ChromHMM Tutorial

Jason Ernst
Assistant Professor
University of California, Los Angeles

Talk Outline

- Chromatin states analysis and ChromHMM
- Accessing chromatin state annotations for ENCODE2 and Roadmap Epigenomics
- Running the ChromHMM software

Talk Outline

- Chromatin states analysis and ChromHMM
- Accessing chromatin state annotations for ENCODE2 and Roadmap Epigenomics
- Running the ChromHMM software

Chromatin Marks for Genome Annotation

100+ histone modifications
Specificity in:

- Histone protein
- Amino acid residue
- Chemical modification (e.g. methyl, acetylation)
- Number of occurrence of the modifications

Examples
H3K4me1 – Enhancers
H3K4me3 – Promoters
H3K27me3 – Repressive
H3K9me3 – Repressive
H3K36me3 – Transcribed

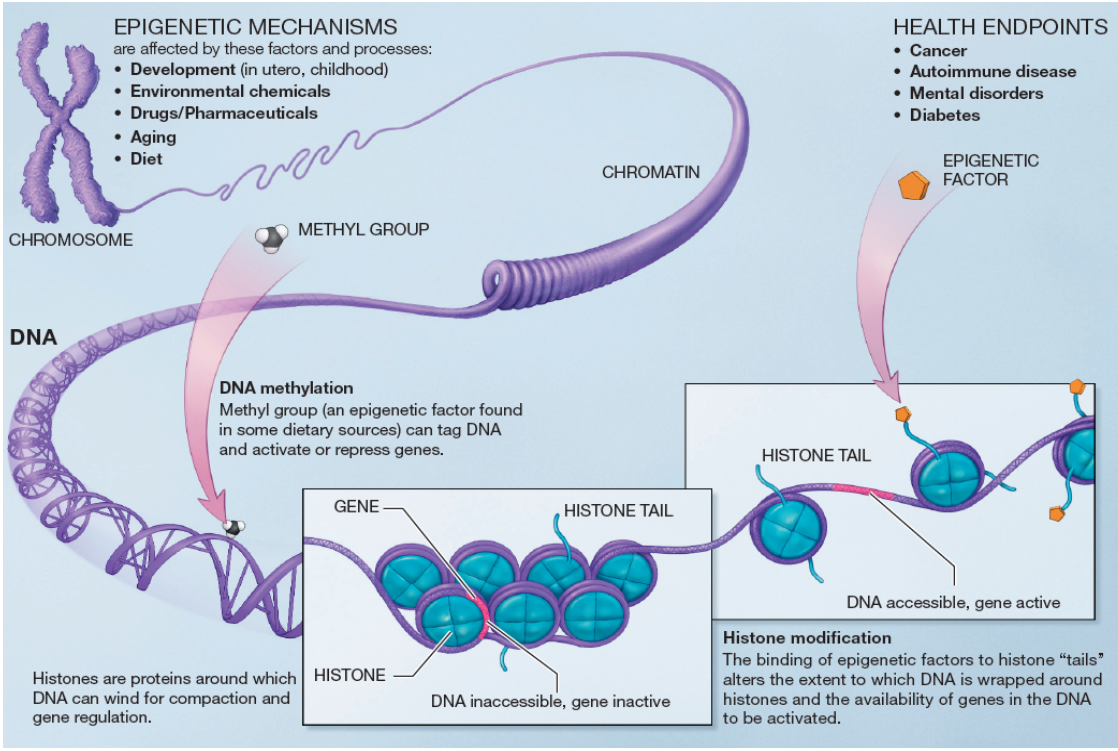
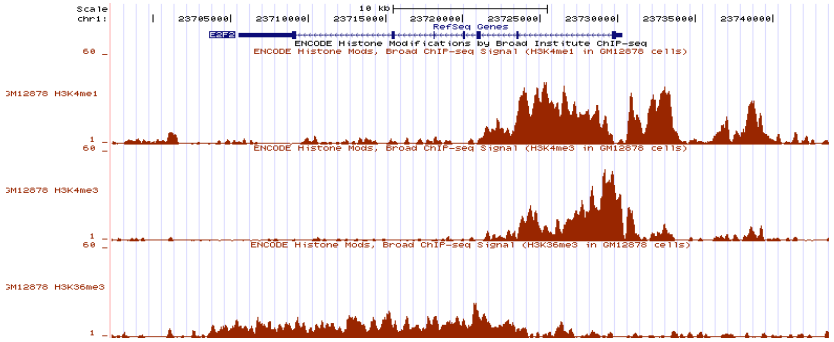
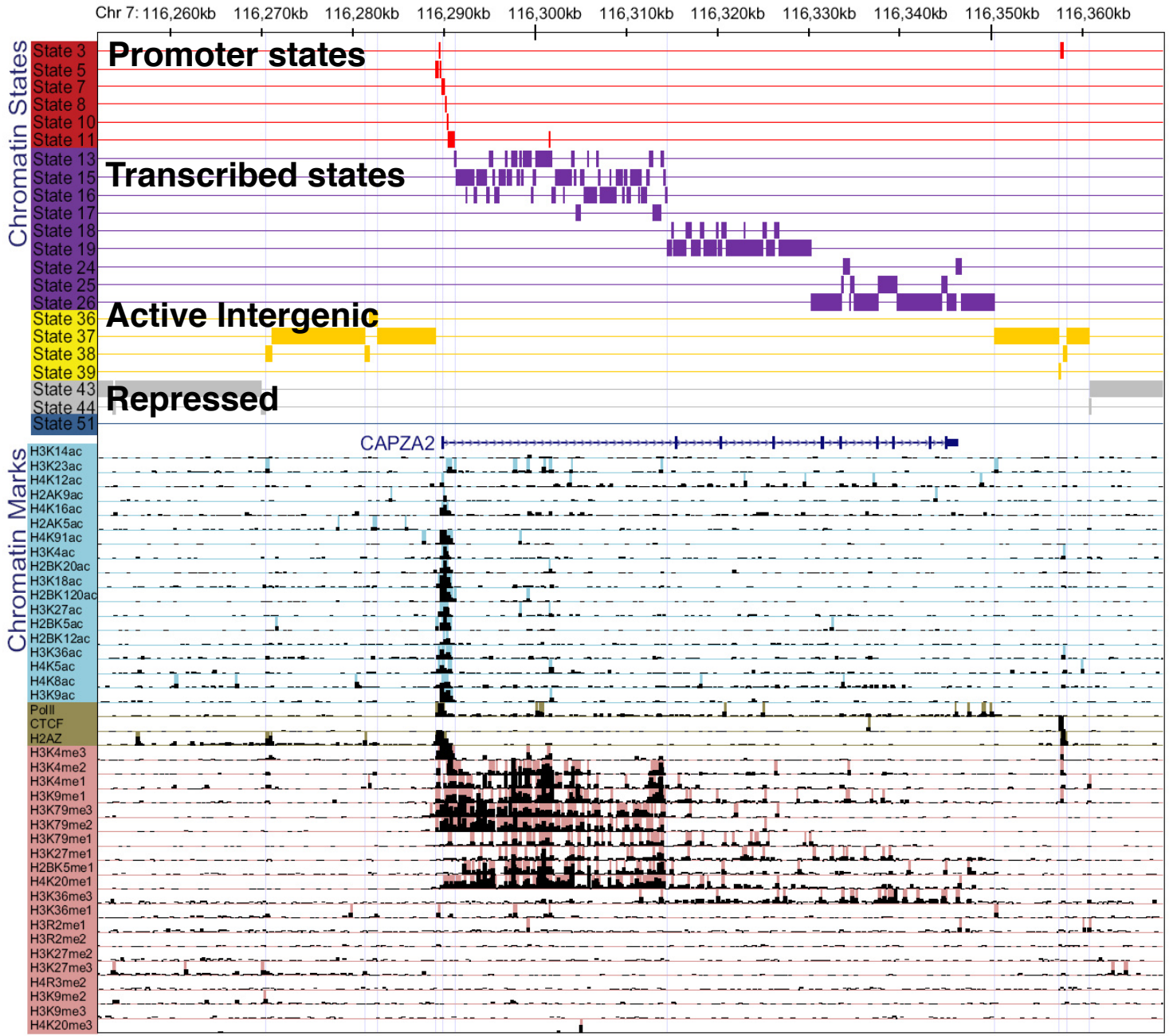


Image source: http://nihroadmap.nih.gov/epigenomics/

Histone Modifications can be Mapped Genome-wide with ChIP-seq



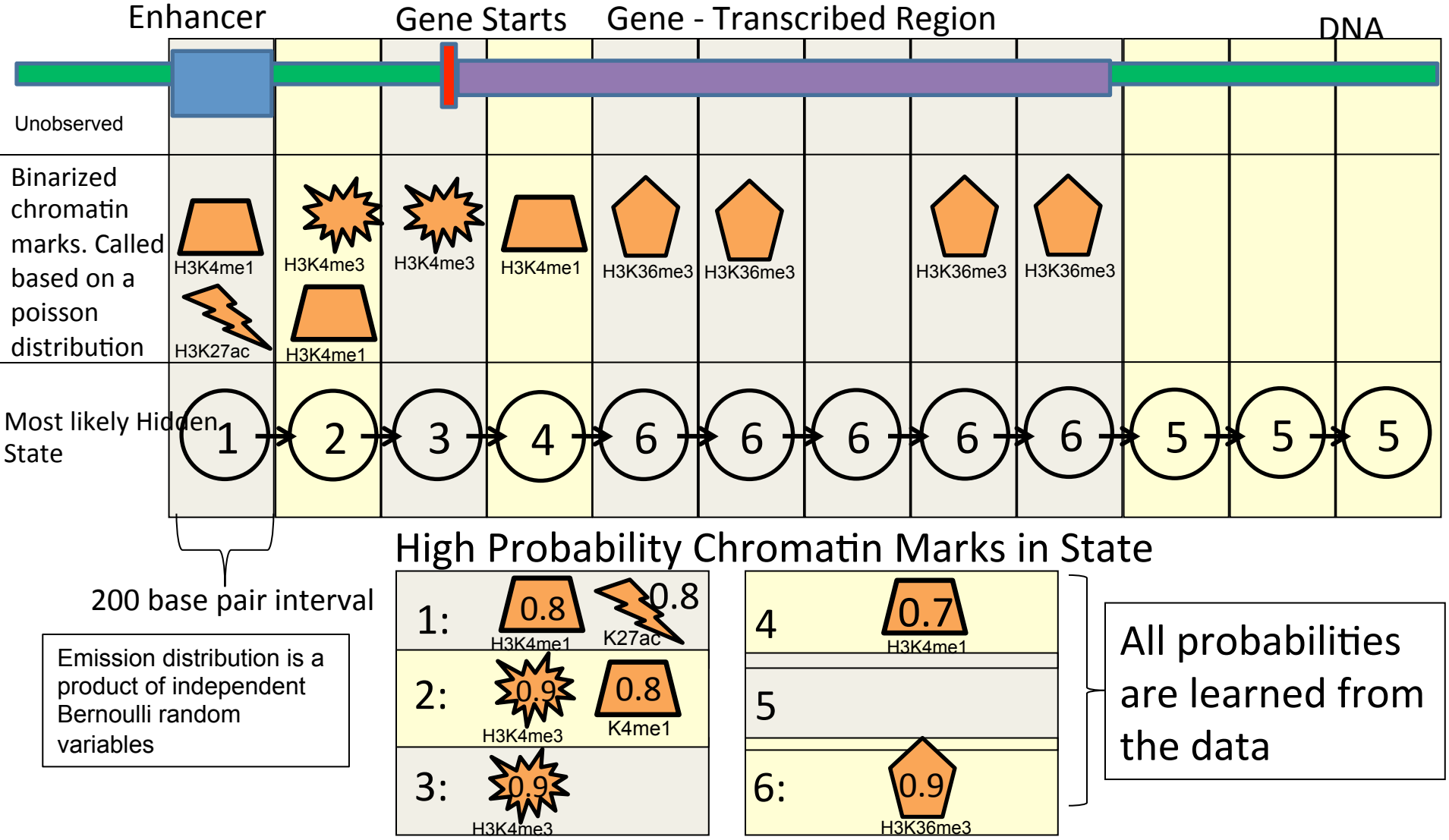
From ‘chromatin marks’ to ‘chromatin states’



- Learn *de novo* significant combinational and spatial patterns of chromatin marks
- Reveal functional elements, even without looking at sequence
- Use for genome annotation

Ernst and Kellis, Nat Biotech 2010

Our approach: Multivariate Hidden Markov Model (ChromHMM)



Binarization leads to explicit modeling of mark combinations and interpretable parameters

Ernst and Kellis, Nat Biotech 2010 ; Ernst and Kellis, Nature Methods 2012

ENCODE: Study nine marks in nine human cell lines

9 marks

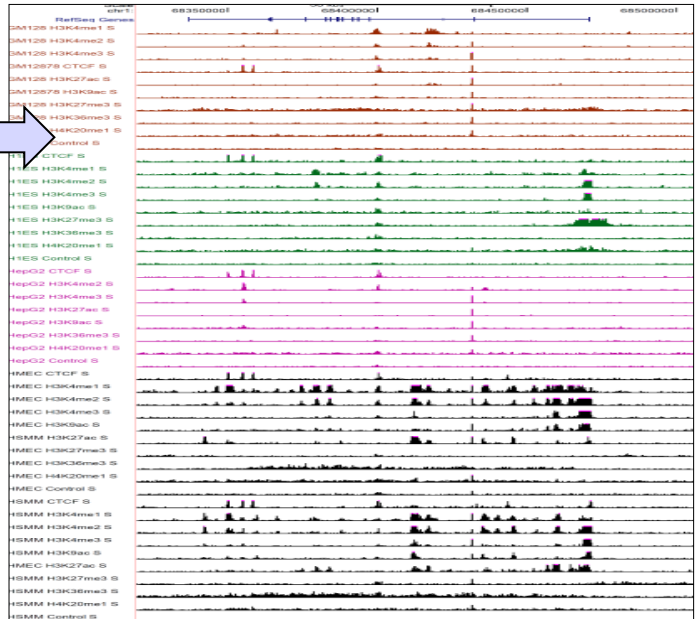
- H3K4me1
- H3K4me2
- H3K4me3
- H3K27ac
- H3K9ac
- H3K27me3
- H4K20me1
- H3K36me3
- CTCF
- +WCE
- +RNA

X

9 human cell types

HUVEC	Umbilical vein endothelial
NHEK	Keratinocytes
GM12878	Lymphoblastoid
K562	Myelogenous leukemia
HepG2	Liver carcinoma
NHLF	Normal human lung fibroblast
HMEC	Mammary epithelial cell
HSMM	Skeletal muscle myoblasts
H1	Embryonic

81 Chromatin Mark Tracks



Brad Bernstein ENCODE Group

HUVEC | NHEK | GM12878 | K562 | HepG2 | NHLF | HMEC | HSMM | H1

d

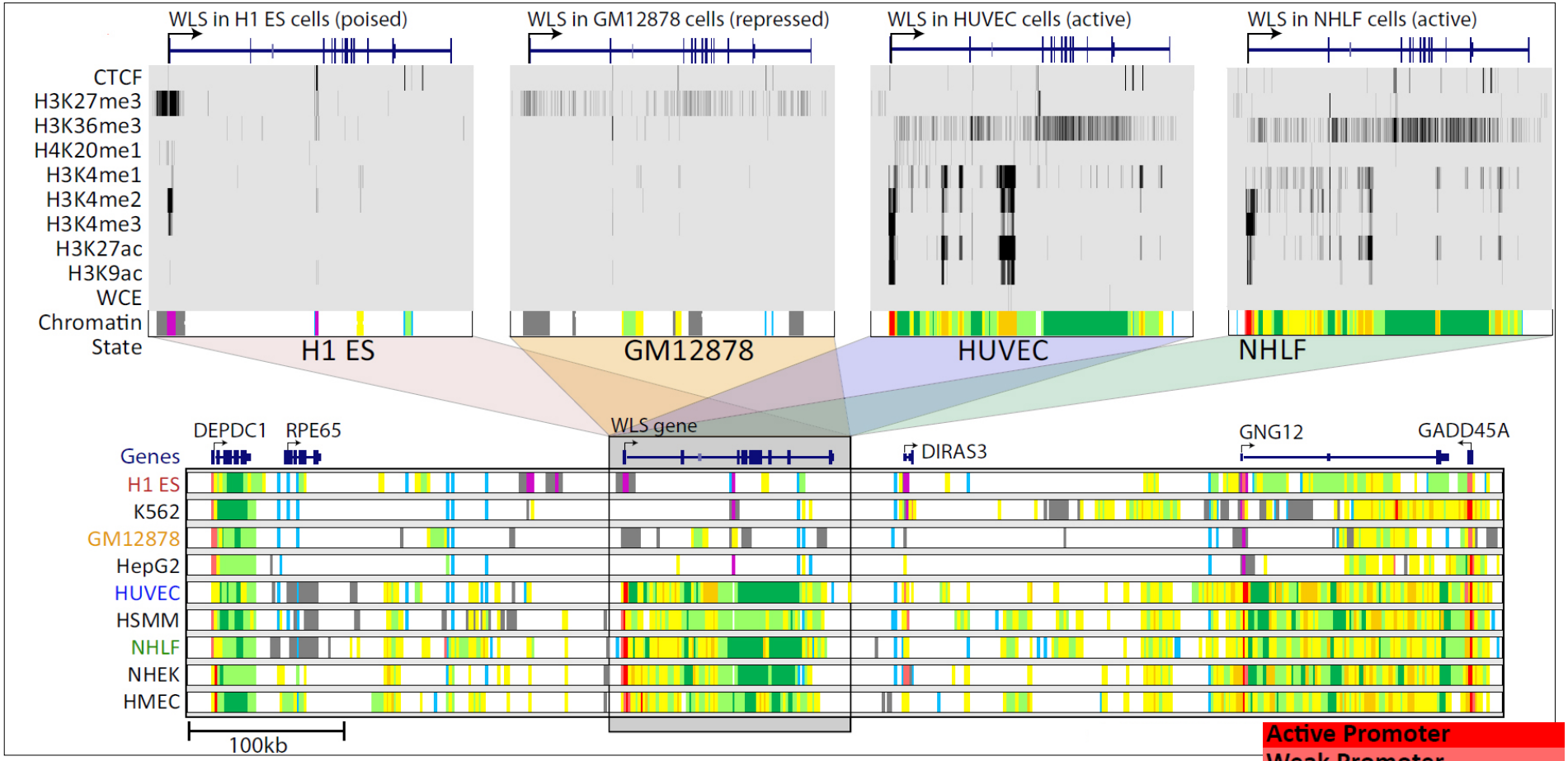
State	CTCF	H3K27me3	H3K36me3	H4K20me1	H3K4me1	H3K4me2	H3K4me3	H3K27ac	H3K9ac	WCE	Coverage: Median	Coverage: H1 ES	Coverage: GM	Median Length	+/-2kb TSS	Conserved non-exon	DNase (K562)	C-Myc (K562)	NF-κB (GM12878)	Transcript	Nuclear Lamina (NHLF)	Candidate state annotation
1	16	2	2	6	17	93	99	96	98	2	0.6	0.5	1.2	1.0	83	3.8	23.3	82.0	40.7	0.2	0.15	Active Promoter
2	12	2	6	9	53	94	95	14	44	1	0.5	1.2	1.3	0.4	58	2.8	15.3	12.6	5.8	0.6	0.30	Weak Promoter
3	13	72	0	9	48	78	49	1	10	1	0.2	4.0	1.0	0.6	49	4.3	10.8	3.1	1.0	0.4	0.68	Inactive/poised Promoter
4	11	1	15	11	96	99	75	97	86	4	0.7	0.1	1.1	0.6	23	2.7	23.1	31.8	49.0	1.3	0.05	Strong enhancer
5	5	0	10	3	88	57	5	84	25	1	1.2	0.2	0.7	0.6	3	1.8	13.6	6.3	15.8	1.4	0.10	Strong enhancer
6	7	1	1	3	58	75	8	6	5	1	0.9	1.3	1.0	0.2	17	2.4	11.9	5.7	7.0	1.1	0.31	Weak/poised enhancer
7	2	1	2	1	56	3	0	6	2	1	1.9	1.2	1.1	0.4	4	1.5	5.1	0.6	2.4	1.3	0.20	Weak/poised enhancer
8	92	2	1	3	6	3	0	0	1	1	0.5	1.4	1.0	0.4	3	1.5	12.8	2.5	1.2	1.1	0.61	Insulator
9	5	0	43	43	37	11	2	9	4	1	0.7	1.3	1.0	0.8	4	1.1	4.5	0.7	0.8	2.4	0.02	Transcriptional transition
10	1	0	47	3	0	0	0	0	0	1	4.3	0.6	1.2	3.0	1	0.9	0.3	0.0	0.0	2.5	0.11	Transcriptional elongation
11	0	0	3	2	0	0	0	0	0	0	12.5	1.3	0.8	2.6	2	0.9	0.3	0.0	0.1	1.9	0.24	Weak transcribed
12	1	27	0	2	0	0	0	0	0	0	4.1	0.3	0.7	2.8	5	1.4	0.3	0.0	0.1	0.8	0.63	Polycomb-repressed
13	0	0	0	0	0	0	0	0	0	0	71.4	1.0	1.0	10.0	1	0.9	0.1	0.0	0.0	0.7	1.30	Heterochrom; low signal
14	22	28	19	41	6	5	26	5	13	37	0.1	0.9	1.2	0.6	3	0.4	1.9	0.3	0.2	0.4	1.44	Repetitive/CNV
15	85	85	91	88	76	77	91	73	85	78	0.1	0.9	1.0	0.2	1	0.2	5.9	9.5	7.4	0.4	1.30	Repetitive/CNV

