

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

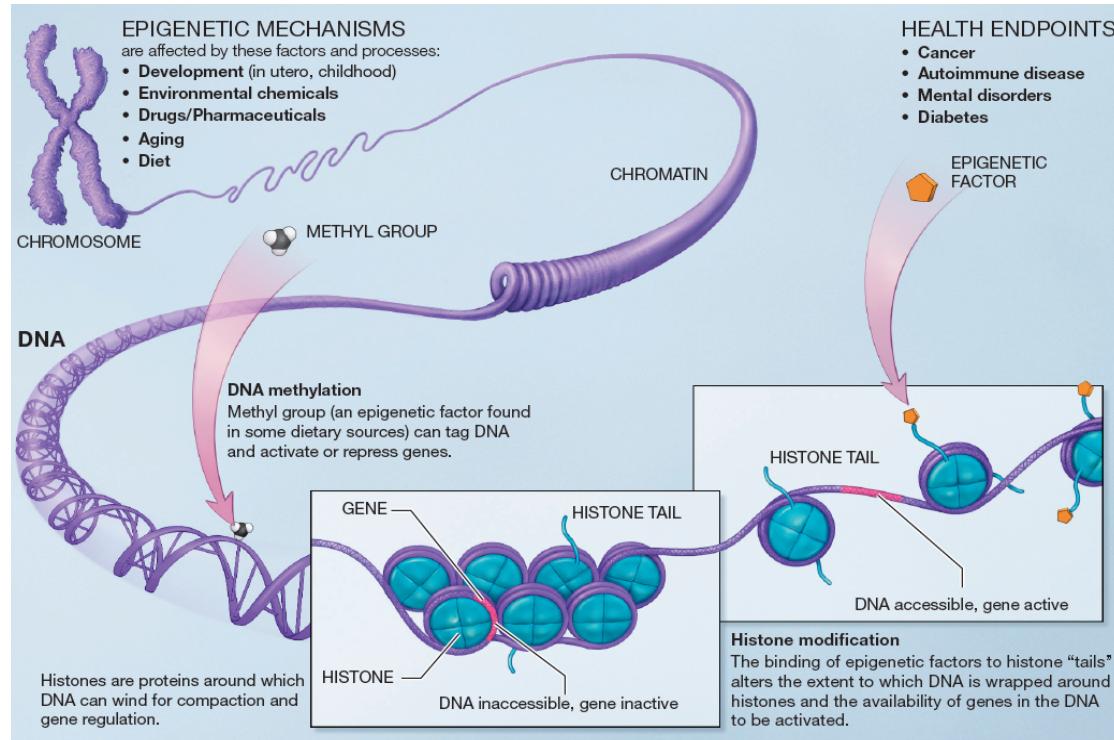
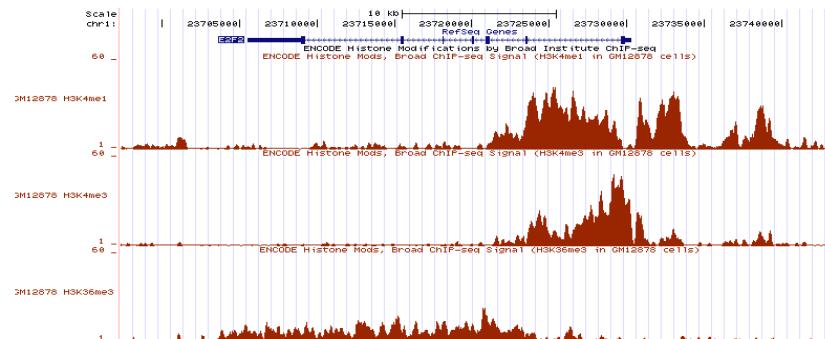


