28:16AM	06:28:23AM	06:28:29AM	10/30/2015 06:28:41AM
Final IDR peak calls					< 1m Log

## INPUTS

IDR peaks from replicate 2 self-pseudoreplicates (r2pr\_peaks)  
[ENCFF001RUMvENCFF001RUM.IDRv2.IDR0.02.narrowPeak.gz](#)

IDR peaks from replicate 1 self-pseudoreplicates (r1pr\_peaks)  
[ENCFF001RULvENCFF001RUL.IDRv2.IDR0.02.narrowPeak.gz](#)

IDR peaks from pooled self-pseudoreplicates (pooledpr\_peaks)  
[ENCFF001RULvENCFF001RUL.IDRv2.IDR0.02.narrowPeak.gz](#)

chrom.sizes for bedToBigBed (chrom\_sizes)  
[ENCODE - Production runs ▶ male.hg19.chrom.sizes](#)

Blacklist (gzipped BED file) (blacklist)  
[ENCODE Reference Files ▶ wgEncodeDacMapabilityConsensusExcludable.bed.gz](#)

IDR peaks from true replicates (reps\_peaks)  
[ENCFF001RULvENCFF001RUM.IDRv2.IDR0.02.narrowPeak.gz](#)

.as file for bedToBigBed (as\_file)  
[ENCODE Reference Files ▶ narrowPeak.as](#)

Output files  
(downloadable)

Atomic outputs

## OUTPUTS

Final peak calls - conservative set (conservative\_set)  
[IDR\\_final\\_conservative.narrowPeak.gz](#)

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](#)

Final peak calls - optimal set bigBed (optimal\_set\_bb)  
[IDR\\_final\\_optimal.narrowPeak.bb](#)

Result of the reproducibility test (reproducibility\_test)  
pass

Rescue ratio (rescue\_ratio)  
1.4036567105366853

Number of peaks from pooled pseudoreplicates (Np)  
17811

Number of peaks from replicate 1 self-pseudoreplicates (N1)  
9167

Number of peaks from replicate 2 self-pseudoreplicates (N2)  
17815

Number of peaks from true replicates (Nt)  
12689

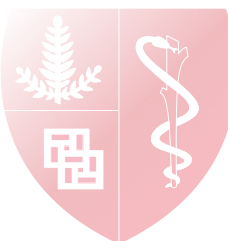
Self-consistency ratio (self\_consistency\_ratio)  
1.9433838769499292



# Summary and Development Plans

---

- DCC have implemented ChIP analysis pipelines as specified by the DAC.
- Signal tracks and peak calls are produced, along with QC metrics.
- First deployment is to the DNAnexus cloud platform.
- All ChIP experiments from the October 2015 freeze have been run through the mapping, and TF or histone pipelines.
- Anyone can run these pipelines on DNAnexus.
- Currently under development (early 2016 release):
  - Add GEM and PeakSeq peak callers to TF pipeline.
  - Add motif analysis.
  - Refine IDR metrics for histone pipeline.
  - Improved documentation.
  - GRCh38 processing of all experiments.