Chromatin States

Chromatin Mark Observation Frequency (%)

(%) (fold) (kb) (%) Functional enrichments (fold)

Chromatin states dynamics across nine ENCODE cell types



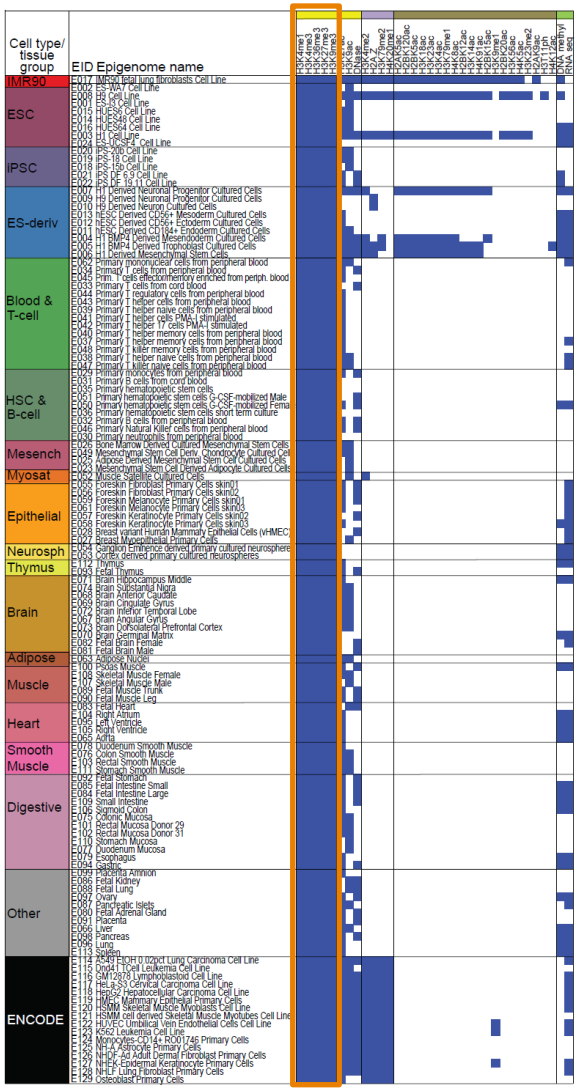
- **Single annotation track for each cell type**
- **Summarize cell-type activity at a glance**
- **Can study 9-cell activity pattern across ↓**

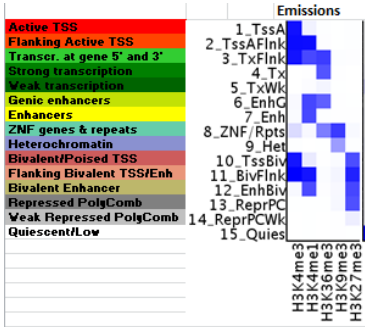
Ernst et al, Nature 2011

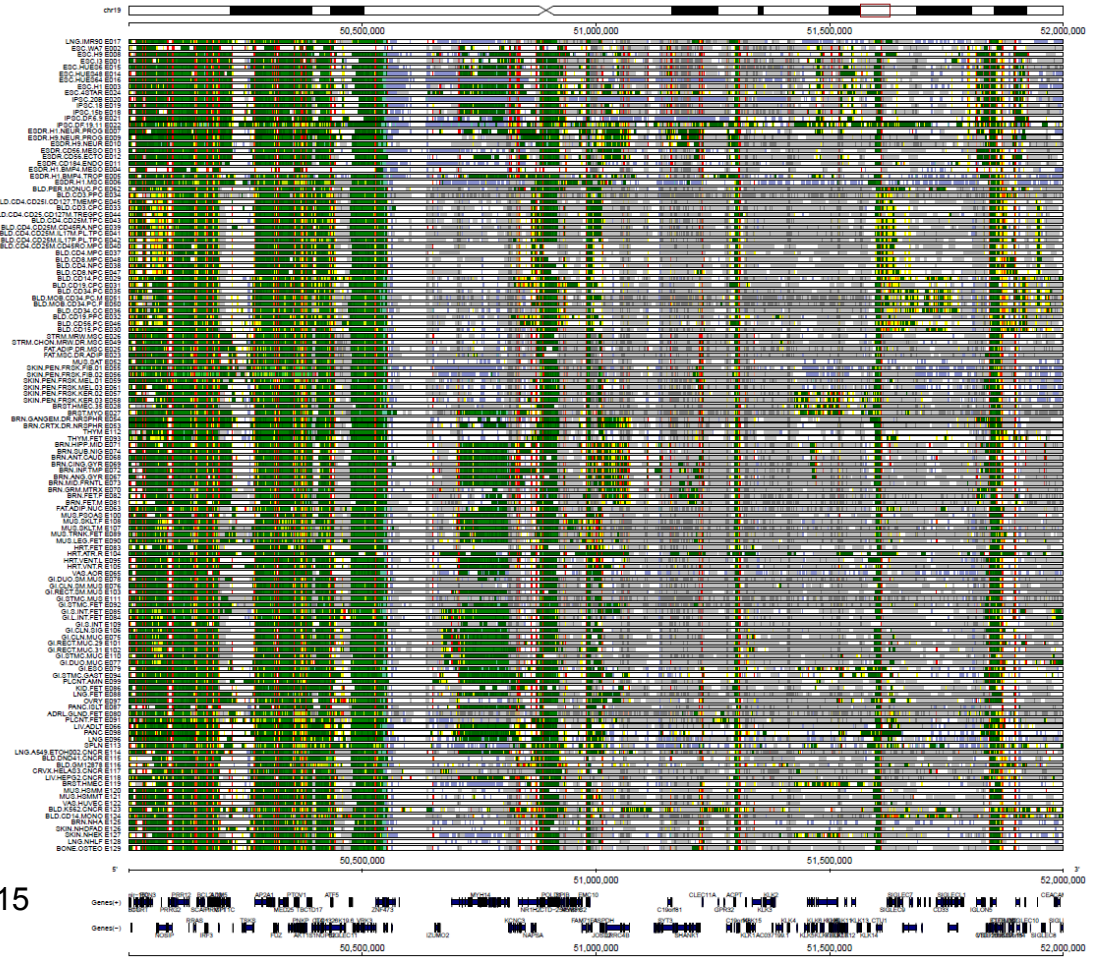
Talk Outline

- Chromatin states analysis and ChromHMM
- Accessing chromatin state annotations for ENCODE2 and Roadmap Epigenomics
- Running the ChromHMM software

Chromatin States Defined Across 127 Cell/Tissues Types



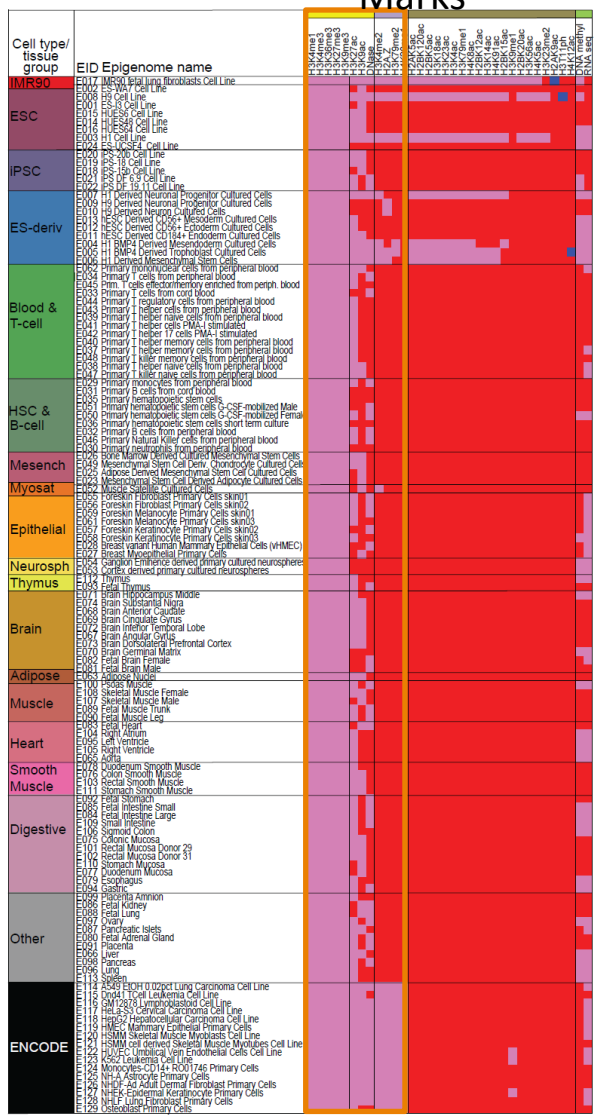


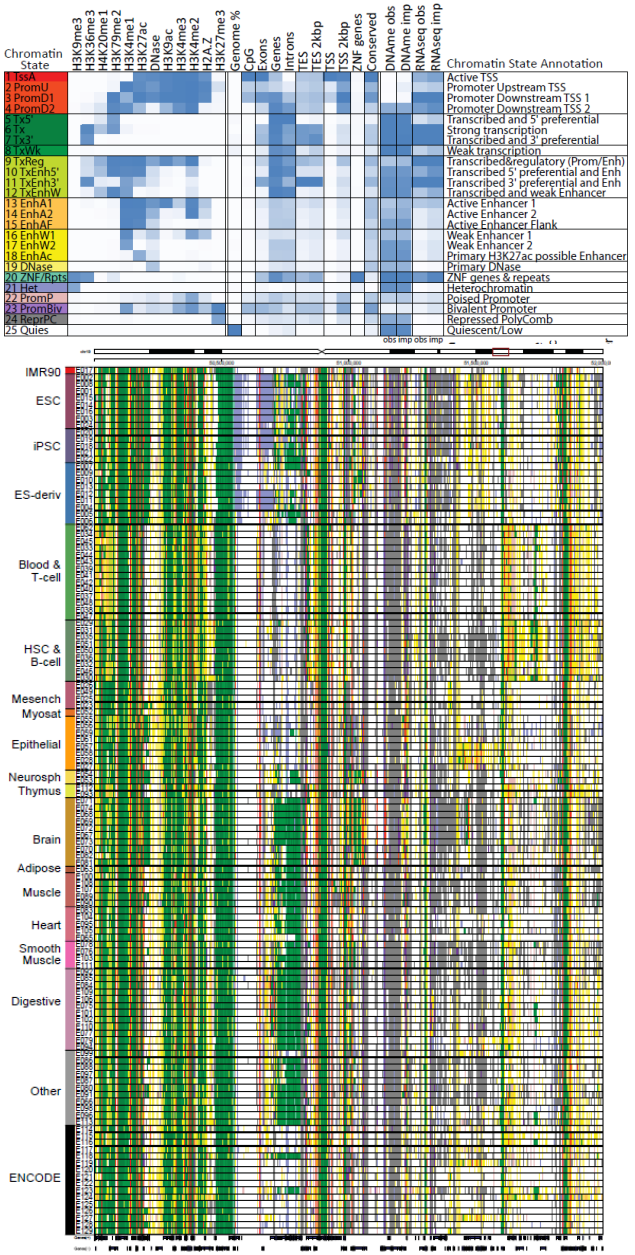


16 epigenomes from ENCODE 2

Roadmap Epigenomics Consortium et al, *Nature* 2015

Chromatin States Defined on Imputed Data





Observed + Imputed | Imputed Only | Observed Only

ChromImpute method

Ernst and Kellis, *Nature Biotech* 2015

ChromHMM Models across Many Roadmap/ ENCODE Cell and Tissue Types

Primary Core Marks segmentation

- State 1 - Red - TssA (Active_TSS)
- State 2 - OrangeRed - TssAFlnk (Flanking_Active_TSS)
- State 3 - LimeGreen - TxFlnk (Transcr_at_gene_5_and_3primer)
- State 4 - Green - Tx (Strong_transcription)
- State 5 - DarkGreen - TxWk (Weak_transcription)
- State 6 - GreenYellow - EnhG (Genic_enhancers)
- State 7 - Yellow - Enh (Enhancers)
- State 8 - MediumAquamarine - ZNF/Rpts (ZNF_genes&repeats)
- State 9 - PaleTurquoise - Het (Heterochromatin)
- State 10 - IndianRed - TssBiv (Bivalent/Poised_TSS)
- State 11 - DarkSalmon - BivFlnk (Flanking_Bivalent_TSS/Enh)
- State 12 - DarkKhaki - EnhBiv (Bivalent_Enhancer)
- State 13 - Silver - ReprPC (Repressed_PolyComb)
- State 14 - Gainsboro - ReprPCWk (Weak_Repressed_PolyComb)
- State 15 - White - Quies (Quiescent/Low)

127 Cell/Tissue Types

H3K4me1
H3K4me3
H3K27me3
H3K9me3
H3K36me3

Auxiliary Core Marks + K27ac segmentation

- State 1 - Red - TssA (Active_TSS)
- State 2 - Orange_Red - TssFlnk (Flanking_TSS)
- State 3 - Orange_Red - TssFlnkU (Flanking_TSS_Upstream)
- State 4 - Orange_Red - TssFlnkD (Flanking_TSS_Downstream)
- State 5 - Green - Tx (Strong_transcription)
- State 6 - DarkGreen - TxWk (Weak_transcription)
- State 7 - GreenYellow - EnhG1 (Genic_enhancer1)
- State 8 - GreenYellow - EnhG2 (Genic_enhancer2)
- State 9 - Orange - EnhA1 (Active_Enhancer1)
- State 10 - Orange - EnhA2 (Active_Enhancer2)
- State 11 - Yellow - EnhWk (Weak_Enhancer)
- State 12 - Medium_Aquamarine - ZNF/Rpts (ZNF_genes&repeats)
- State 13 - PaleTurquoise - Het (Heterochromatin)
- State 14 - IndianRed - TssBiv (Bivalent/Poised_TSS)
- State 15 - DarkKhaki - EnhBiv (Bivalent_Enhancer)
- State 16 - Silver - ReprPC (Repressed_PolyComb)
- State 17 - Gainsboro - ReprPCWk (Weak_Repressed_PolyComb)
- State 18 - White - Quies (Quiescent/Low)

98 Cell/Tissue Types

H3K4me1
H3K4me3
H3K27me3
H3K9me3
H3K36me3
H3K27ac

Imputed Marks Segmentation

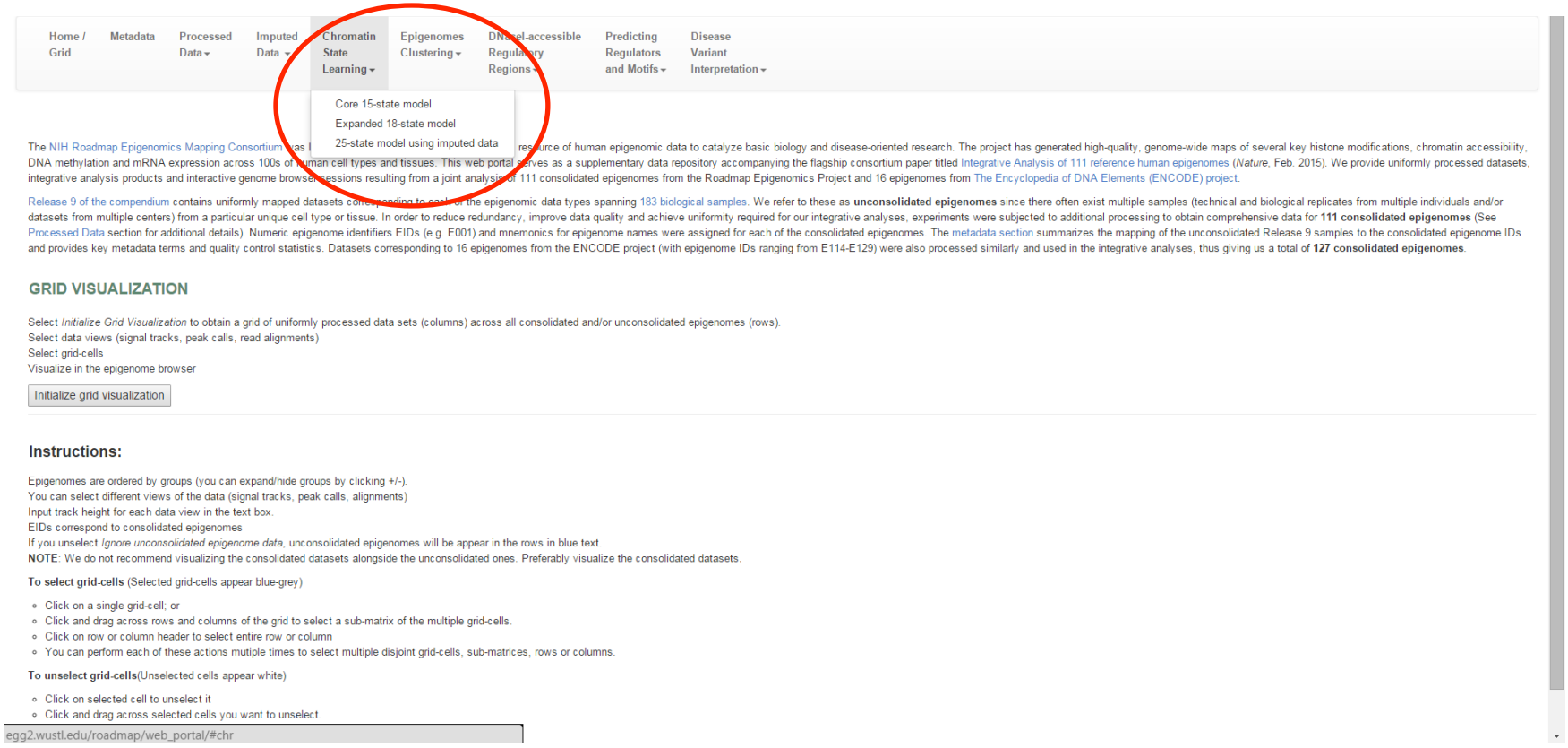
- State 1 - Red - TssA (Active TSS)
- State 2 - Orange Red - PromU (Promoter Upstream TSS)
- State 3 - Orange Red - PromD1 (Promoter Downstream TSS with DNase)
- State 4 - Orange Red - PromD2 (Promoter Downstream TSS)
- State 5 - Green - Tx5' (Transcription 5')
- State 6 - Green - Tx (Transcription)
- State 7 - Green - Tx3' (Transcription 3')
- State 8 - Light Green - TxWk (Weak transcription)
- State 9 - GreenYellow - TxReg (Transcription Regulatory)
- State 10 - GreenYellow - TxEnh5' (Transcription 5' Enhancer)
- State 11 - GreenYellow - TxEnh3' (Transcription 3' Enhancer)
- State 12 - GreenYellow - TxEnhW (Transcription Weak Enhancer)
- State 13 - Orange - EnhA1 (Active Enhancer 1)
- State 14 - Orange - EnhA2 (Active Enhancer 2)
- State 15 - Orange - EnhAF (Active Enhancer Flank)
- State 16 - Yellow - EnhW1 (Weak Enhancer 1)
- State 17 - Yellow - EnhW2 (Weak Enhancer 2)
- State 18 - Yellow - EnhAc (Enhancer Acetylation Only)
- State 19 - Light Yellow - DNase (DNase only)
- State 20 - Medium Aquamarine - ZNF/Rpts (ZNF genes & repeats)
- State 21 - PaleTurquoise - Het (Heterochromatin)
- State 22 - Light Purple - PromP (Poised Promoter)
- State 23 - Purple - PromBiv (Bivalent Promoter)
- State 24 - Silver - ReprPC (Repressed PolyComb)
- State 25 - White - Quies (Quiescent/Low)