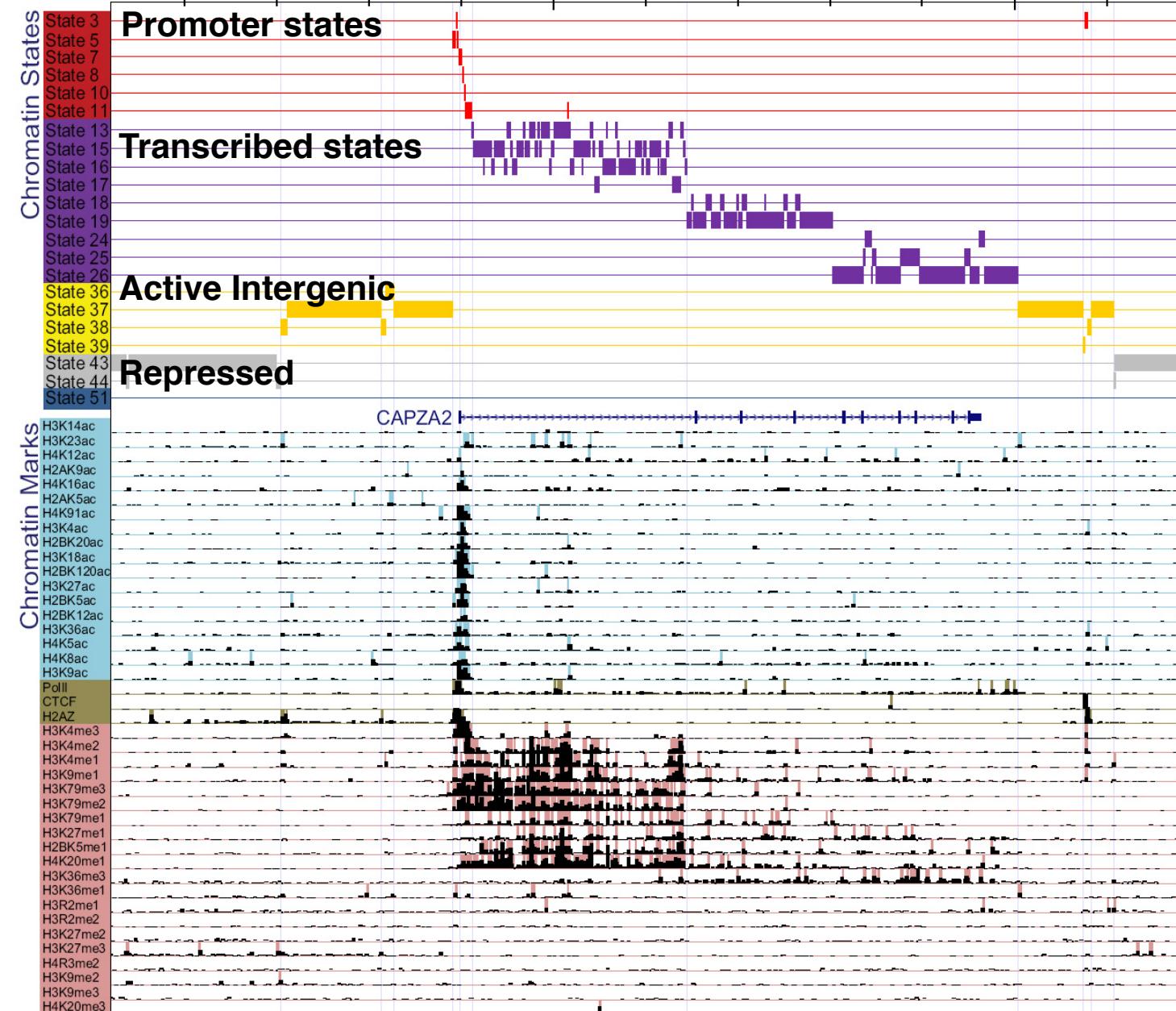
Image source: <http://nihroadmap.nih.gov/epigenomics/>

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



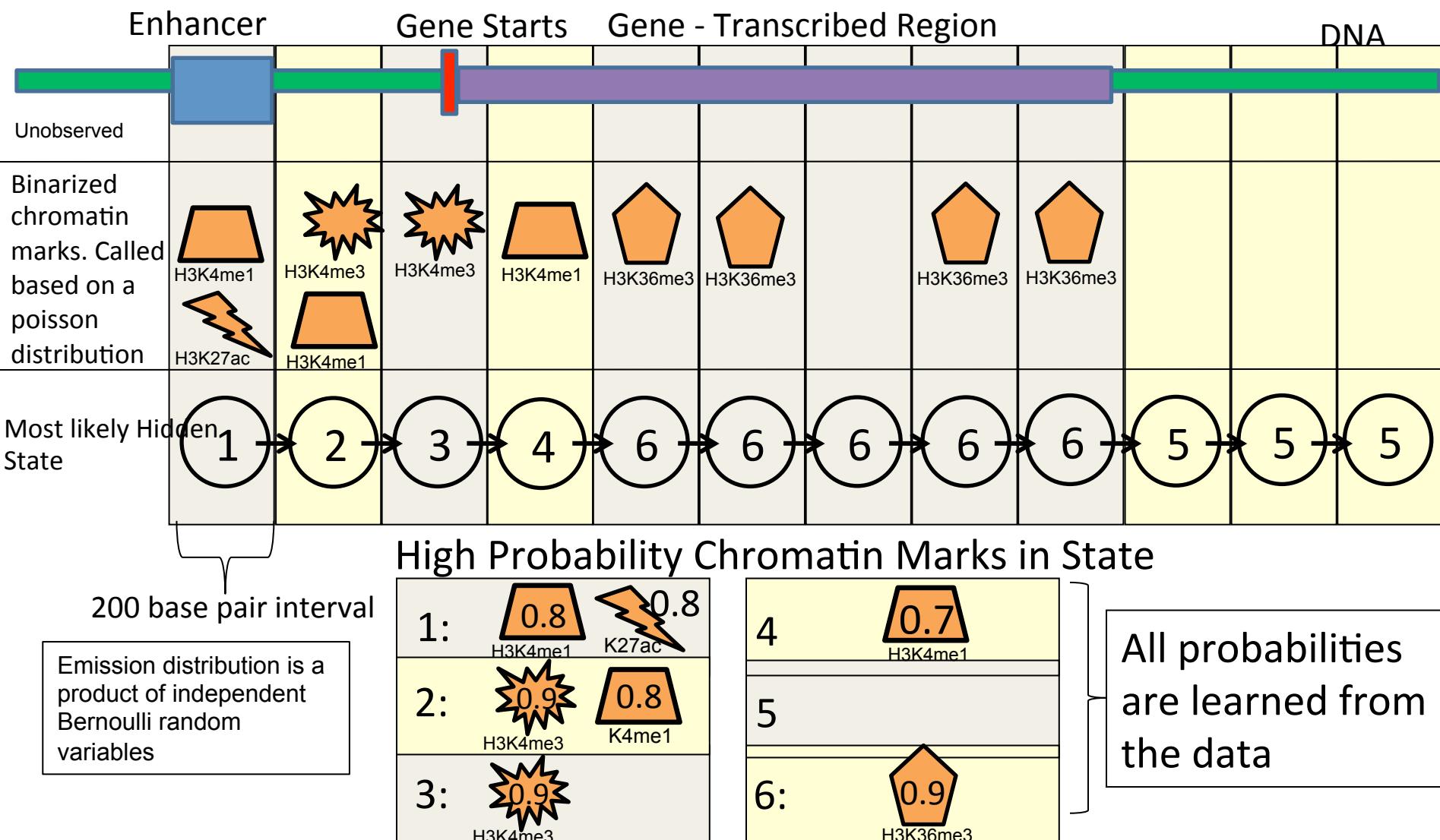
From ‘chromatin marks’ to ‘chromatin states’