# 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 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



@encodedcc



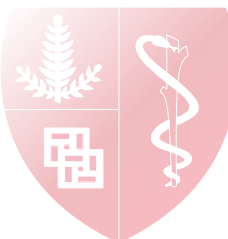
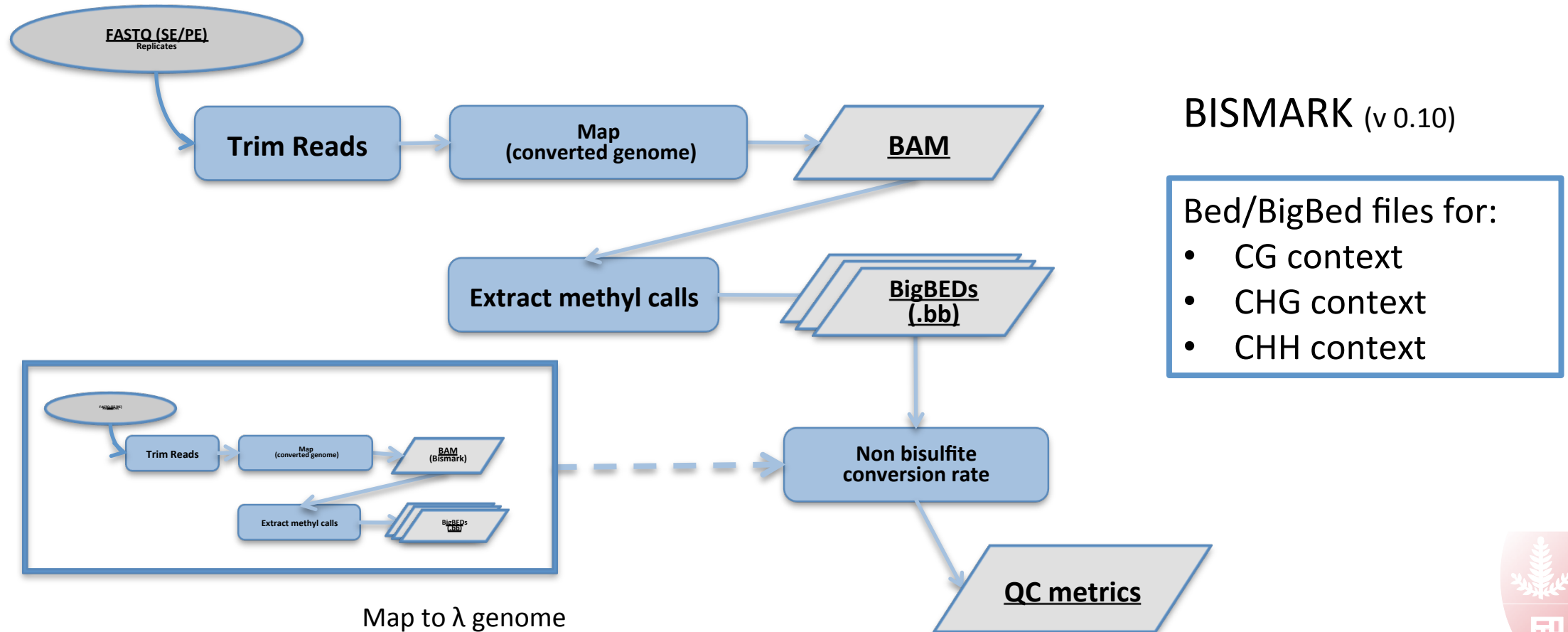
encode-help@lists.stanford.edu