127 Cell/Tissue Types

H3K4me1
H3K4me3
H3K27me3
H3K9me3
H3K36me3
H3K27ac
H3K9ac
H4K20me1
H3K79me2
H3K4me2
H2A.Z
DNase

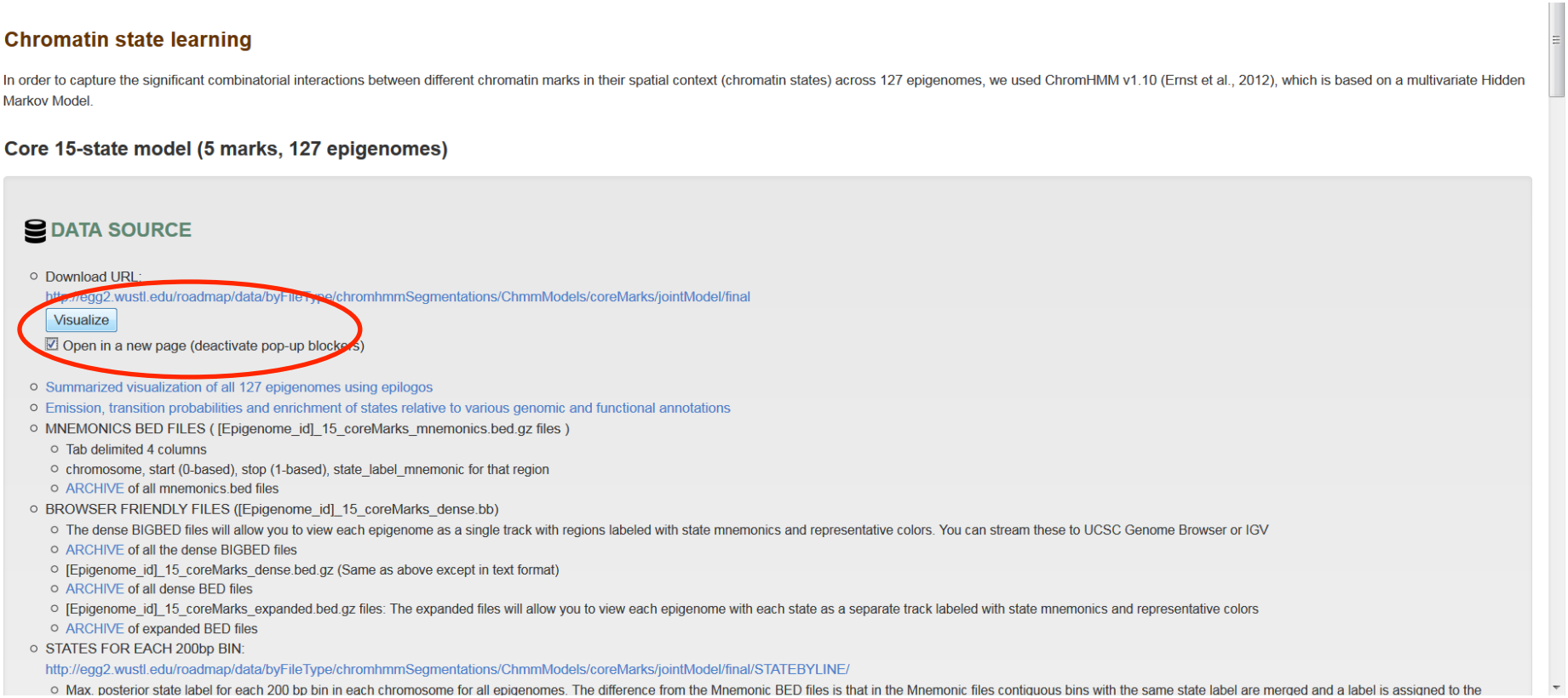
Roadmap Epigenomics Integrative Analysis Portal

http://compbio.mit.edu/roadmap



Roadmap Epigenomics Integrative Analysis Portal

http://compbio.mit.edu/roadmap



Chromatin state learning

In order to capture the significant combinatorial interactions between different chromatin marks in their spatial context (chromatin states) across 127 epigenomes, we used ChromHMM v1.10 (Ernst et al., 2012), which is based on a multivariate Hidden Markov Model.

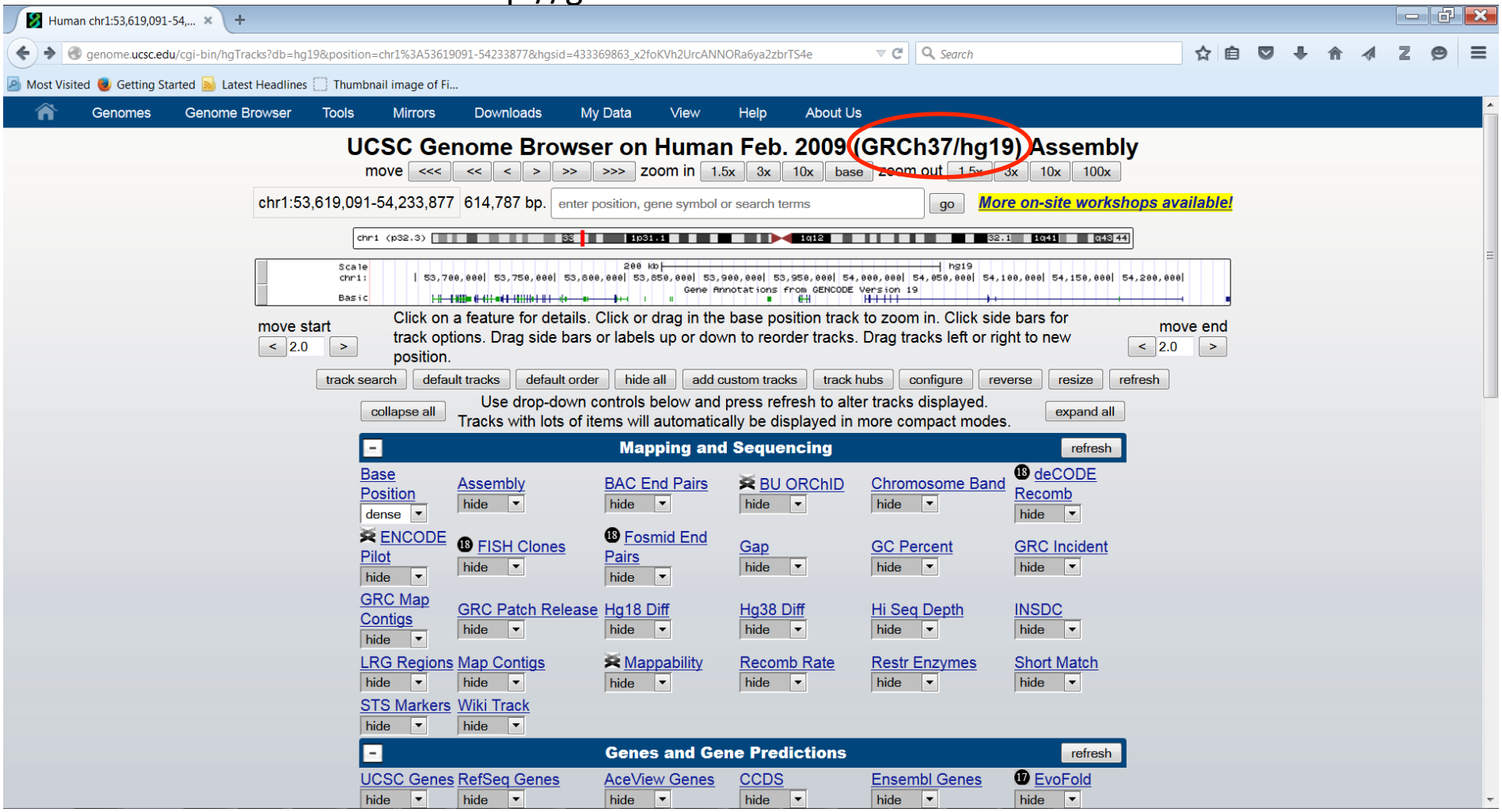
Core 15-state model (5 marks, 127 epigenomes)

DATA SOURCE

- Download URL:
  http://egg2.wustl.edu/roadmap/data/byFileType/chromhmmSegmentations/ChmmModels/coreMarks/jointModel/final
  - Visualize
  - ☑ Open in a new page (deactivate pop-up blockers)
- Summarized visualization of all 127 epigenomes using epilogos
- Emission, transition probabilities and enrichment of states relative to various genomic and functional annotations
- MNEMONICS BED FILES ([Epigenome_id]_15_coreMarks_mnemonics.bed.gz files)
  - Tab delimited 4 columns
  - chromosome, start (0-based), stop (1-based), state_label_mnemonic for that region
  - ARCHIVE of all mnemonics.bed files
- BROWSER FRIENDLY FILES ([Epigenome_id]_15_coreMarks_dense.bb)
  - The dense BIGBED files will allow you to view each epigenome as a single track with regions labeled with state mnemonics and representative colors. You can stream these to UCSC Genome Browser or IGV
  - ARCHIVE of all the dense BIGBED files
  - [Epigenome_id]_15_coreMarks_dense.bed.gz (Same as above except in text format)
  - ARCHIVE of all dense BED files
  - [Epigenome_id]_15_coreMarks_expanded.bed.gz files: The expanded files will allow you to view each epigenome with each state as a separate track labeled with state mnemonics and representative colors
  - ARCHIVE of expanded BED files
- STATES FOR EACH 200bp BIN:
  http://egg2.wustl.edu/roadmap/data/byFileType/chromhmmSegmentations/ChmmModels/coreMarks/jointModel/final/STATEBYLINE/
  - Max. posterior state label for each 200 bp bin in each chromosome for all epigenomes. The difference from the Mnemonic BED files is that in the Mnemonic files contiguous bins with the same state label are merged and a label is assigned to the

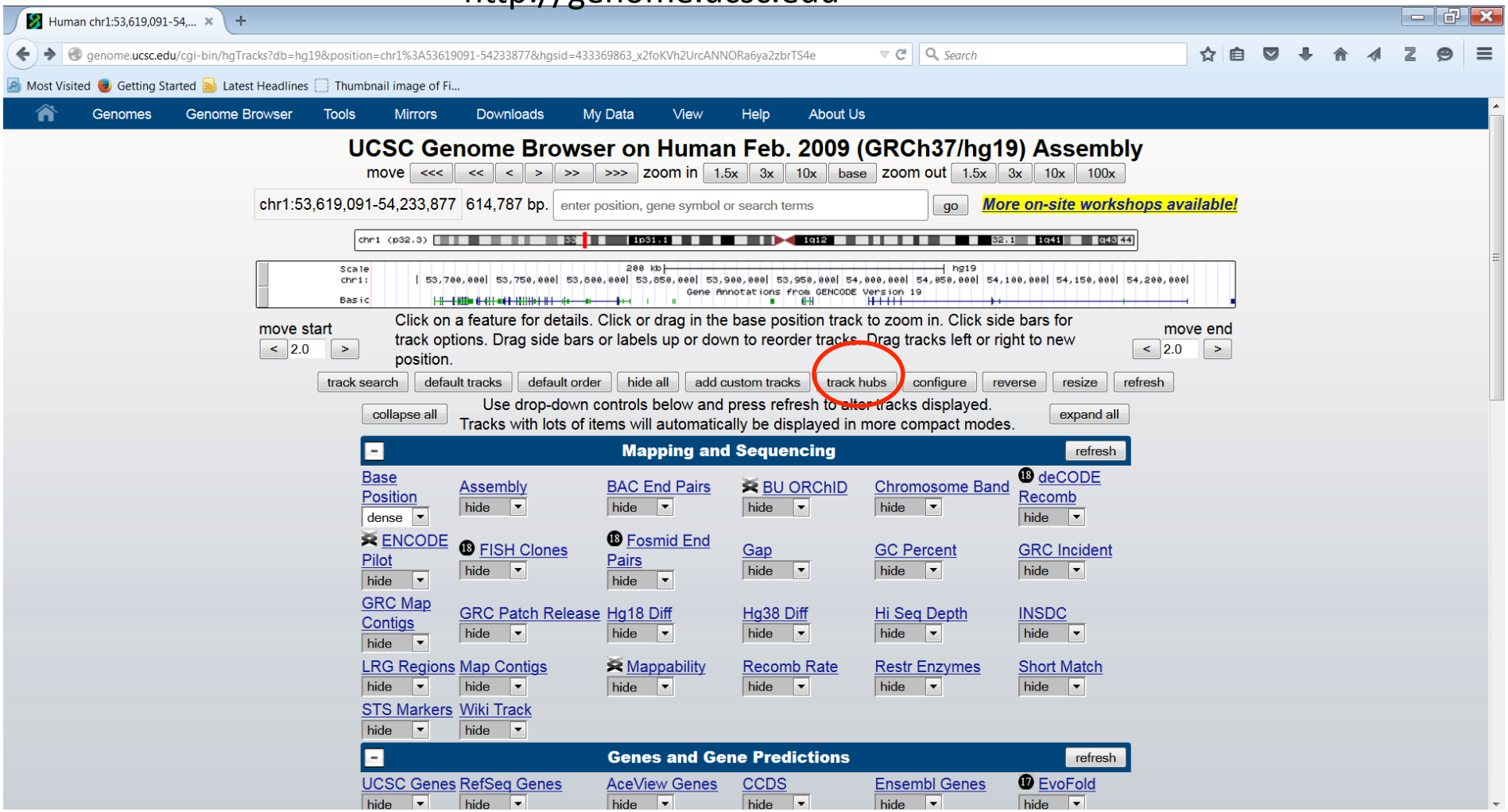
Accessing Roadmap ChromHMM through the UCSC Genome Browser Track Hubs