Chr 7: 116,260kb 116,270kb 116,280kb 116,290kb 116,300kb 116,310kb 116,320kb 116,330kb 116,340kb 116,350kb 116,360kb



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

Our approach: Multivariate Hidden Markov Model (ChromHMM)



Binarization leads to explicit modeling of mark combinations and interpretable parameters

ENCODE: Study nine marks in nine human cell lines

9 marks

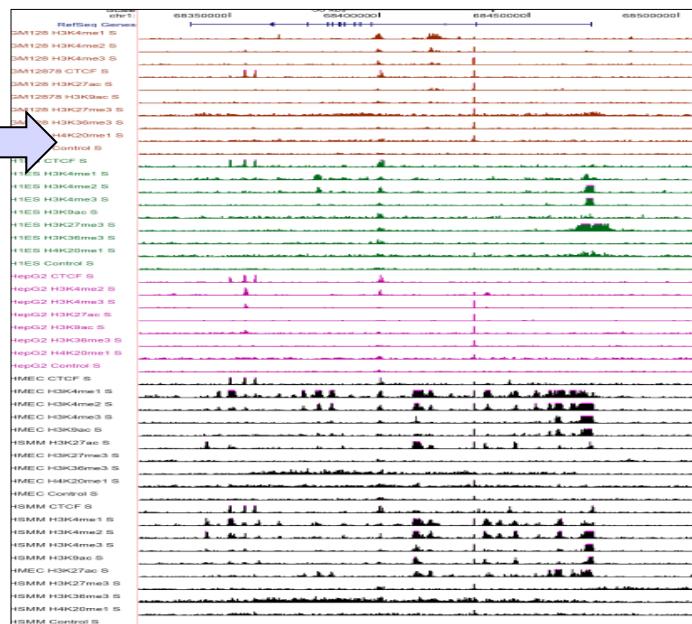
H3K4me1
H3K4me2
H3K4me3
H3K27ac
H3K9ac
H3K27me3
H4K20me1
H3K36me3
CTCF
+WCE
+RNA