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



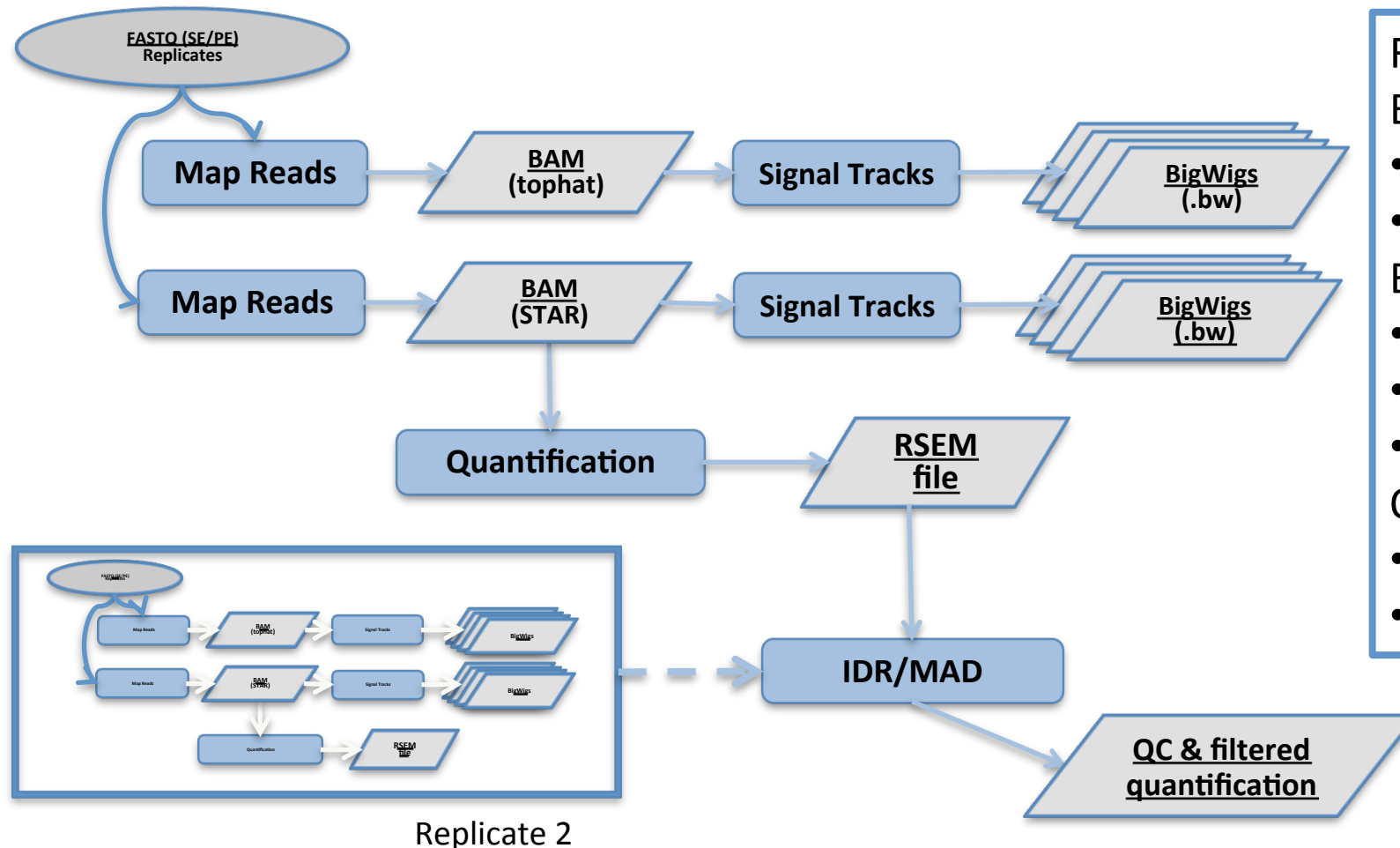
# Schema: ENCODE WGBS Pipeline

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



# 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



# Pick up the Results

ENCORE RNA-Seq (Long) Pipeline - 1 replicate (paired-end)

Launch as new Analysis Save as new Workflow View Info

ENCORE RNA-Seq (Long) Pipeline - 1 replicate (paired-end) analysis-Bf6k7zQ08BgPg4Zbxjzy966Y 06/26/2015 11:03 am Benjamin Hitz 1h 9m \$2.58  
executable execution id launched on launched by ran for cost

11:08AM	11:25AM	11:37AM	11:49AM	06/26/2015 12:12:16PM
✓ Long-RNA-Seq-align-star-pe				37m Log
✓ Long-RNA-Seq-align-tophat-pe				36m Log
✓ Long-RNA-Seq-quantitate-RSEM				26m Log
✓ Long-RNA-Seq-BAM-to-BW-stranded				7m Log
✓ Long-RNA-Seq-BAM-to-BW-stranded				4m Log

INPUTS

Pair 1 Reads to align (fastq.gz) (stage-Bf6k3Vj0VBPYbJV2Y5F8kPP.reads\_1)  
ENCFF646CCF\_1-chr21hemi.fastq.gz

Pair 2 Reads to align (fastq.gz) (stage-Bf6k3Vj0VBPYbJV2Y5FX8kPP.reads\_2)  
ENCFF646CCF\_2-chr21hemi.fastq.gz