http://genome.ucsc.edu

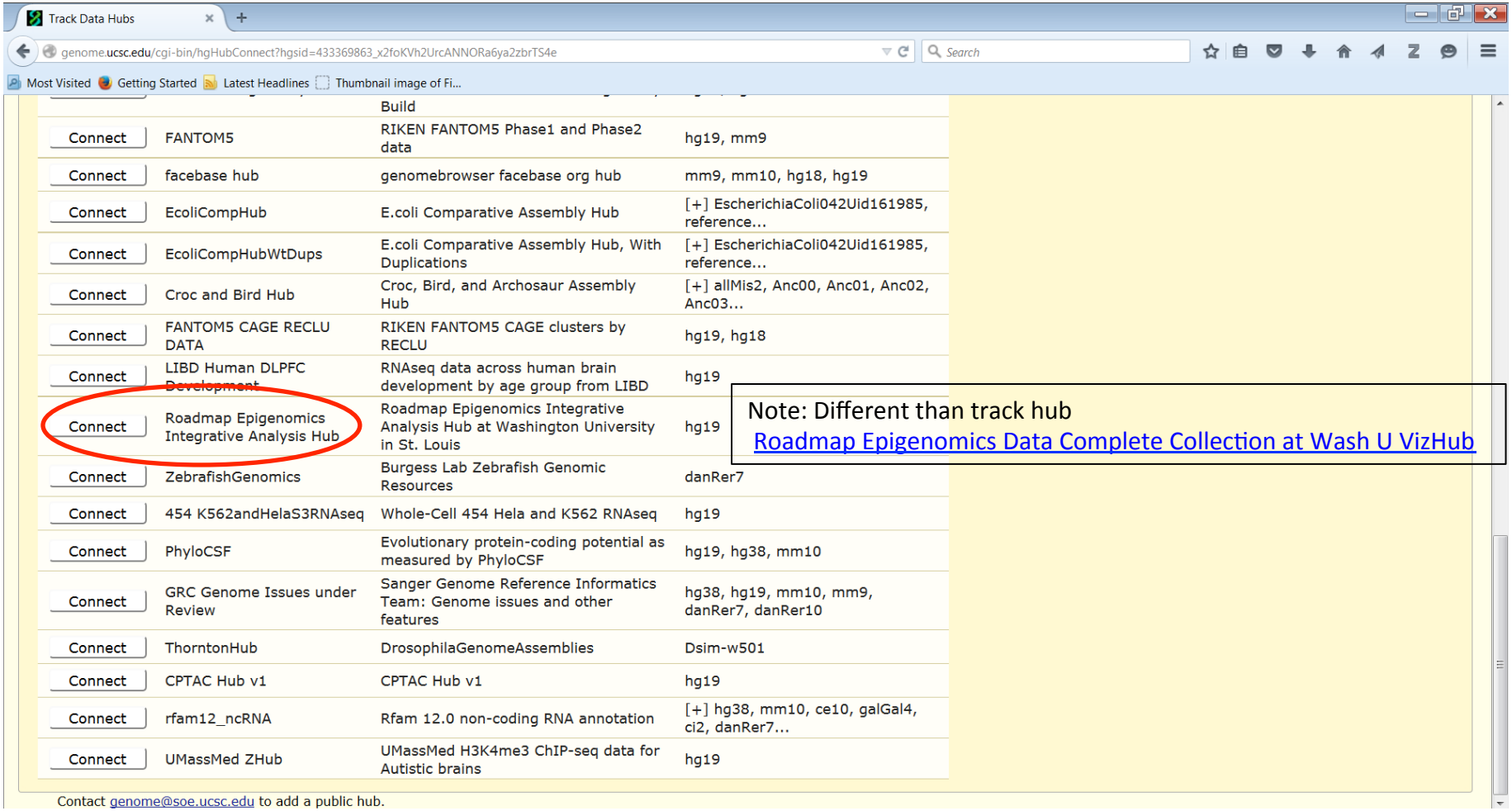


Accessing Roadmap ChromHMM through the UCSC Genome Browser Track Hubs

http://genome.ucsc.edu



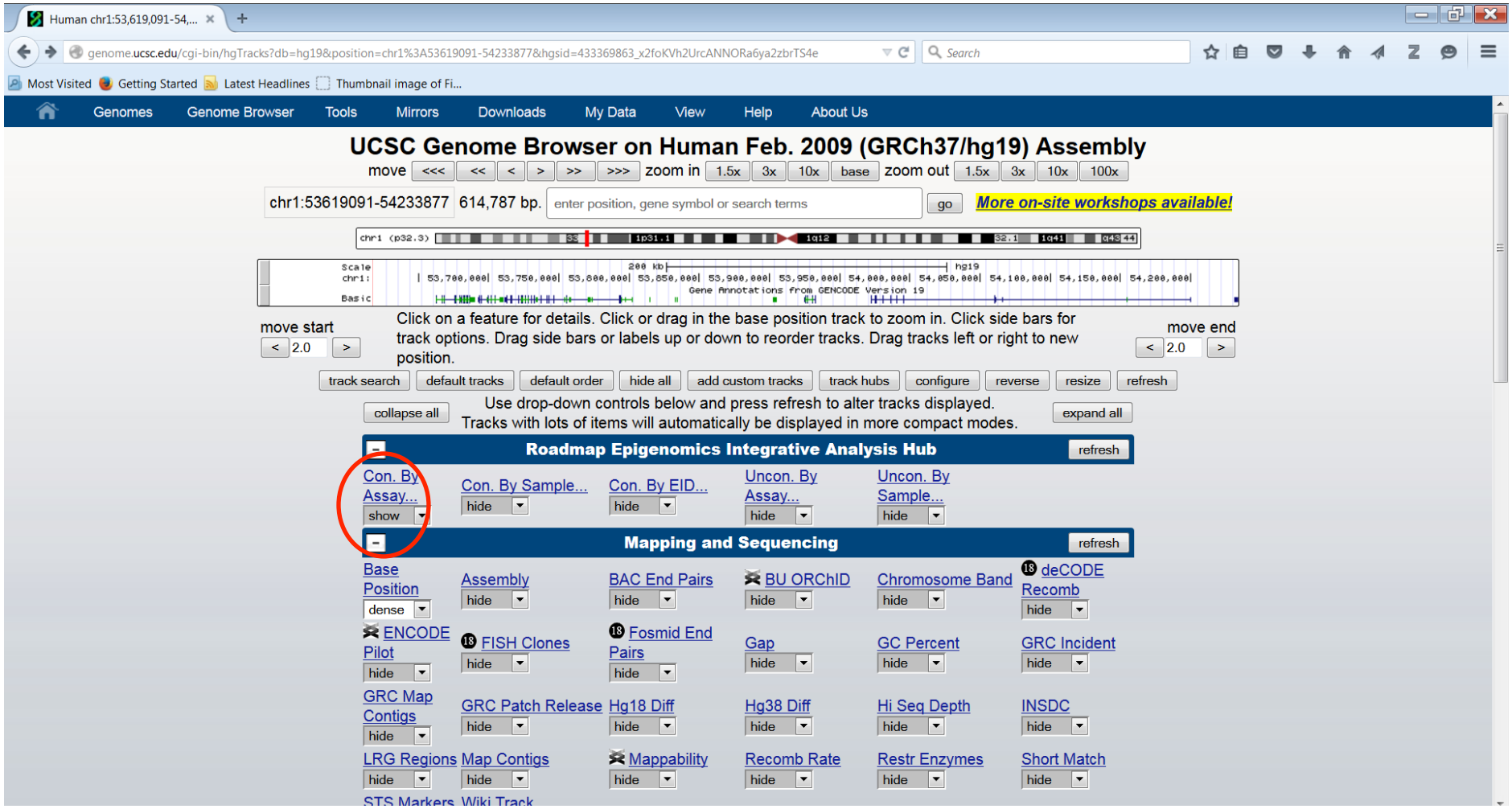
Accessing Roadmap ChromHMM through the UCSC Genome Browser Track Hubs



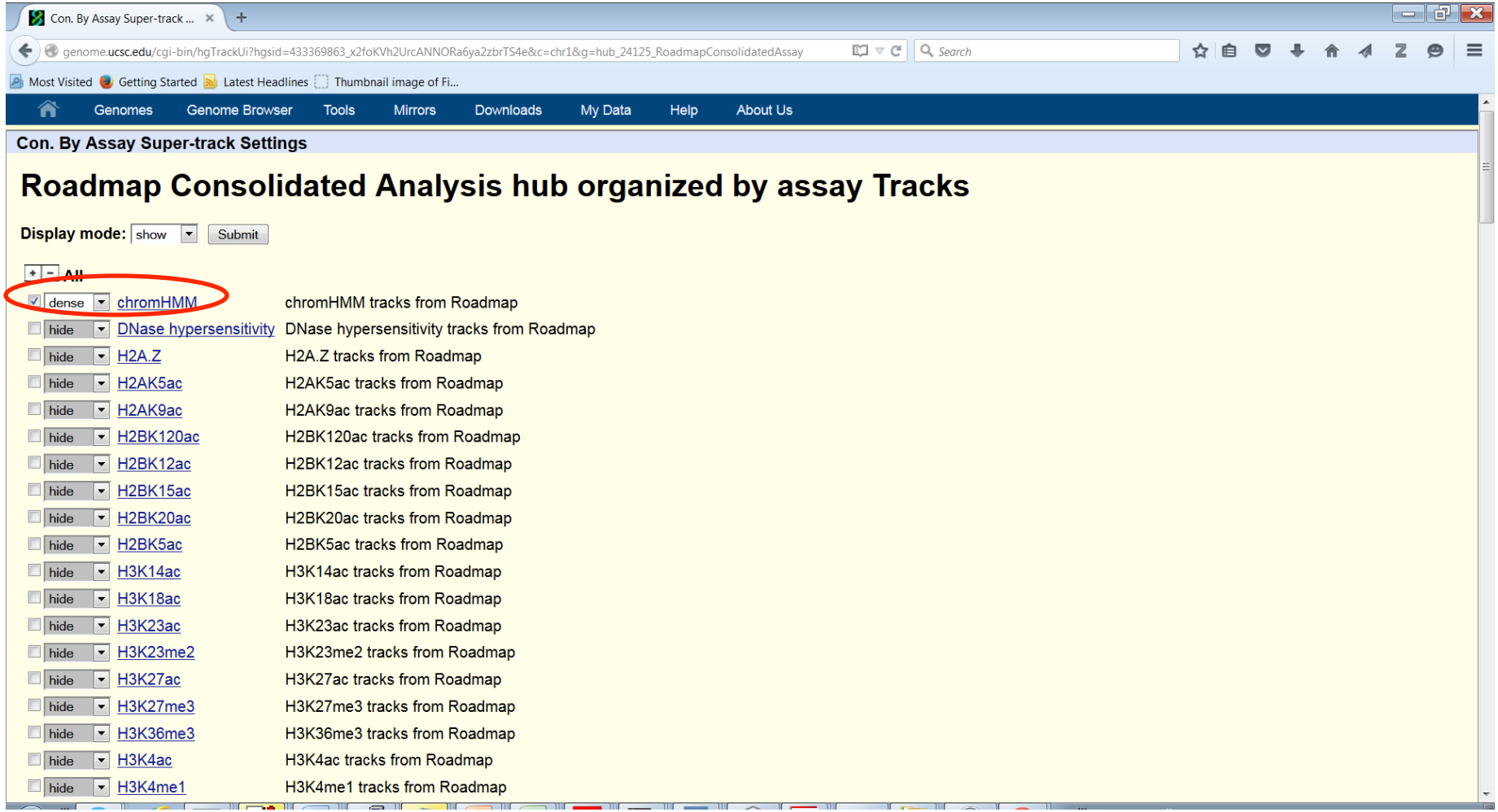
Note: Different than track hub

[Roadmap Epigenomics Data Complete Collection at Wash U VizHub]

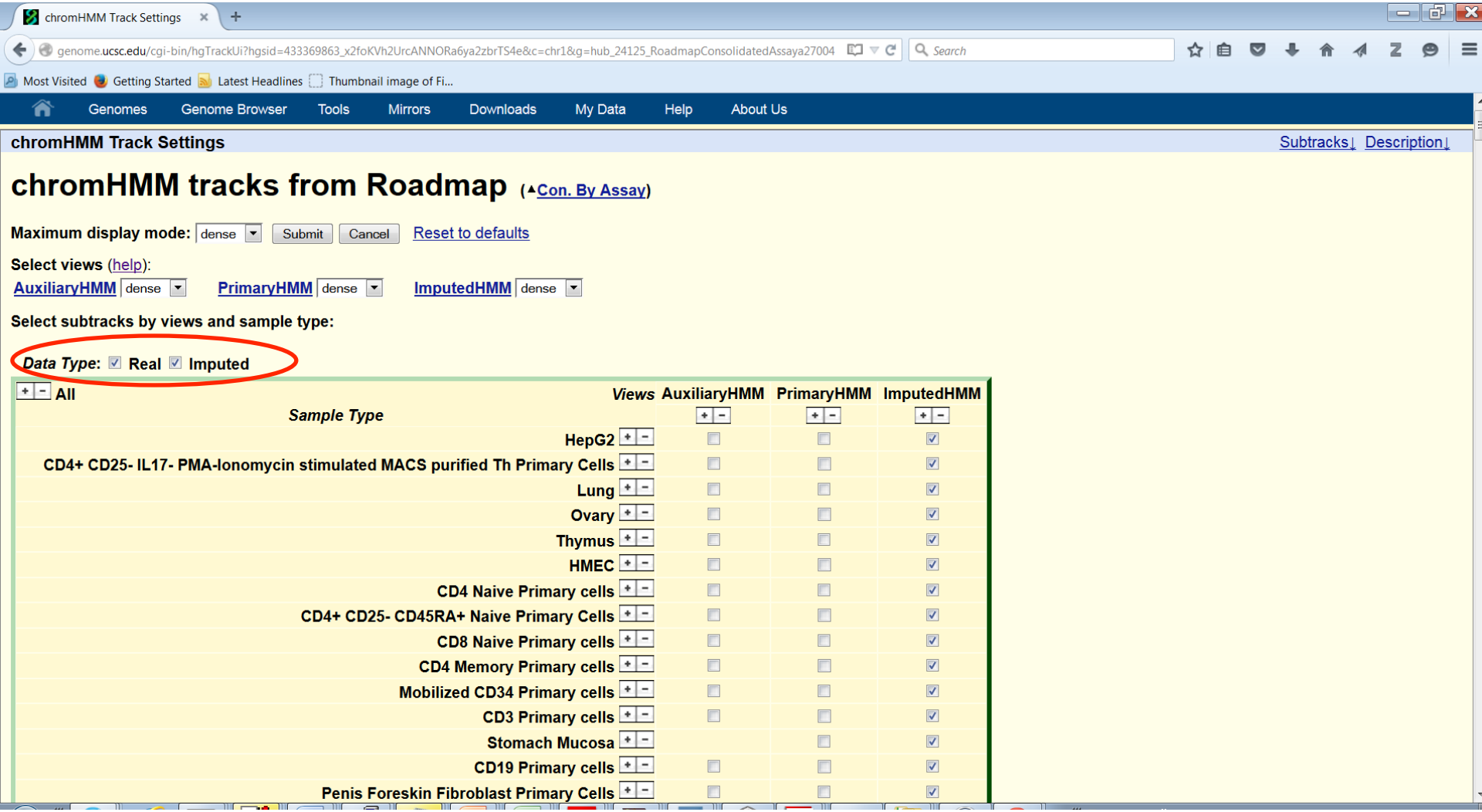
Accessing Roadmap ChromHMM through the UCSC Genome Browser Track Hubs



Accessing Roadmap ChromHMM through the UCSC Genome Browser Track Hubs



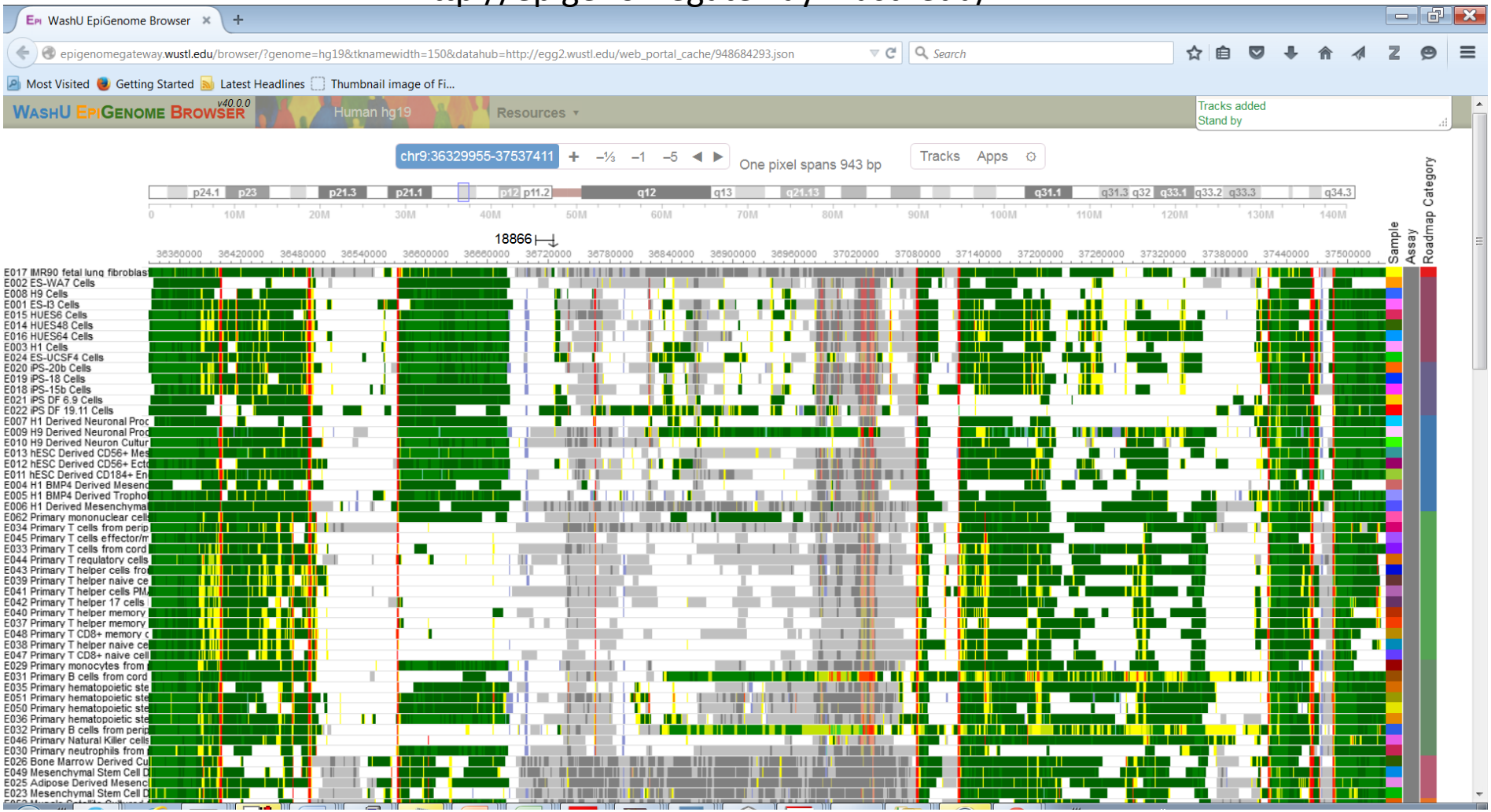
Accessing Roadmap ChromHMM through the UCSC Genome Browser Track Hubs



Accessing Roadmap ChromHMM through the UCSC Genome Browser Track Hubs

Human Epigenome Browser at Washington University

http://epigenomegateway.wustl.edu/

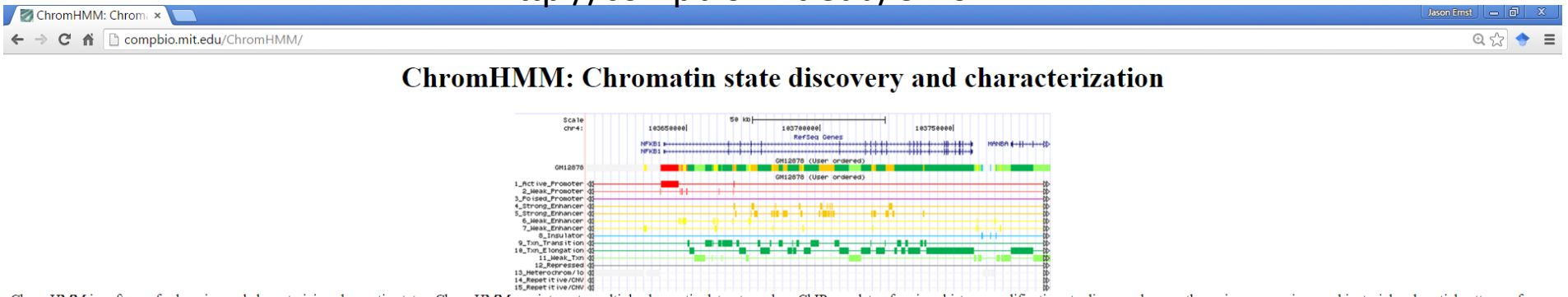


Talk Outline

- Chromatin states analysis and ChromHMM
- Accessing chromatin state annotations for ENCODE2 and Roadmap Epigenomics
- Running the ChromHMM software

ChromHMM Website

http://compbio.mit.edu/ChromHMM



ChromHMM: Chromatin state discovery and characterization

ChromHMM is software for learning and characterizing chromatin states. ChromHMM can integrate multiple chromatin datasets such as ChIP-seq data of various histone modifications to discover de novo the major re-occuring combinatorial and spatial patterns of marks. ChromHMM is based on a multivariate Hidden Markov Model that explicitly models the presence or absence of each chromatin mark. The resulting model can then be used to systematically annotate a genome in one or more cell types. By automatically computing state enrichments for large-scale functional and annotation datasets ChromHMM facilitates the biological characterization of each state. ChromHMM also produces files with genome-wide maps of chromatin state annotations that can be directly visualized in a genome browser.

- ChromHMM software v1.11 (version log)
- ChromHMM manual

Software download

Quick instructions on running ChromHMM:
1. Install Java 1.5 or later if not already installed.
2. Unzip the file ChromHMM.zip
3. To try out ChromHMM learning a 10-state model on the sample data enter from a command line in the directory with the ChromHMM.jar file the command:

java -mx1600M -jar ChromHMM.jar LearnModel SAMPLEDATA_HG18 OUTPUTSAMPLE 10 hg18

After termination in ~5-10 minutes a file in OUTPUTSAMPLE/webpage_10.html will be created showing output images and linking to all the output files created. If a web browser is found on the computer the webpage will automatically be opened in it.
In general binarized input for the *LearnModel* command can be generated by first running the *BinarizeBed* command on bed files with coordinates of aligned reads or the *BinarizeBam* command on bam files with the coordinates of aligned reads.
New in version 1.11: ChromHMM has a BinarizeBam command which allows binarizing bam files of aligned reads.
New in version 1.10: ChromHMM has the option for parallel training with multiprocessors leading to significantly reduced training times. Add the '-p 0' option to the LearnModel command to have ChromHMM to try to use as many processors as available or specify the maximum it should use.

- The ChromHMM software is described in:

Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. *Nature Methods*, 9:215-216, 2012.

- Here are links to some existing ChromHMM annotations in hg19 available for 127 Reference Epigenomes (Roadmap Epigenomics), 9-ENCODE cell types (from Ernst et al. Nature 2011), and 6-ENCODE cell types (from ENCODE Integrative Analysis).
- Contact Jason Ernst (jason.ernst at ucla dot edu) with any questions, comments, or bug reports.
- Subscribe to a mailing list for announcements of new versions
- ChromHMM is released under a GPL 3 license.
- ChromHMM source code is available on GitHub here.
- Funding for ChromHMM provided by NSF Postdoctoral Fellowship 0905968 to JE and grants from the National Institutes of Health (NIH 1-RC1-HG005334 and NIH 1 U54 HG004570).