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

X

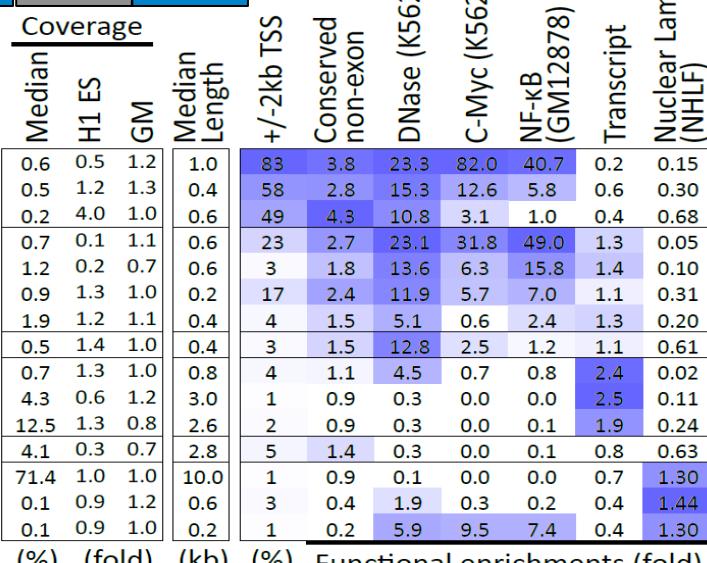
81 Chromatin Mark Tracks



Brad Bernstein ENCODE Group

HUVEC NHEK GM12878 K562 HepG2 NHLF HMEC HSMM H1

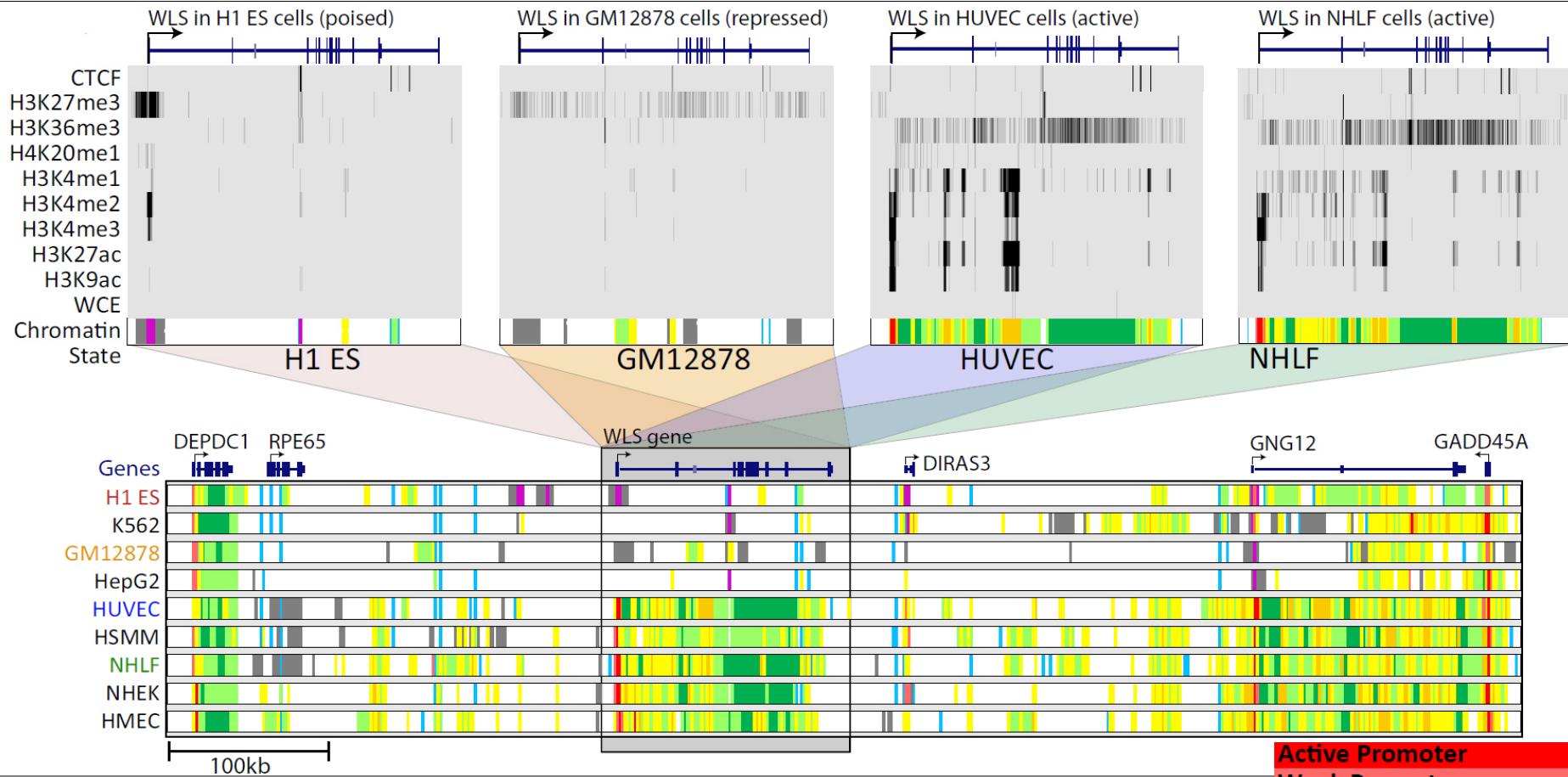
Chromatin States	State	Chromatin Mark Observation Frequency (%)									
		CTCF	H3K27me3	H3K36me3	H4K20me1	H3K4me1	H3K4me2	H3K4me3	H3K27ac	H3K9ac	WCE
1	16	2	2	6	17	93	99	96	98	2	
2	12	2	6	9	53	94	95	14	44	1	
3	13	72	0	9	48	78	49	1	10	1	
4	11	1	15	11	96	99	75	97	86	4	
5	5	0	10	3	88	57	5	84	25	1	
6	7	1	1	3	58	75	8	6	5	1	
7	2	1	2	1	56	3	0	6	2	1	
8	92	2	1	3	6	3	0	0	1	1	
9	5	0	43	43	37	11	2	9	4	1	
10	1	0	47	3	0	0	0	0	0	1	
11	0	0	3	2	0	0	0	0	0	0	
12	1	27	0	2	0	0	0	0	0	0	
13	0	0	0	0	0	0	0	0	0	0	
14	22	28	19	41	6	5	26	5	13	37	
15	85	85	91	88	76	77	91	73	85	78	



Chromatin Mark Observation Frequency (%)