Genome indexed for STAR (stage-Bf6k3Vj0VBPYbJV2Y5FX8kPP.star\_index)  
hg19\_male\_v19\_ERCC\_starIndex.tgz

<input type="checkbox"/>	ENCFF646CCF_1-chr21hemi-ENCFF646CCF_2-chr21hemi_star_genome_minusUniq.bw	File 1.6...	Jun 26, 2015 3:50 PM
<input checked="" type="checkbox"/>	ENCFF646CCF_1-chr21hemi-ENCFF646CCF_2-chr21hemi_star_genome_plusAll.bw	File 1.7...	Jun 26, 2015 3:50 PM
<input type="checkbox"/>	ENCFF646CCF_1-chr21hemi-ENCFF646CCF_2-chr21hemi_star_genome_plusUniq.bw	File 1.5...	Jun 26, 2015 3:50 PM
<input type="checkbox"/>	ENCFF646CCF_1-chr21hemi-ENCFF646CCF_2-chr21hemi_star_Log.final.out	File 1.6...	Jun 26, 2015 3:45 PM
<input type="checkbox"/>	ENCFF646CCF_1-chr21hemi-ENCFF646CCF_2-chr21hemi_tophat.bam	File 44....	Jun 26, 2015 4:04 PM
<input type="checkbox"/>	ENCFF646CCF_1-chr21hemi-ENCFF646CCF_2-chr21hemi_tophat_minusAll.bw	File 1.6...	Jun 26, 2015 4:09 PM

Copy  
Delete  
Info  
Download  
Open in New Tab






# Visualize!

**Get Your Data**  
⊙ For security reasons, these download links will expire in 1m 19s

 Download file

Name	Link
ENCFF646CCF_1-chr21hemi-ENCFF646CCF_2-chr21hemi_star_genome_plusAll.bw	<a href="#">Get URL</a> <input type="button" value="Download"/>

**Get Your Data**  
⊙ For security reasons, these download links will expire in 1m 5s



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/g5xqzjXqzBk5v0vqj11jg8Pqg5Xg8x8V0YqzjPZj/ENCFF646CCF\\_1-chr21hemi-ENCFF646CCF\\_2-chr21hemi\\_star\\_genome\\_plusAll.bw](https://dl.dnanex.us/F/D/g5xqzjXqzBk5v0vqj11jg8Pqg5Xg8x8V0YqzjPZj/ENCFF646CCF_1-chr21hemi-ENCFF646CCF_2-chr21hemi_star_genome_plusAll.bw)

**Warning:** Anyone with this link can access this file without additional authentication.

genome.ucsc.edu

## UCSC Genome Bioinformatics

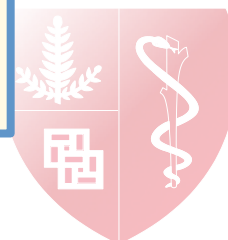
Genomes Genome Browser Tools Mirrors Downloads My Data Help About Us

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/g5xqzjXqzBk5v0vqj11jg8Pqg5Xg8x8V0YqzjPZj/ENCFF646CCF_1-chr21hemi-ENCFF646CCF_2-chr21hemi_star_genome_plusAll.bw`



# Visualize!

## Manage Custom Tracks

genome: Human assembly: Feb. 2009 (GRCh37/hg19) [hg19]

Name	Description	Type	Doc	delete
<a href="#">ENCFF646CCF_1-chr21hemi-ENCFF646CCF_2-chr21hemi_star_genome_plusAll</a>	ENCFF646CCF_1-chr21hemi-ENCFF646CCF_2-chr21hemi_star_genome_plusAll	bigWig		<input type="checkbox"/>

- add custom tracks
- go to genome browser
- go to table browser
- go to variant annotation integrator

tracks that are enabled will automatically be displayed in more compact mode.

**Custom Tracks** refresh

[ENCFF646CCF\\_1-chr21hemi-ENCFF646CCF\\_2-chr21hemi\\_star\\_genome\\_plusAll](#)

dense ▾