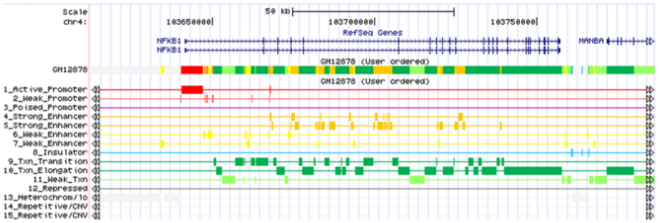
ChromHMM Website

http://compbio.mit.edu/ChromHMM

ChromHMM: Chrom ×

compbio.mit.edu/ChromHMM/

ChromHMM: Chromatin state discovery and characterization



ChromHMM is software for learning and characterizing chromatin states. ChromHMM can integrate multiple chromatin datasets such as ChIP-seq data of various histone modifications to discover de novo the major re-occuring combinatorial and spatial patterns of marks. ChromHMM is based on a multivariate Hidden Markov Model that explicitly models the presence or absence of each chromatin mark. The resulting model can then be used to systematically annotate a genome in one or more cell types. By automatically computing state enrichments for large-scale functional and annotation datasets ChromHMM facilitates the biological characterization of each state. ChromHMM also produces files with genome-wide maps of chromatin state annotations that can be directly visualized in a genome browser.

- ChromHMM software v1.11 (version log)
- ChromHMM manual

Software manual

Quick instructions on running ChromHMM:
1. Install Java 1.5 or later if not already installed.
2. Unzip the file ChromHMM.zip
3. To try out ChromHMM learning a 10-state model on the sample data enter from a command line in the directory with the ChromHMM.jar file the command:

java -mx1600M -jar ChromHMM.jar LearnModel SAMPLEDATA_HG18 OUTPUTSAMPLE 10 hg18

After termination in ~5-10 minutes a file in OUTPUTSAMPLE/webpage_10.html will be created showing output images and linking to all the output files created. If a web browser is found on the computer the webpage will automatically be opened in it.
In general binarized input for the *LearnModel* command can be generated by first running the *BinarizeBed* command on bed files with coordinates of aligned reads or the *BinarizeBam* command on bam files with the coordinates of aligned reads.
New in version 1.11: ChromHMM has a BinarizeBam command which allows binarizing bam files of aligned reads.
New in version 1.10: ChromHMM has the option for parallel training with multiprocessors leading to significantly reduced training times. Add the '-p 0' option to the LearnModel command to have ChromHMM to try to use as many processors as available or specify the maximum it should use.

- The ChromHMM software is described in:

Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. *Nature Methods*, 9:215-216, 2012.

- Here are links to some existing ChromHMM annotations in hg19 available for 127 Reference Epigenomes (Roadmap Epigenomics), 9-ENCODE cell types (from Ernst et al. Nature 2011), and 6-ENCODE cell types (from ENCODE Integrative Analysis).
- Contact Jason Ernst (jason.ernst at ucla dot edu) with any questions, comments, or bug reports.
- Subscribe to a mailing list for announcements of new versions
- ChromHMM is released under a GPL 3 license.
- ChromHMM source code is available on GitHub here.
- Funding for ChromHMM provided by NSF Postdoctoral Fellowship 0905968 to JE and grants from the National Institutes of Health (NIH 1-RC1-HG005334 and NIH 1 U54 HG004570).

Try to Run ChromHMM on Sample Data on Your Computer

(Java needs to already be installed)

1. Download

http://compbio.mit.edu/ChromHMM/ChromHMM.zip

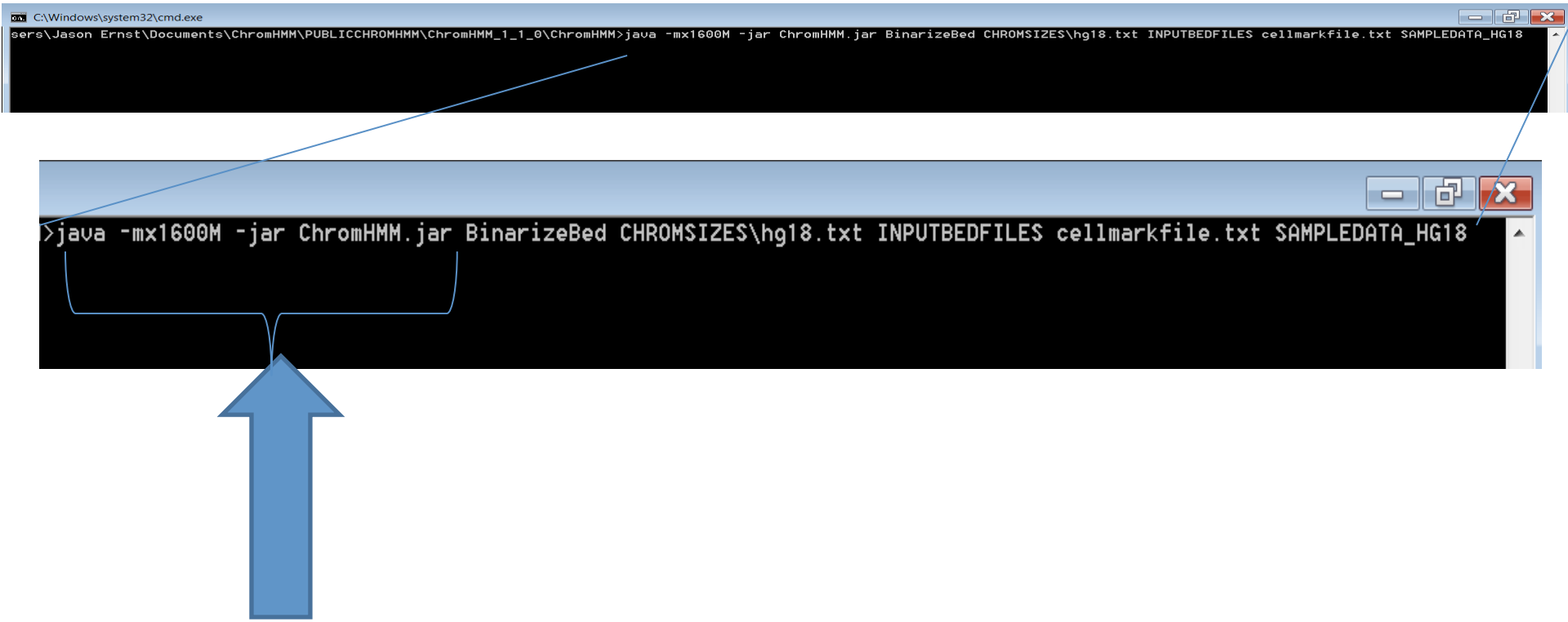
2. Unzip ChromHMM.zip
3. Open a command line
4. Change into the ChromHMM directory
5. Enter the command:

```
java -mx1600M -jar ChromHMM.jar LearnModel -p 0 SAMPLEDATA_HG18 OUTPUTSAMPLE 10 hg18
```

Input to ChromHMM

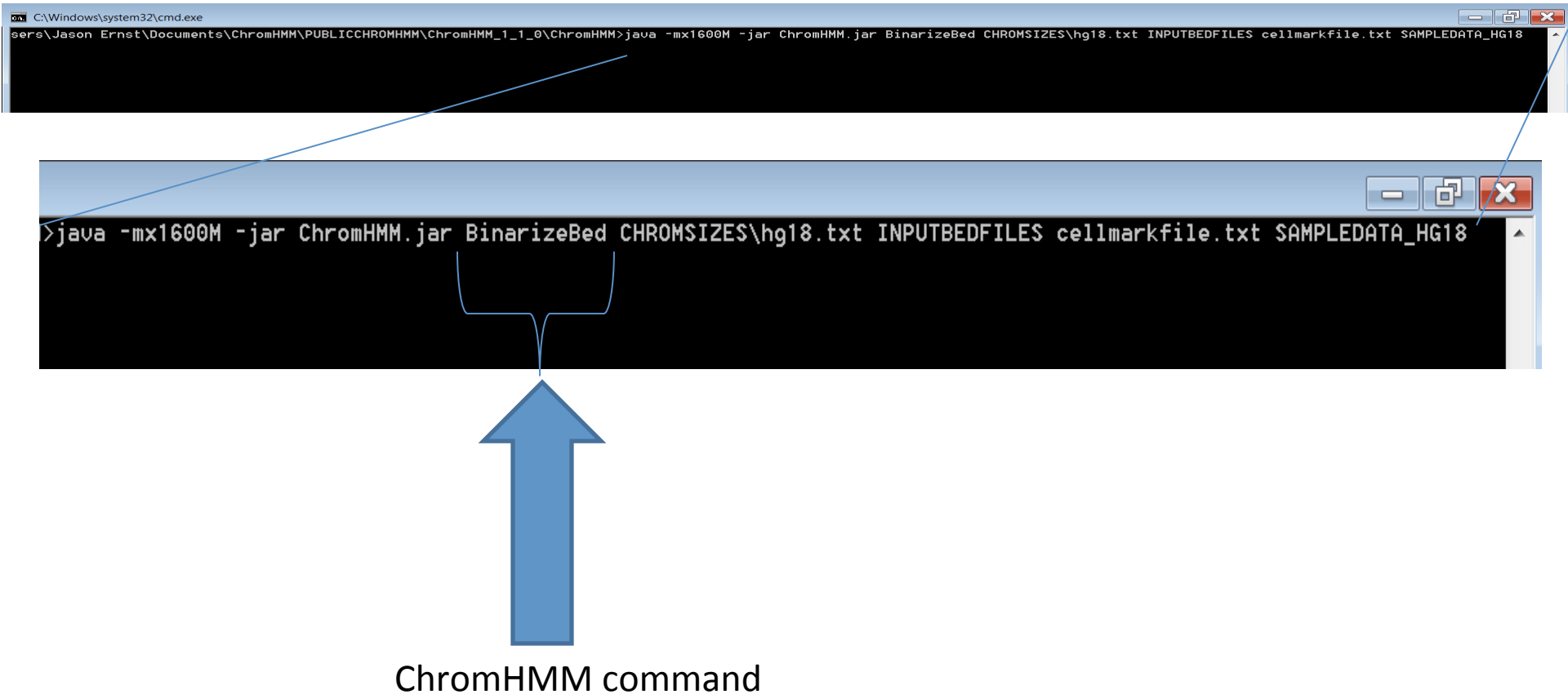
- ChromHMM models are learned from binarized data using its `LearnModel` command
- Binarized data is typically obtained starting from aligned reads.
  - Apply `BinarizeBed` if reads are in BED format
  - Apply `BinarizeBam` if reads are in BAM format

BinarizeBed

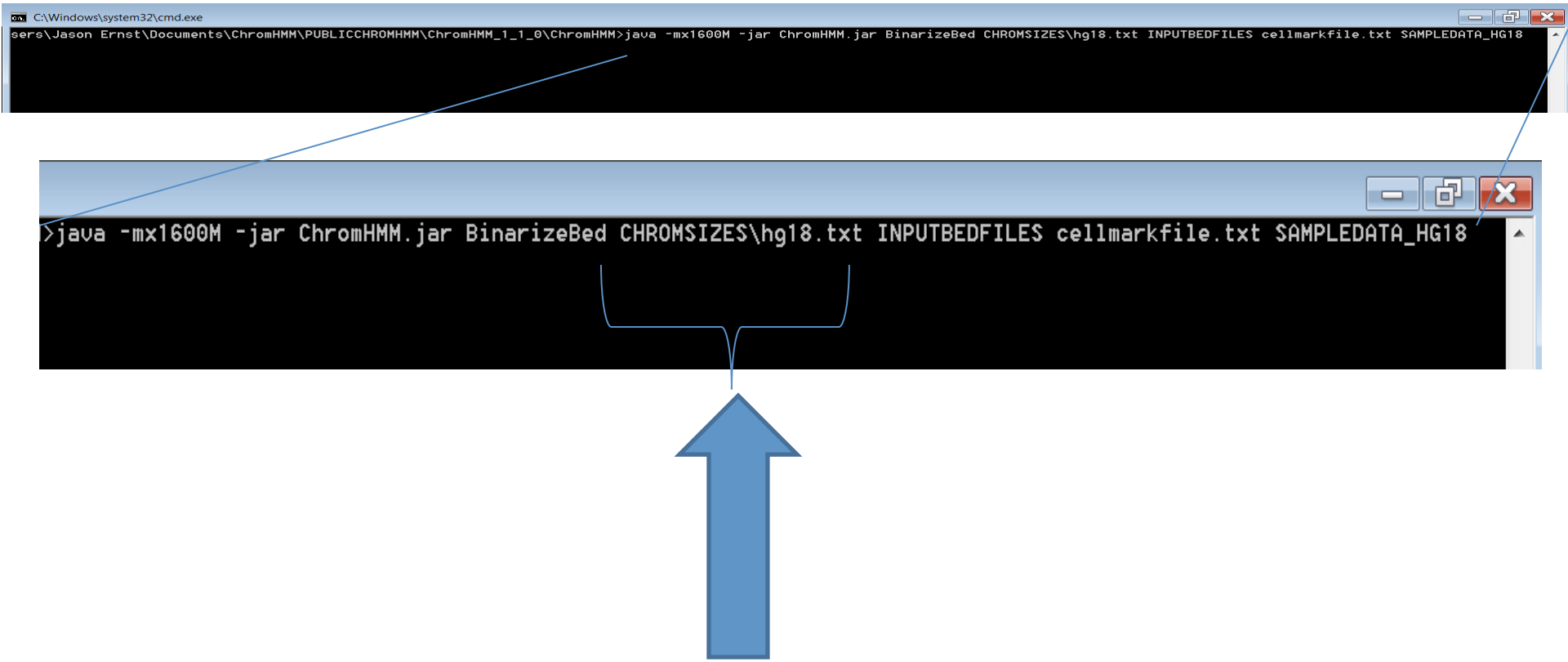


Java command '-mx1600M' specifies memory to Java

BinarizeBed

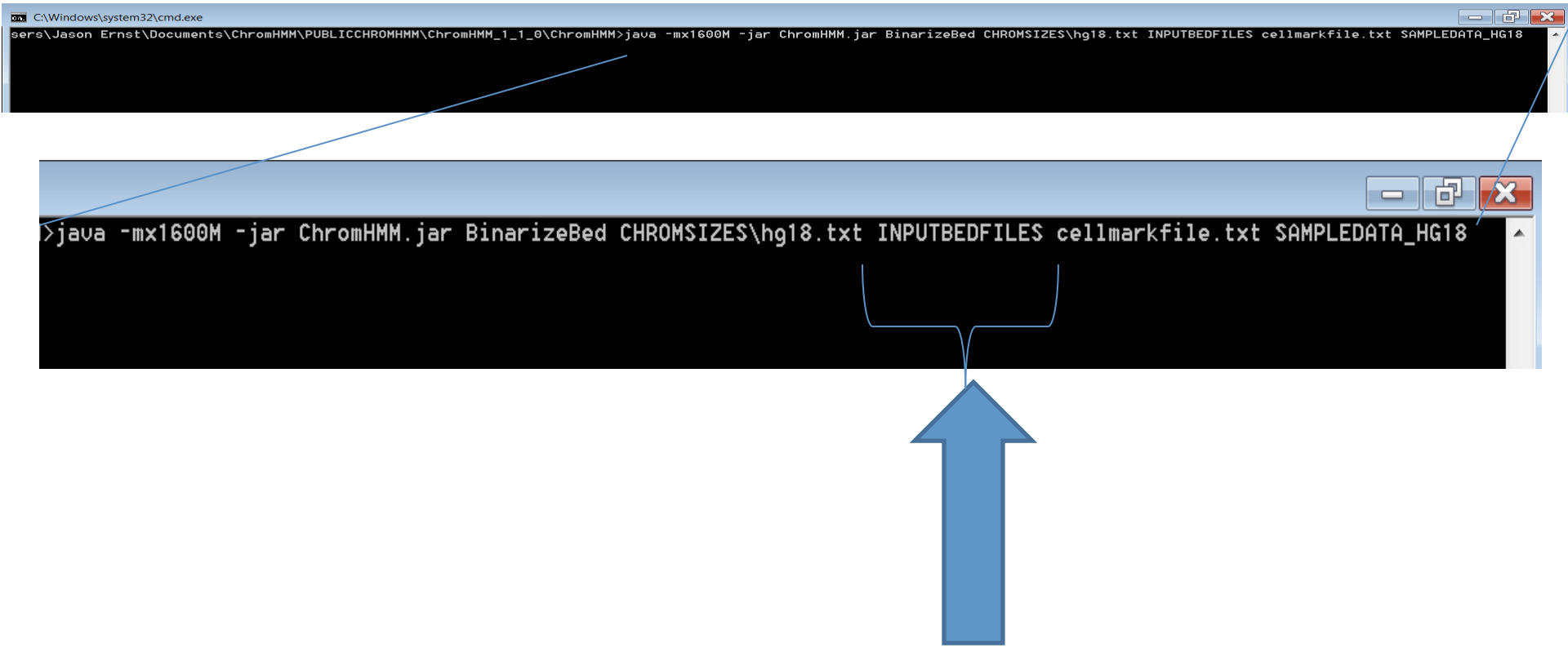


BinarizeBed



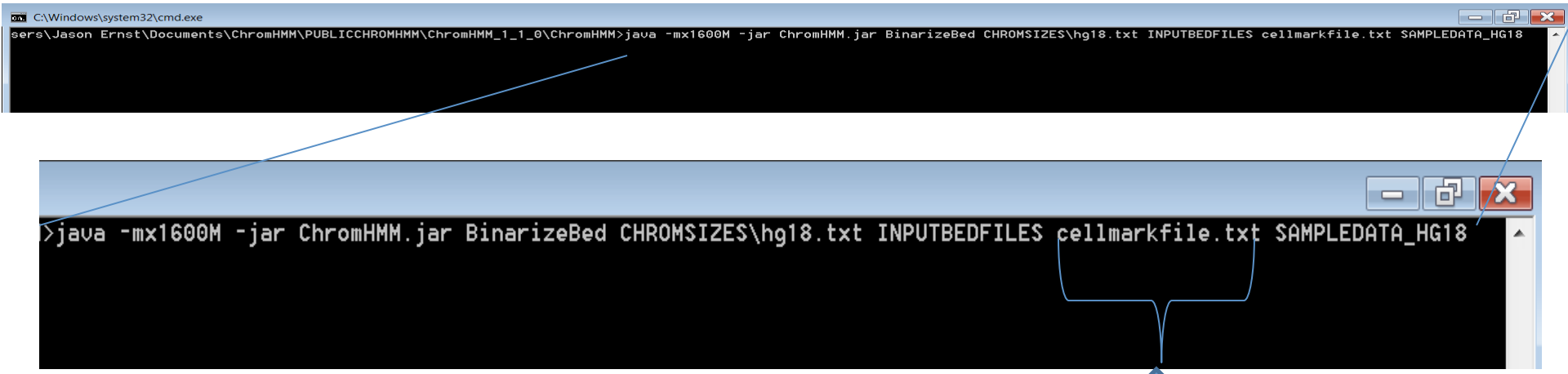
File with the chromosome lengths for the assembly