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

Chromatin states dynamics across nine ENCODE cell types

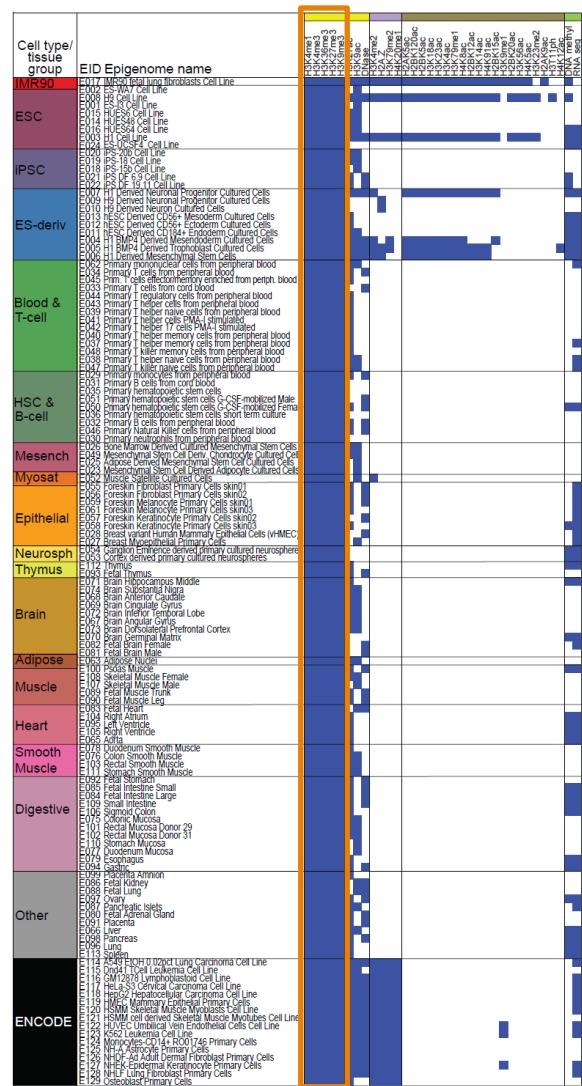


- Single annotation track for each cell type
- Summarize cell-type activity at a glance
- Can study 9-cell activity pattern across ↓

Talk Outline

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

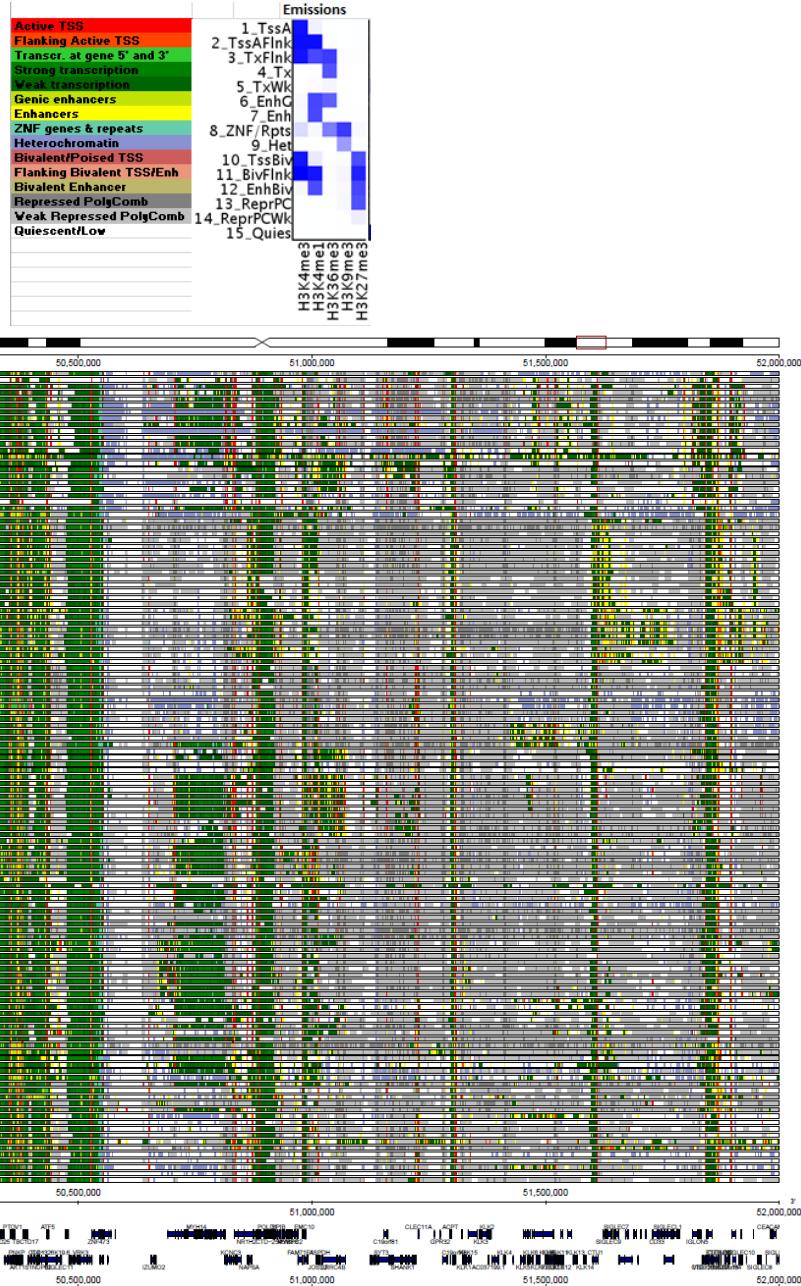
Chromatin States Defined Across 127 Cell/Tissues Types



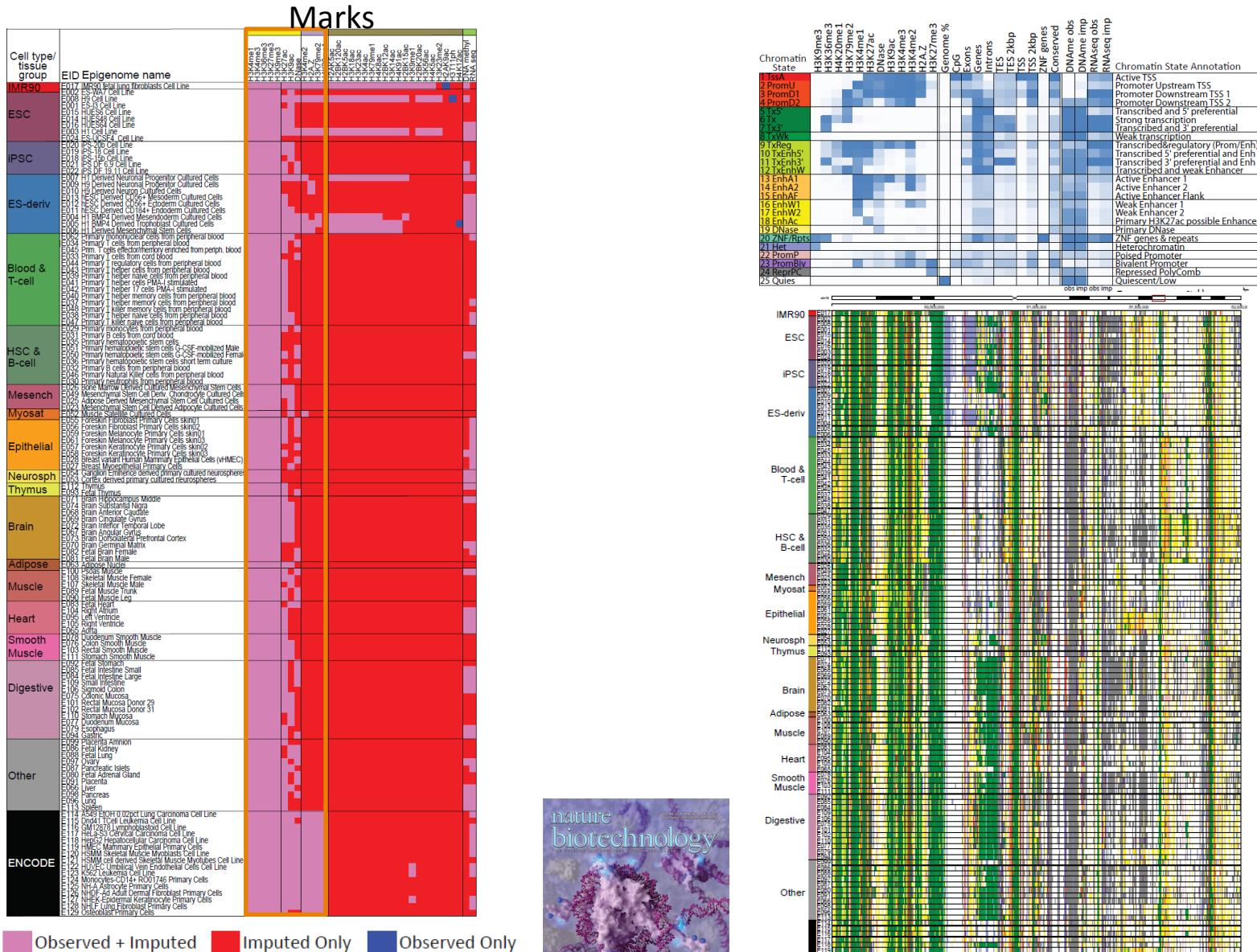
Observed

16 epigenomes from ENCODE 2

Roadmap Epigenomics Consortium et al, *Nature* 2015



Chromatin States Defined on Imputed Data



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_enancers)
- 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_enancer1)
- State 8 - GreenYellow - EnhG2 (Genic_enancer2)
- 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	H3K9ac
H3K4me3	H4K20me1
H3K27me3	H3K79me2
H3K9me3	H3K4me2
H3K36me3	H2A.Z
H3K27ac	DNase

Roadmap Epigenomics Integrative Analysis Portal

<http://compbio.mit.edu/roadmap>

The screenshot shows the main navigation bar of the portal. The items in the bar are: Home / Grid, Metadata, Processed Data ▾, Imputed Data ▾, Chromatin State Learning ▾ (which is highlighted with a red circle), Epigenomes Clustering ▾, DNase-accessible Regulatory Regions ▾, Predicting Regulators and Motifs ▾, and Disease Variant Interpretation ▾.

The NIH Roadmap Epigenomics Mapping Consortium was I resource of human epigenomic data to catalyze basic biology and disease-oriented research. The project has generated high-quality, genome-wide maps of several key histone modifications, chromatin accessibility, DNA methylation and mRNA expression across 100s of human cell types and tissues. This web portal serves as a supplementary data repository accompanying the flagship consortium paper titled *Integrative Analysis of 111 reference human epigenomes* (*Nature*, Feb. 2015). We provide uniformly processed datasets, integrative analysis products and interactive genome browser sessions resulting from a joint analysis of 111 consolidated epigenomes from the Roadmap Epigenomics Project and 16 epigenomes from *The Encyclopedia of DNA Elements (ENCODE)* project.