BinarizeBed



DIRECTORY of BED files

BinarizeBed



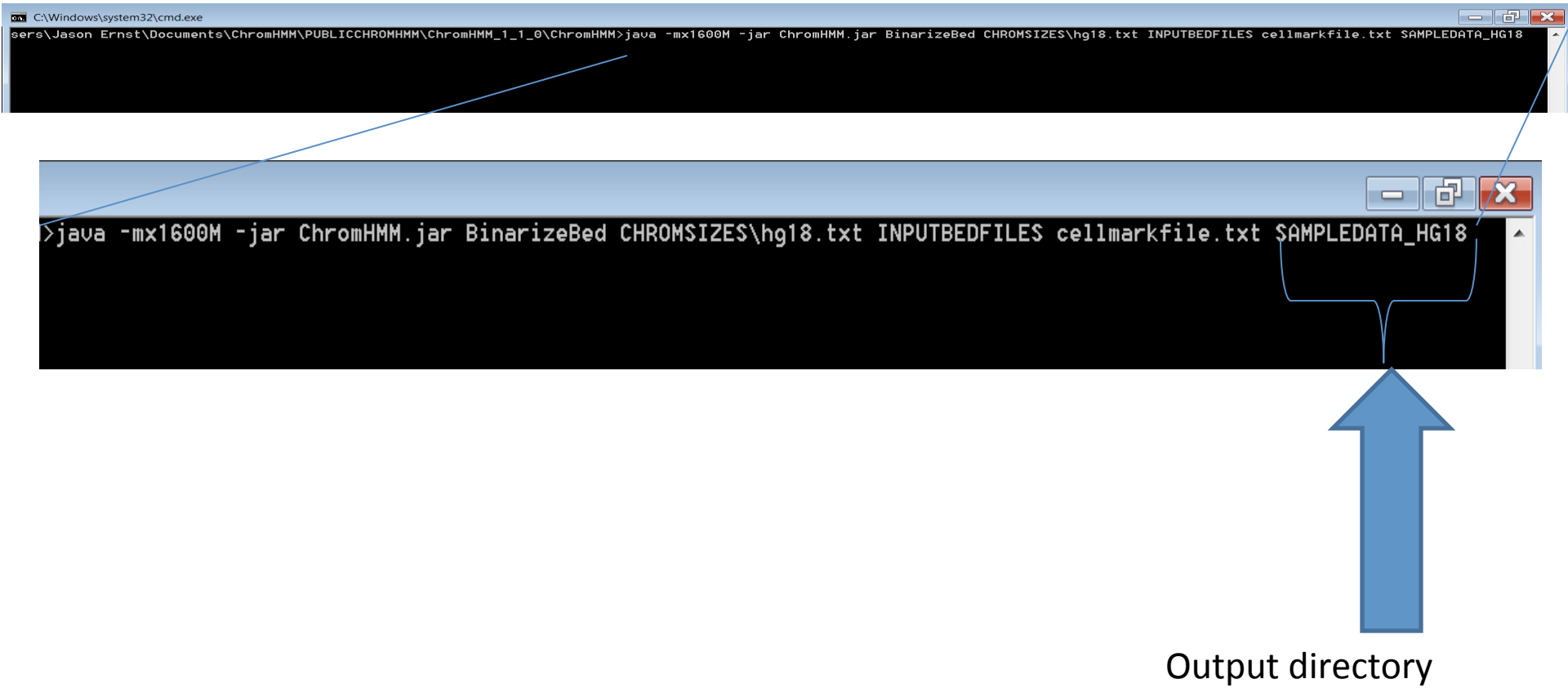

```
cell1 mark1 cell1_mark1.bed   cell1_control.bed
cell1 mark2 cell1_mark2.bed   cell1_control.bed
cell2 mark1 cell2_mark1.bed   cell2_control.bed
cell2 mark2 cell2_mark2.bed   cell2_control.bed
```

cell mark cell-mark

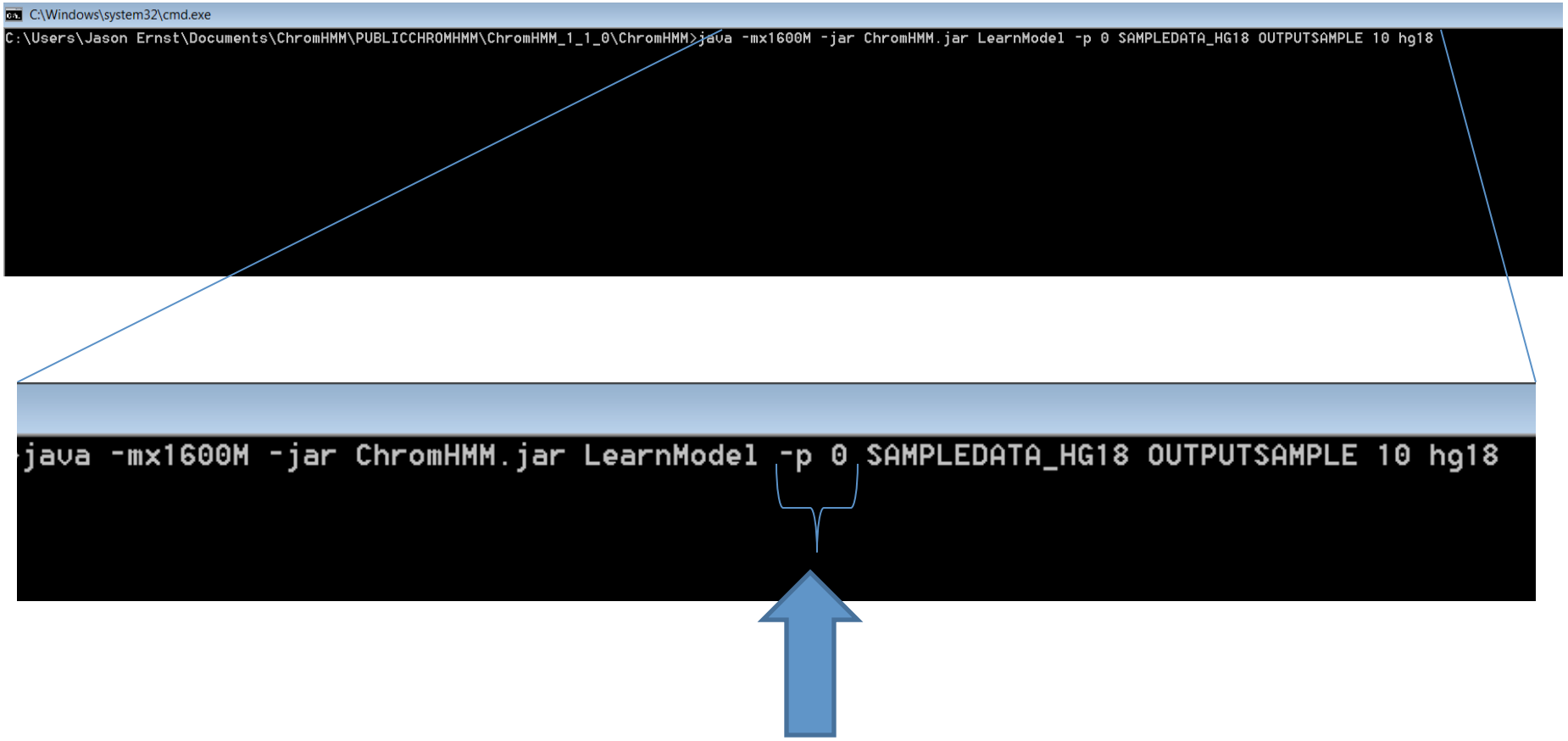
Cell-mark –file table

Control data – is optional and can also be treated as a mark

BinarizeBed

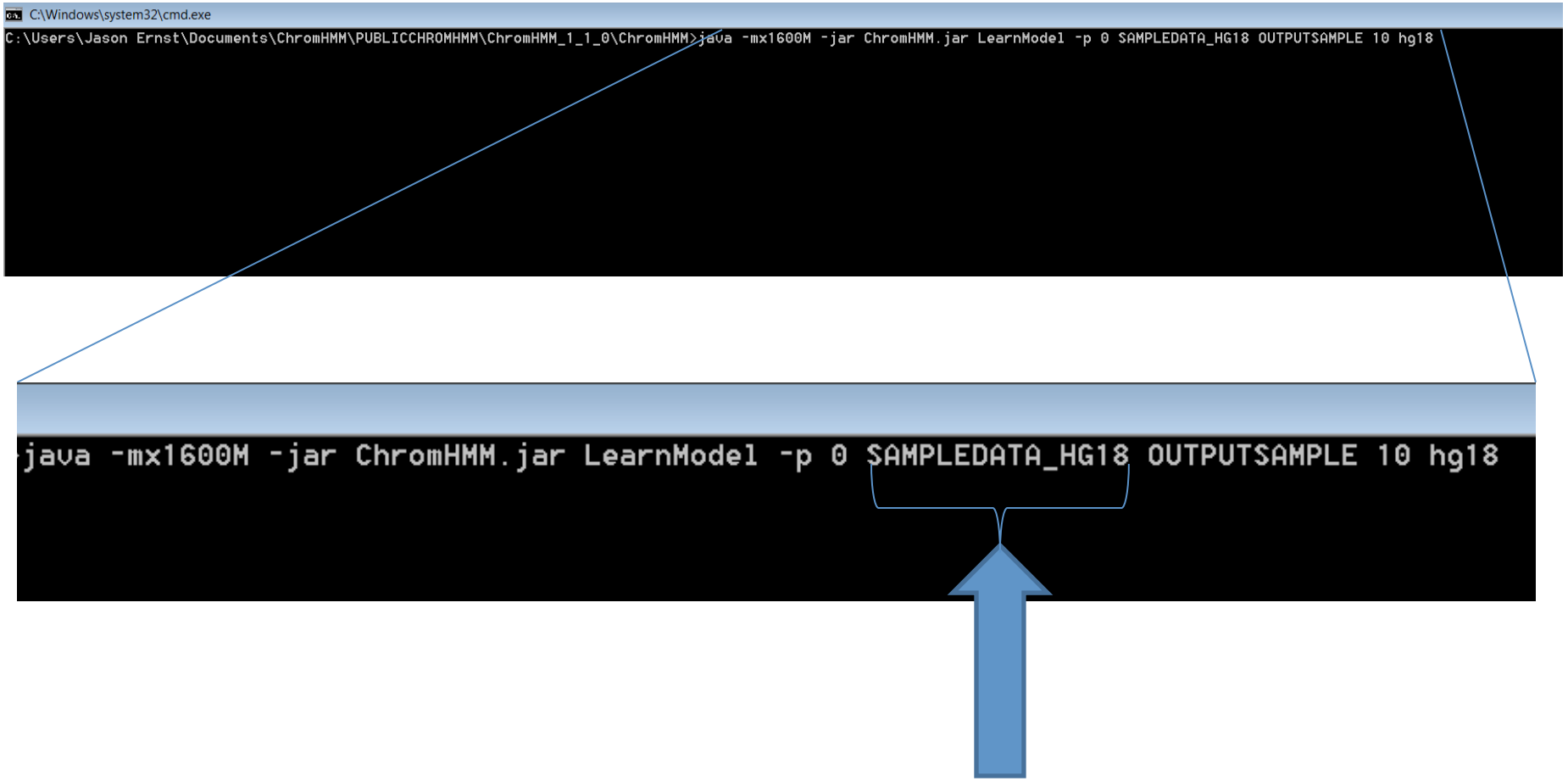


LearnModel



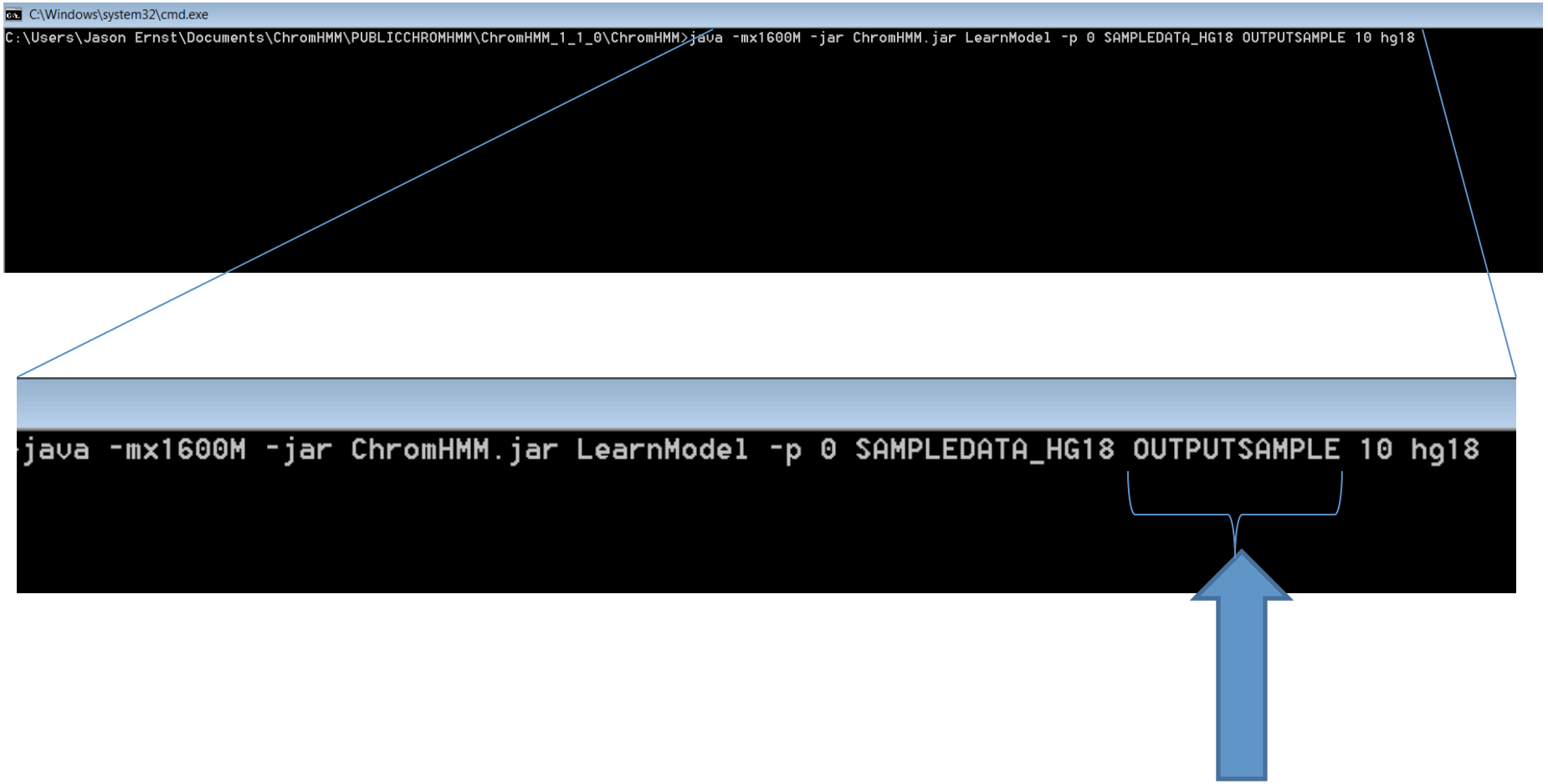
‘-p 0’ Use as many processors as available
‘-p N’ Use up to N processors (default N=1)

LearnModel

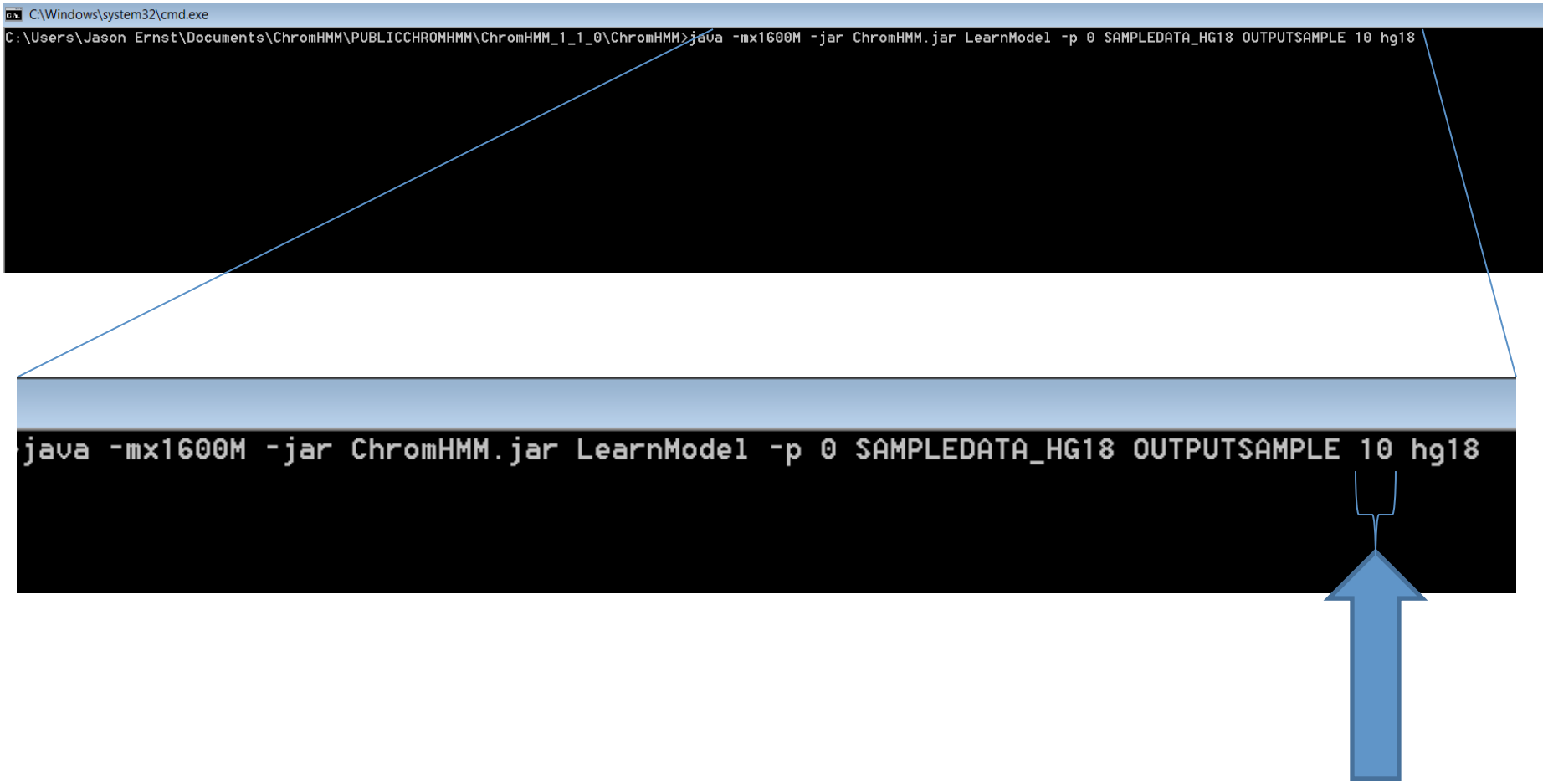


Directory with the Binarized Input

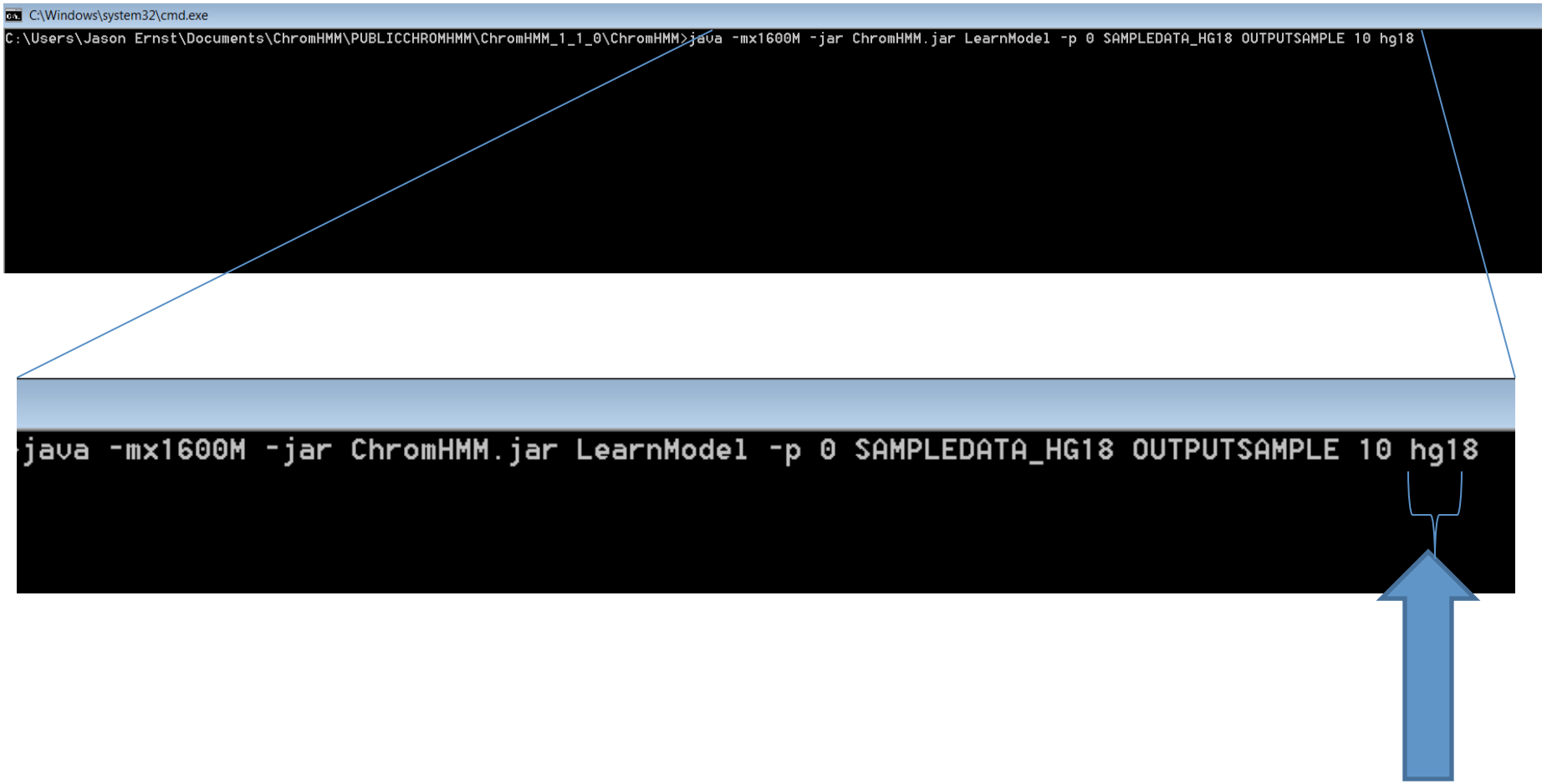
LearnModel


C:\Windows\system32\cmd.exe
C:\Users\Jason Ernst\Documents\ChromHMM\PUBLICCHROMHMM\ChromHMM_1_1_0\ChromHMM>java -mx1600M -jar ChromHMM.jar LearnModel -p 0 SAMPLEDATA_HG18 OUTPUTSAMPLE 10 hg18
java -mx1600M -jar ChromHMM.jar LearnModel -p 0 SAMPLEDATA_HG18 OUTPUTSAMPLE 10 hg18
Directory where the output goes

LearnModel


C:\Windows\system32\cmd.exe
C:\Users\Jason Ernst\Documents\ChromHMM\PUBLICCHROMHMM\ChromHMM_1_1_0\ChromHMM>java -mx1600M -jar ChromHMM.jar LearnModel -p 0 SAMPLEDATA_HG18 OUTPUTSAMPLE 10 hg18
java -mx1600M -jar ChromHMM.jar LearnModel -p 0 SAMPLEDATA_HG18 OUTPUTSAMPLE 10 hg18
Number of states

LearnModel


C:\Windows\system32\cmd.exe
C:\Users\Jason Ernst\Documents\ChromHMM\PUBLICCHROMHMM\ChromHMM_1_1_0\ChromHMM>java -mx1600M -jar ChromHMM.jar LearnModel -p 0 SAMPLEDATA_HG18 OUTPUTSAMPLE 10 hg18
java -mx1600M -jar ChromHMM.jar LearnModel -p 0 SAMPLEDATA_HG18 OUTPUTSAMPLE 10 hg18
Genome assembly

ChromHMM Report

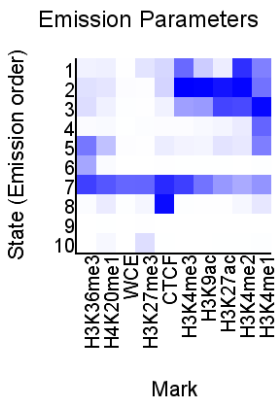
webpage_10.html

file:///C:/Users/Jason%20Ernst/Documents/ChromHMM/PUBLICCHROMHMM/ChromHMM_1_1_0/ChromHMM/OUTPUTSAMPLE/webpage_10.html

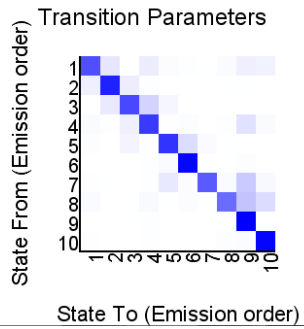
ChromHMM Report

Input Directory: SAMPLEDATA_HG18
Output Directory: OUTPUTSAMPLE
Number of States: 10
Assembly: hg18
Full ChromHMM command: LearnModel -p 0 SAMPLEDATA_HG18 OUTPUTSAMPLE 10 hg18

Model Parameters



- Emission Parameter SVG File
- Emission Parameter Tab-Delimited Text File



ChromHMM Report

webpage_10.html

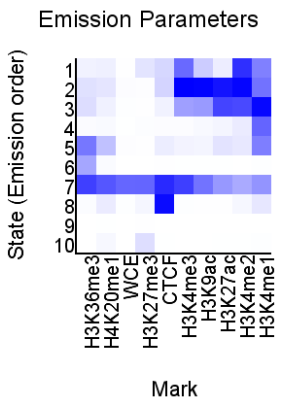
file:///C:/Users/Jason%20Ernst/Documents/ChromHMM/PUBLICCHROMHMM/ChromHMM_1_1_0/ChromHMM/OUTPUTSAMPLE/webpage_10.html

ChromHMM Report

Input Directory: SAMPLEDATA_HG18
Output Directory: OUTPUTSAMPLE
Number of States: 10
Assembly: hg18
Full ChromHMM command: LearnModel -p 0 SAMPLEDATA_HG18 OUTPUTSAMPLE 10 hg18

Model Parameters

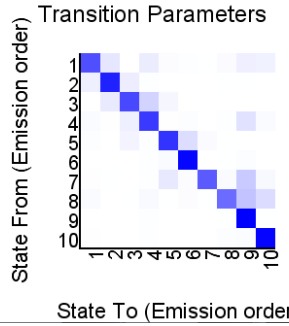
Emission Parameters



Input Directory: SAMPLEDATA_HG18
Output Directory: OUTPUTSAMPLE
Number of States: 10
Assembly: hg18
Full ChromHMM command: LearnModel -p 0 SAMPLEDATA_HG18 OUTPUTSAMPLE 10 hg18

- Emission Parameter SVG File
- Emission Parameter Tab-Delimited Text File

Transition Parameters



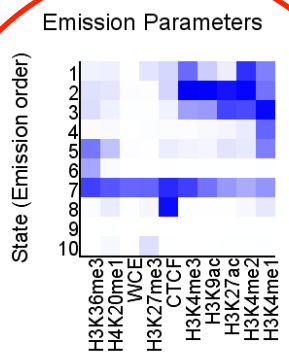
Emission Parameters

webpage_10.html

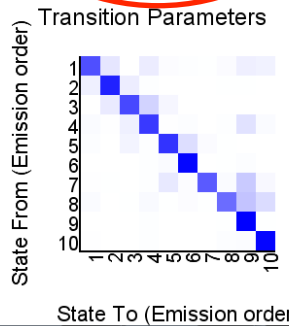
file:///C:/Users/Jason%20Ernst/Documents/ChromHMM/PUBLICCHROMHMM/ChromHMM_1_1_0/ChromHMM/OUTPUTSAMPLE/webpage_10.html

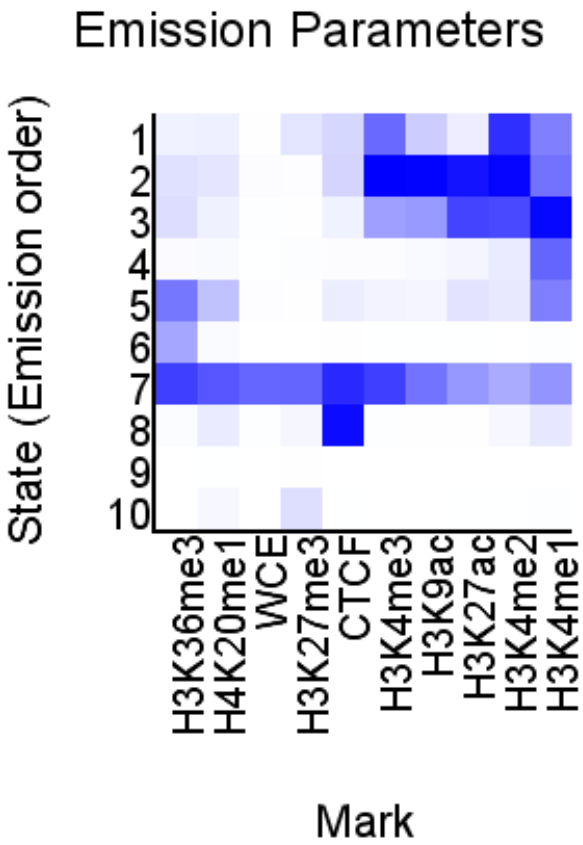
Input Directory: SAMPLEDATA_HG18
Output Directory: OUTPUTSAMPLE
Number of States: 10
Assembly: hg18
Full ChromHMM command: LearnModel -p 0 SAMPLEDATA_HG18 OUTPUTSAMPLE 10 hg18

Model Parameters



- Emission Parameter SVG File
- Emission Parameter Tab-Delimited Text File





- Emission Parameter SVG File
- Emission Parameter Tab-Delimited Text File

Transition Parameters

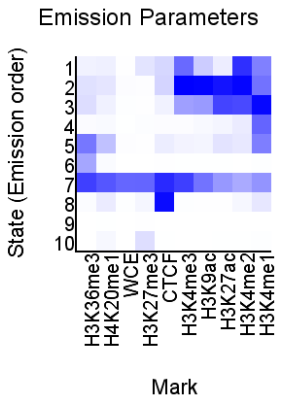
webpage_10.html

file:///C:/Users/Jason%20Ernst/Documents/ChromHMM/PUBLICCHROMHMM/ChromHMM_1_1_0/ChromHMM/OUTPUTSAMPLE/webpage_10.html

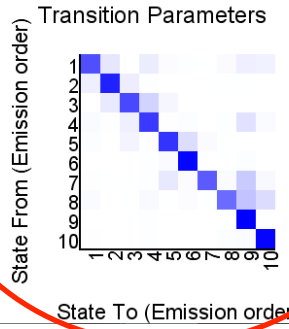
ChromHMM Report

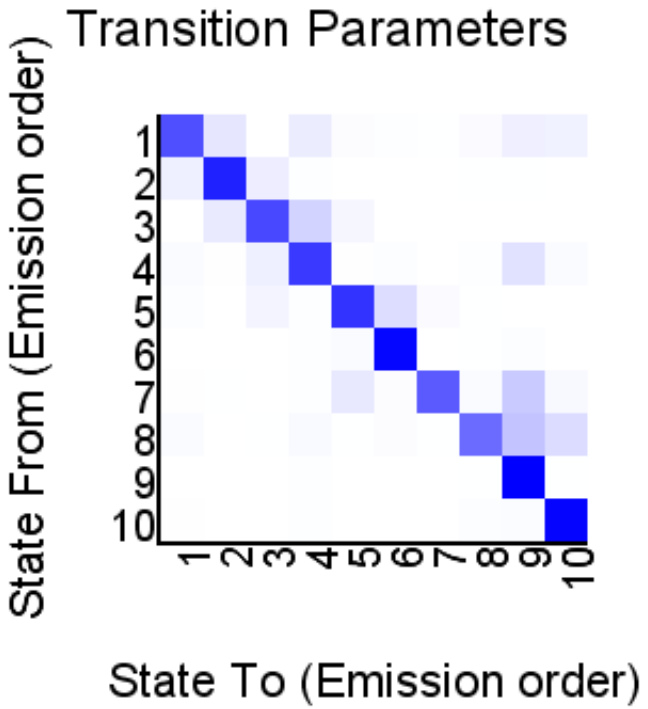
Input Directory: SAMPLEDATA_HG18
Output Directory: OUTPUTSAMPLE
Number of States: 10
Assembly: hg18
Full ChromHMM command: LearnModel -p 0 SAMPLEDATA_HG18 OUTPUTSAMPLE 10 hg18

Model Parameters



- Emission Parameter SVG File
- Emission Parameter Tab-Delimited Text File





- Transition Parameter SVG File
- Transition Parameter Tab-Delimited Text File

Model Parameter File

webpage_10.html

file:///C:/Users/Jason%20Ernst/Documents/ChromHMM/PUBLICCHROMHMM/ChromHMM_1_1_0/ChromHMM/OUTPUTSAMPLE/webpage_10.html

- Transition Parameter SVG File
- Transition Parameter Tab-Delimited Text File

- All Model Parameters Tab-Delimited Text File

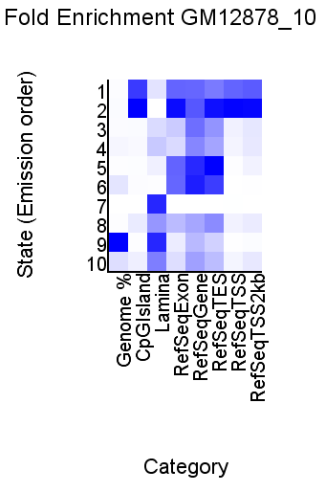
Genome Segmentation Files

- GM12878_10 Segmentation File (Four Column Bed File)
- K562_10 Segmentation File (Four Column Bed File)

Custom Tracks for loading into the UCSC Genome Browser:
- GM12878_10 Browser Custom Track Dense File
- GM12878_10 Browser Custom Track Expanded File
- K562_10 Browser Custom Track Dense File
- K562_10 Browser Custom Track Expanded File

State Enrichments

GM12878_10 Enrichments



- GM12878_10 Overlap Enrichment SVG File
- GM12878_10 Overlap Enrichment Tab-Delimited Text File

Segmentation File

webpage_10.html

file:///C:/Users/Jason%20Ernst/Documents/ChromHMM/PUBLICCHROMHMM/ChromHMM_1_1_0/ChromHMM/OUTPUTSAMPLE/webpage_10.html

- Transition Parameter SVG File
- Transition Parameter Tab-Delimited Text File

- All Model Parameters Tab-Delimited Text File

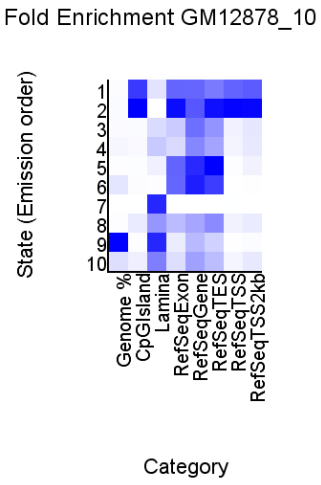
Genome Segmentation Files

- GM12878_10 Segmentation File (Four Column Bed File)
- K562_10 Segmentation File (Four Column Bed File)

Custom Tracks for loading into the UCSC Genome Browser:
- GM12878_10 Browser Custom Track Dense File
- GM12878_10 Browser Custom Track Expanded File
- K562_10 Browser Custom Track Dense File
- K562_10 Browser Custom Track Expanded File

State Enrichments

GM12878_10 Enrichments



- GM12878_10 Overlap Enrichment SVG File
- GM12878_10 Overlap Enrichment Tab-Delimited Text File

GM12878_10_segments.bed - Notepad

File Edit Format View Help

chr11	0	58000	E9
chr11	58000	58400	E8
chr11	58400	61400	E9
chr11	61400	61800	E8
chr11	61800	66800	E9
chr11	66800	67600	E4
chr11	67600	75600	E9
chr11	75600	76000	E8
chr11	76000	87600	E9
chr11	87600	88600	E10
chr11	88600	116200	E9
chr11	116200	116400	E1
chr11	116400	116600	E10
chr11	116600	117000	E8
chr11	117000	120600	E9
chr11	120600	121000	E8
chr11	121000	165400	E9
chr11	165400	166000	E7
chr11	166000	170000	E10
chr11	170000	171000	E8
chr11	171000	173200	E9
chr11	173200	173600	E7
chr11	173600	174000	E9
chr11	174000	174400	E8
chr11	174400	177000	E9
chr11	177000	177600	E4
chr11	177600	178800	E3
chr11	178800	179800	E5
chr11	179800	180600	E3
chr11	180600	181000	E2
chr11	181000	181600	E1
chr11	181600	182200	E2
chr11	182200	182800	E1
chr11	182800	183000	E4
chr11	183000	184600	E10
chr11	184600	184800	E7
chr11	184800	186600	E10
chr11	186600	189800	E9
chr11	189800	193000	E6
chr11	193000	195600	E5
chr11	195600	195800	E3
chr11	195800	200400	E2
chr11	200400	201200	E3
chr11	201200	201800	E5
chr11	201800	205800	E6
chr11	205800	210600	E9
chr11	210600	215400	E6
chr11	215400	215600	E4
chr11	215600	216400	E2
chr11	216400	224200	E9
chr11	224200	224400	E4
chr11	224400	225200	E9
chr11	225200	225600	E1
chr11	225600	228200	E2
chr11	228200	228800	E1

Browser Files

webpage_10.html

file:///C:/Users/Jason%20Ernst/Documents/ChromHMM/PUBLICCHROMHMM/ChromHMM_1_1_0/ChromHMM/OUTPUTSAMPLE/webpage_10.html

- Transition Parameter SVG File
- Transition Parameter Tab-Delimited Text File

- All Model Parameters Tab-Delimited Text File

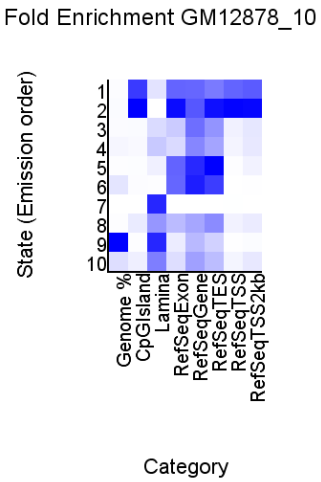
Genome Segmentation Files

- GM12878_10 Segmentation File (Four Column Bed File)
- K562_10 Segmentation File (Four Column Bed File)

Custom Tracks for loading into the UCSC Genome Browser:
- GM12878_10 Browser Custom Track Dense File
- GM12878_10 Browser Custom Track Expanded File
- K562_10 Browser Custom Track Dense File
- K562_10 Browser Custom Track Expanded File

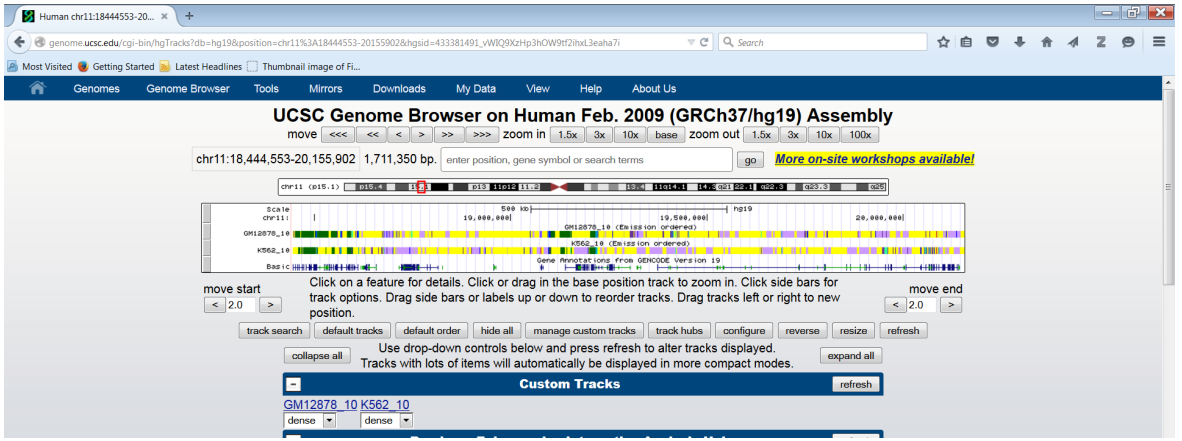
State Enrichments

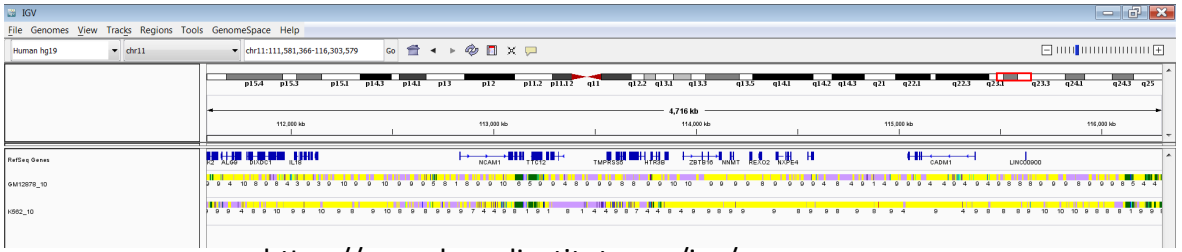
GM12878_10 Enrichments



- GM12878_10 Overlap Enrichment SVG File
- GM12878_10 Overlap Enrichment Tab-Delimited Text File

Can load into browser UCSC Genome, IGV





https://www.broadinstitute.org/igv/

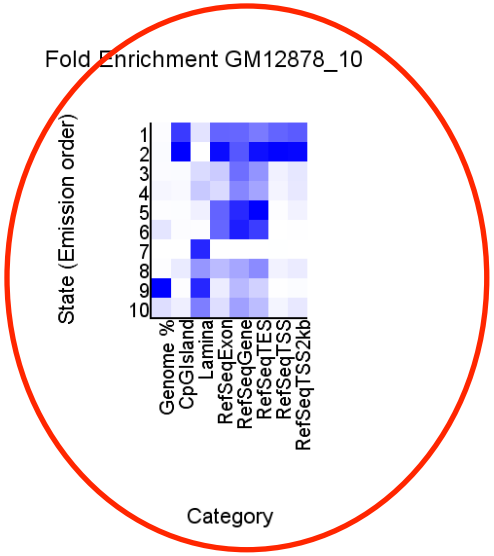
Enrichments

webpage_10.html

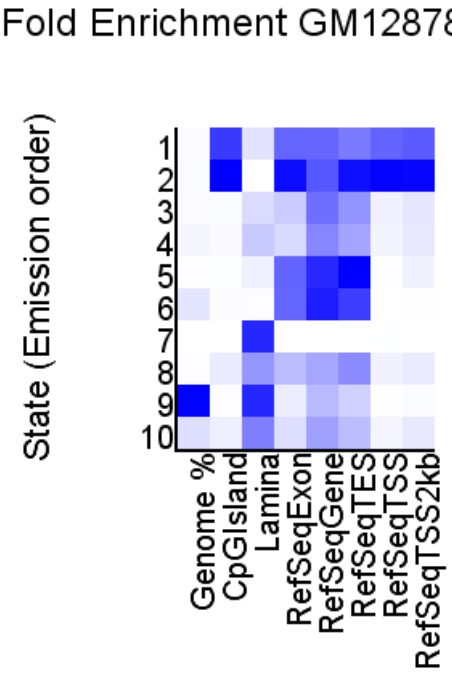
file:///C:/Users/Jason%20Ernst/Documents/ChromHMM/PUBLICCHROMHMM/ChromHMM_1_1_0/ChromHMM/OUTPUTSAMPLE/webpage_10.html

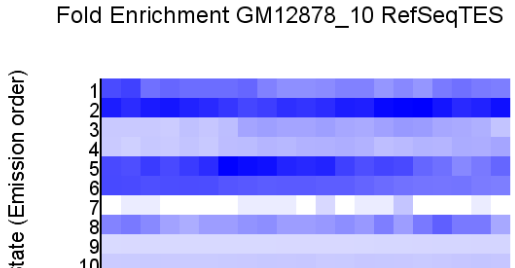
State Enrichments

GM12878_10 Enrichments



- GM12878_10 Overlap Enrichment SVG File
- GM12878_10 Overlap Enrichment Tab-Delimited Text File





Category

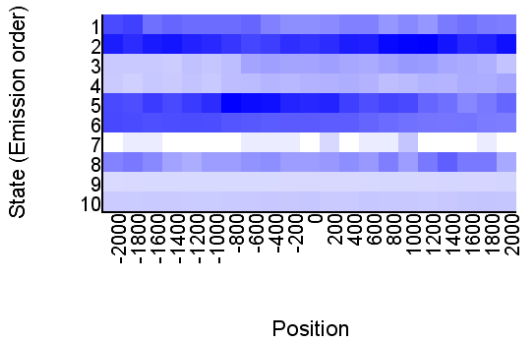
- GM12878_10 Overlap Enrichment SVG File
- GM12878_10 Overlap Enrichment Tab-Delimited Text File

Positional Plots

webpage_10.html

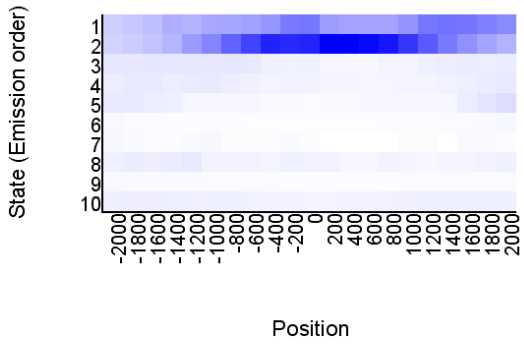
file:///C:/Users/Jason%20Ernst/Documents/ChromHMM/PUBLICCHROMHMM/ChromHMM_1_1_0/ChromHMM/OUTPUTSAMPLE/webpage_10.html

Fold Enrichment GM12878_10 RefSeqTES



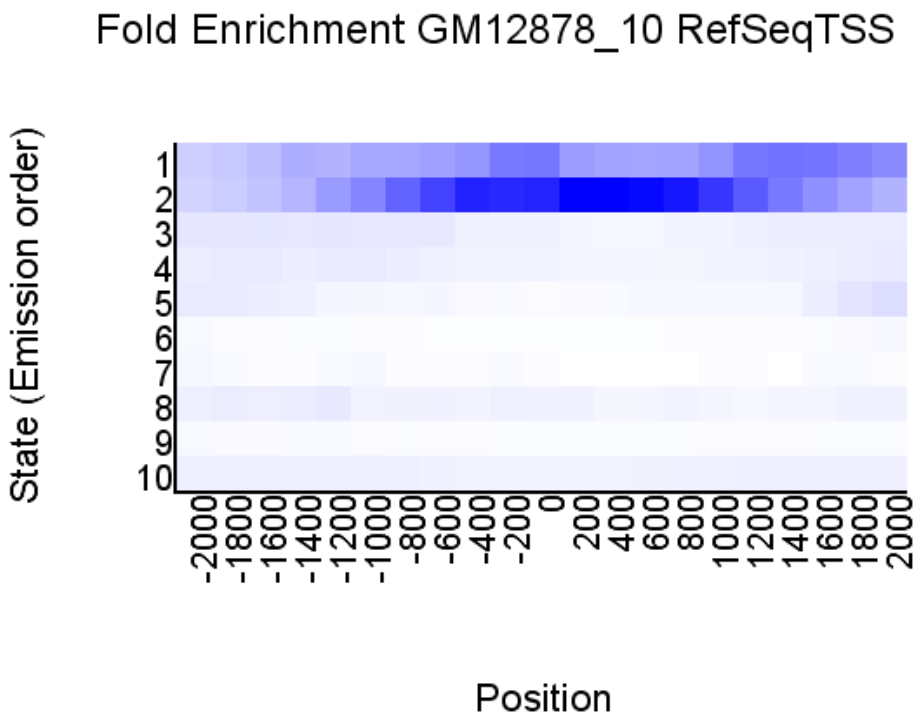
- GM12878_10_RefSeqTES_neighborhood Enrichment SVG File
- GM12878_10_RefSeqTES_neighborhood Enrichment Tab-Delimited Text File

Fold Enrichment GM12878_10 RefSeqTSS



- GM12878_10_RefSeqTSS_neighborhood Enrichment SVG File
- GM12878_10_RefSeqTSS_neighborhood Enrichment Tab-Delimited Text File

Fold Enrichment GM12878_10 RefSeqTSS



- GM12878_10_RefSeqTSS_neighborhood Enrichment SVG File
- GM12878_10_RefSeqTSS_neighborhood Enrichment Tab-Delimited Text File

Enrichments for Additional Cell Types

webpage_10.html

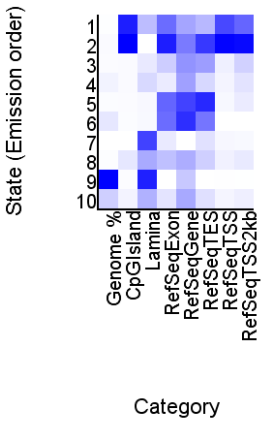
file:///C:/Users/Jason%20Ernst/Documents/ChromHMM/PUBLICCHROMHMM/ChromHMM_1_1_0/ChromHMM/OUTPUTSAMPLE/webpage_10.html

Position

- GM12878_10_RefSeqTSS_neighborhood Enrichment SVG File
- GM12878_10_RefSeqTSS_neighborhood Enrichment Tab-Delimited Text File

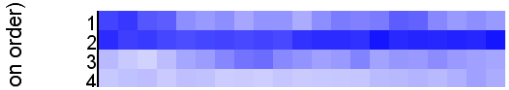
K562_10 Enrichments

Fold Enrichment K562_10

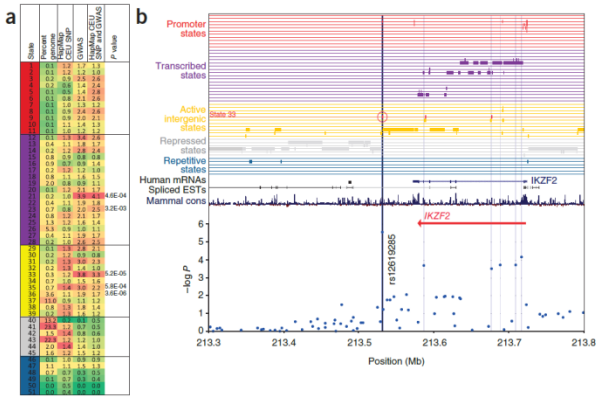


- K562_10 Overlap Enrichment SVG File
- K562_10 Overlap Enrichment Tab-Delimited Text File

Fold Enrichment K562_10 RefSeqTES

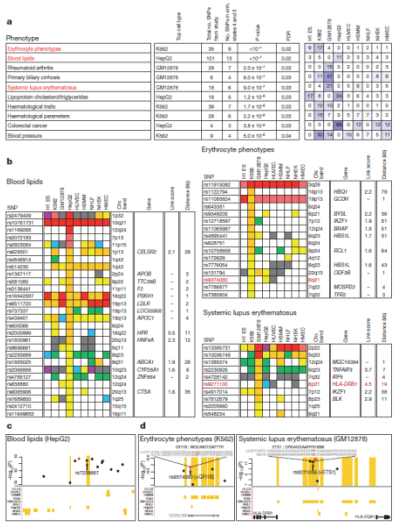


Chromatin states to interpret disease variants



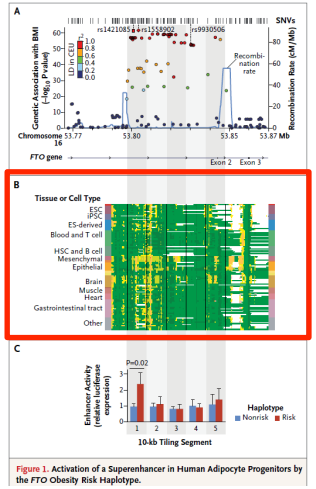
- Specific chromatin states enriched in GWAS catalog

Ernst and Kellis, *Nature Biotech* 2010



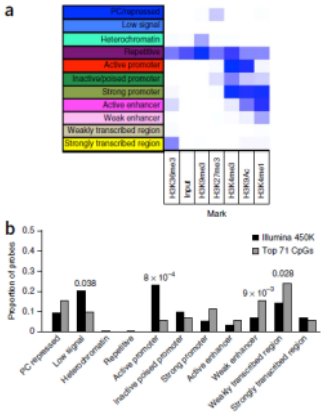
- Enhancers from different cell types enriched in different traits

Ernst et al, *Nature* 2011



- Imputation based chromatin state used in dissection FTO loci

Claussnitzer et al, *NEJM* 2015



- Interpreting epigenetic disease associated variation in Alzheimer's disease

De Jager et al, *Nature Neuroscience* 2014

- Many other examples in the literature

Collaborators and Acknowledgements

- Manolis Kellis

ENCODE consortium

  - Brad Bernstein production group

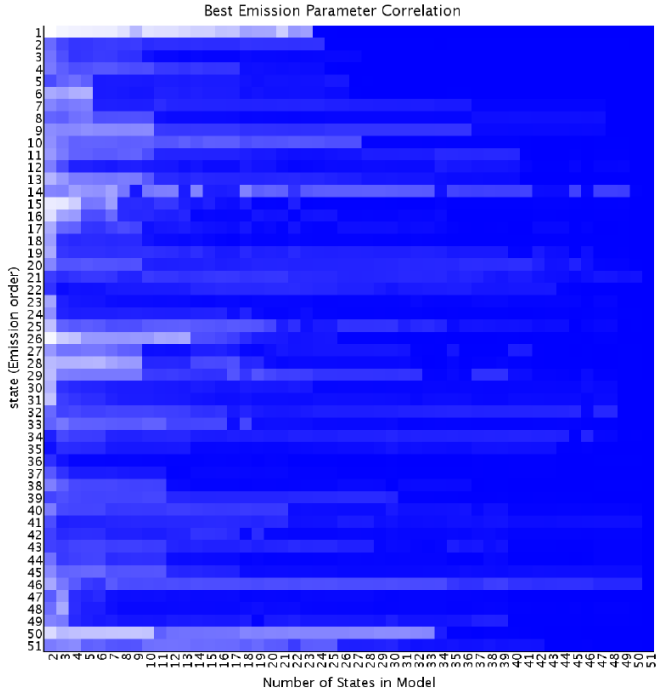
Roadmap Epigenomics consortium

Funding

- NHGRI, NIH, NSF, HHMI, Sloan Foundation

Additional Commands

- CompareModels – the command allows the comparison of the emission parameters of a selected model to a set of models in terms of correlation.

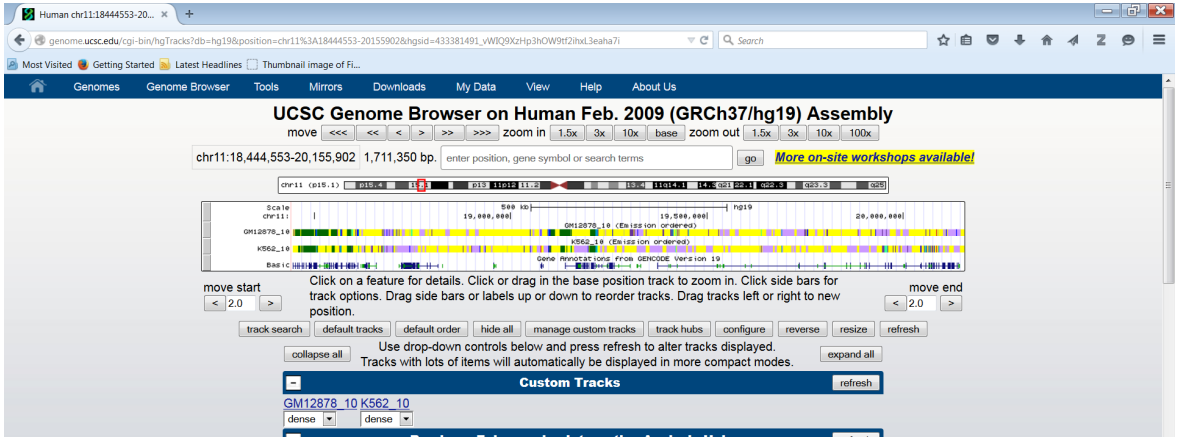



```
CompareModels [-color r,g,b] referencemodel comparedir outputprefix
```

Additional Commands

- MakeBrowserFiles – (re)generates browser files from segmentation files and allows specifying the coloring

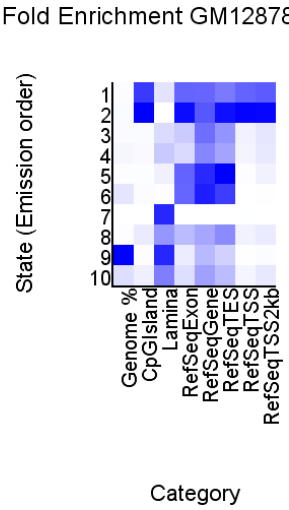
```
MakeBrowserFiles [-c colormappingfile][-m labelmappingfile][-n numstates]
segmentfile segmentationname outputfileprefix
```



Additional Commands

- OverlapEnrichment – (re)computes enrichments of a segmentation for a set of annotations

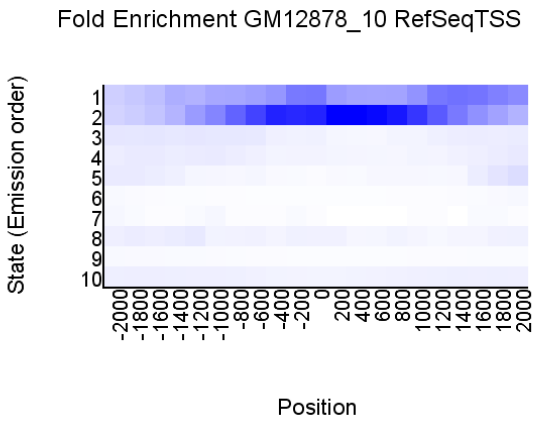
```
OverlapEnrichment [-a cell][-b binsize][-binres][-color r,g,b][-center]
[-colfields chromosome,start,end[,signal]][-e offsetend][-f coordlistfile][-m
labelmappingfile][-multicount][-posterior][-s offsetstart][-signal][-t
title][-uniformscale] inputsegment inputcoorddir outfileprefix
```



Additional Commands

- NeighborhoodEnrichment – (re)computes enrichments of a segmentation around a set of anchor positions

```
usage NeighborhoodEnrichment [-a cell][-b binsize][-color r,g,b]
[-colfields chromosome,position[,optionalcol1|,optionalcol1,optionalcol2]]
[-l numleftintervals][-m labelmappingfile][-nostrand][-o anchoroffset]
[-posterior][-r numrightintervals][-s spacing][-signal][-t title]
inputsegment anchorpositions outfileprefix
```



Additional Commands

- Reorder – reorders the states of the model

```
usage: Reorder [-color r,g,b][-f columnorderingfile][-holdcolumnorder]
[-i outfileID][-m labelmappingfile][-o stateorderingfile][-stateordering
emission|transition] inputmodel outputdir
```