Release 9 of the compendium contains uniformly mapped datasets corresponding to each of the epigenomic data types spanning 183 biological samples. We refer to these as **unconsolidated epigenomes** since there often exist multiple samples (technical and biological replicates from multiple individuals and/or datasets from multiple centers) from a particular unique cell type or tissue. In order to reduce redundancy, improve data quality and achieve uniformity required for our integrative analyses, experiments were subjected to additional processing to obtain comprehensive data for 111 **consolidated epigenomes** (See Processed Data section for additional details). Numeric epigenome identifiers EIDs (e.g. E001) and mnemonics for epigenome names were assigned for each of the consolidated epigenomes. The **metadata** section summarizes the mapping of the unconsolidated Release 9 samples to the consolidated epigenome IDs and provides key metadata terms and quality control statistics. Datasets corresponding to 16 epigenomes from the ENCODE project (with epigenome IDs ranging from E114-E129) were also processed similarly and used in the integrative analyses, thus giving us a total of 127 **consolidated epigenomes**.

GRID VISUALIZATION

Select *Initialize Grid Visualization* to obtain a grid of uniformly processed data sets (columns) across all consolidated and/or unconsolidated epigenomes (rows).

Select data views (signal tracks, peak calls, read alignments)

Select grid-cells

Visualize in the epigenome browser

Instructions:

Epigenomes are ordered by groups (you can expand/hide groups by clicking +/-).

You can select different views of the data (signal tracks, peak calls, alignments)

Input track height for each data view in the text box.

EIDs correspond to consolidated epigenomes

If you unselect *Ignore unconsolidated epigenome data*, unconsolidated epigenomes will appear in the rows in blue text.

NOTE: We do not recommend visualizing the consolidated datasets alongside the unconsolidated ones. Preferably visualize the consolidated datasets.

To select grid-cells (Selected grid-cells appear blue-grey)

- Click on a single grid-cell, or
- Click and drag across rows and columns of the grid to select a sub-matrix of the multiple grid-cells.
- Click on row or column header to select entire row or column
- You can perform each of these actions multiple times to select multiple disjoint grid-cells, sub-matrices, rows or columns.

To unselect grid-cells (Unselected cells appear white)

- Click on selected cell to unselect it
- Click and drag across selected cells you want to unselect.

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>

The screenshot shows the UCSC Genome Browser interface for the Human Feb. 2009 (GRCh37/hg19) Assembly. The browser window has a blue header bar with links for Genomes, Genome Browser, Tools, Mirrors, Downloads, My Data, View, Help, and About Us. Below the header is a search bar and a zoom control panel. The main content area displays a genomic track for chromosome 1, showing gene annotations from GENECODE Version 19. A red circle highlights the title "UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly". Below the title is a navigation bar with buttons for moving left and right, zooming in and out, and a search bar. The genomic track shows a scale from 53,619,091 to 54,233,877 bp. A detailed track below the main track shows gene annotations from GENECODE Version 19. At the bottom of the screen, there are two sections of track hubs: "Mapping and Sequencing" and "Genes and Gene Predictions", each with several tracks listed with "hide" dropdown menus.

Human chr1:53,619,091-54...

genome.ucsc.edu/cgi-bin/hgTracks?db=hg19&position=chr1%3A53619091-54233877&hgsid=433369863_x2foKvh2UrcANNORa6ya2zbrTS4e

Search

Most Visited Getting Started Latest Headlines Thumbnail image of Fi...

Genomes Genome Browser Tools Mirrors Downloads My Data View Help About Us

UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

move <<< << < > >> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x

chr1:53,619,091-54,233,877 614,787 bp enter position, gene symbol or search terms go More on-site workshops available!

chr1 (p32.3) 83 1p31.1 1q12 82.1 1q41 q43-q44

Scale: chr1: 53,700,000 53,750,000 53,800,000 53,850,000 53,900,000 53,950,000 54,000,000 54,050,000 54,100,000 54,150,000 54,200,000

Basic: Gene Annotations from GENECODE Version 19

move start < 2.0 > move end < 2.0 >

Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels up or down to reorder tracks. Drag tracks left or right to new position.

track search default tracks default order hide all add custom tracks track hubs configure reverse resize refresh

collapse all expand all

Use drop-down controls below and press refresh to alter tracks displayed. Tracks with lots of items will automatically be displayed in more compact modes.

Mapping and Sequencing refresh

Base Position Assembly BAC End Pairs BU ORCHID Chromosome Band deCODE Recomb

ENCODE FISH Clones Fosmid End Pairs Gap GC Percent GRC Incident

Pilot hide hide hide hide hide hide

GRC Map Contigs GRC Patch Release Hg18 Diff Hg38 Diff Hi Seq Depth INSDC

Contigs hide hide hide hide hide hide

LRG Regions Map Contigs Mappability Recomb Rate Restr Enzymes Short Match

STS Markers Wiki Track hide hide hide hide hide

Genes and Gene Predictions refresh

UCSC Genes RefSeq Genes AceView Genes CCDS Ensembl Genes EvoFold

hide hide hide hide hide

Accessing Roadmap ChromHMM through the UCSC Genome Browser Track Hubs

<http://genome.ucsc.edu>

The screenshot shows the UCSC Genome Browser interface for the Human Feb. 2009 (GRCh37/hg19) Assembly. The main view displays chromosome 1 (chr1) from position 53,619,091 to 54,233,877, which is 614,787 bp long. A scale bar indicates a 200 kb range. Below the genome track, a legend shows 'Scale' (chr1), 'Basic' (chromosome structure), and 'hg19' (Gene Annotations from Gencode Version 19). A tooltip explains track options: click or drag on a feature for details, click or drag in the base position track to zoom in, click side bars for track options, drag side bars or labels up or down to reorder tracks, and drag tracks left or right to new position. The 'track hubs' button in the track control panel is circled in red.

UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

move <<< << < > >> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x

chr1:53,619,091-54,233,877 614,787 bp enter position, gene symbol or search terms go More on-site workshops available!

chr1 (p32.3) 32.3 1p31.1 1q12 32.1 1q41 q43.44

Scale chr1: hg19 | 53,700,000 53,750,000 53,800,000 53,850,000 53,900,000 53,950,000 54,000,000 54,050,000 54,100,000 54,150,000 54,200,000 Gene annotations from Gencode Version 19

Basic

move start < 2.0 > move end < 2.0 >

track search default tracks default order hide all add custom tracks **track hubs** configure reverse resize refresh

collapse all expand all

Use drop-down controls below and press refresh to alter tracks displayed. Tracks with lots of items will automatically be displayed in more compact modes.

Mapping and Sequencing

Base Position Assembly BAC End Pairs BU ORCHID Chromosome Band Recomb deCODE Recomb

ENCODE Pilot FISH Clones Fosmid End Pairs Gap GC Percent GRC Incident

GRC Map Contigs GRC Patch Release Hg18 Diff Hg38 Diff Hi Seq Depth INSDC

LRG Regions Map Contigs Mappability Recomb Rate Restr Enzymes Short Match

STS Markers Wiki Track

Genes and Gene Predictions

UCSC Genes RefSeq Genes AceView Genes CCDS Ensembl Genes EvoFold

Accessing Roadmap ChromHMM through the UCSC Genome Browser Track Hubs

The screenshot shows a list of track hubs available through the UCSC Genome Browser. The 'Roadmap Epigenomics Integrative Analysis Hub' is circled in red and has a callout box with the following text:

Note: Different than track hub
[Roadmap Epigenomics Data Complete Collection at Wash U VizHub](#)

	Build
Connect FANTOM5	RIKEN FANTOM5 Phase1 and Phase2 data hg19, mm9
Connect facebase hub	genomebrowser facebase org hub mm9, mm10, hg18, hg19
Connect EcoliCompHub	E.coli Comparative Assembly Hub [+] EscherichiaColi042Uid161985, reference...
Connect EcoliCompHubWtDups	E.coli Comparative Assembly Hub, With Duplications [+] EscherichiaColi042Uid161985, reference...
Connect Croc and Bird Hub	Croc, Bird, and Archosaur Assembly Hub [+] allMis2, Anc00, Anc01, Anc02, Anc03...
Connect FANTOM5 CAGE RECLU DATA	RIKEN FANTOM5 CAGE clusters by RECLU hg19, hg18
Connect LIBD Human DLPFC Development	RNAseq data across human brain development by age group from LIBD hg19
Connect Roadmap Epigenomics Integrative Analysis Hub	Roadmap Epigenomics Integrative Analysis Hub at Washington University in St. Louis hg19
Connect ZebrafishGenomics	Burgess Lab Zebrafish Genomic Resources danRer7
Connect 454 K562andHelaS3RNAseq	Whole-Cell 454 Hela and K562 RNAseq hg19
Connect PhyloCSF	Evolutionary protein-coding potential as measured by PhyloCSF hg19, hg38, mm10
Connect GRC Genome Issues under Review	Sanger Genome Reference Informatics Team: Genome issues and other features hg38, hg19, mm10, mm9, danRer7, danRer10
Connect ThorntonHub	DrosophilaGenomeAssemblies Dsim-w501
Connect CPTAC Hub v1	CPTAC Hub v1 hg19
Connect rfam12_ncRNA	Rfam 12.0 non-coding RNA annotation [+] hg38, mm10, ce10, galGal4, ci2, danRer7...
Connect UMassMed ZHub	UMassMed H3K4me3 ChIP-seq data for Autistic brains hg19

Accessing Roadmap ChromHMM through the UCSC Genome Browser Track Hubs

The screenshot shows the UCSC Genome Browser interface for the Human Feb. 2009 (GRCh37/hg19) Assembly. The main genome track displays chromosome 1 (chr1) from p32.3 to q43.44, with a zoomed-in view of the 614,787 bp region chr1:53619091-54233877. Below the genome track, a track hub for the "Roadmap Epigenomics Integrative Analysis Hub" is visible. This hub contains several tracks, with the "Con. By Assay..." track highlighted and circled in red. Other tracks shown include "Assembly", "BAC End Pairs", "BU ORCHID", "Chromosome Band", "deCODE Recomb", "FISH Clones", "Gap", "GC Percent", "GRC Incident", "GRC Map Contigs", "GRC Patch Release", "Hg38 Diff", "Hi Seq Depth", "INSDC", "LRG Regions Map Contigs", "Mappability", "Recomb Rate", "Restr Enzymes", "Short Match", "STS Markers", and "Wiki Track".

Accessing Roadmap ChromHMM through the UCSC Genome Browser Track Hubs

Con. By Assay Super-track ... + -

genome.ucsc.edu/cgi-bin/hgTrackUi?hgSID=433369863_x2foKvh2UrcANNORa6ya2zbrTS4e&c=chr1&g=hub_24125_RoadmapConsolidatedAssay

Most Visited Getting Started Latest Headlines Thumbnail image of Fi...

Home Genomes Genome Browser Tools Mirrors Downloads My Data Help About Us

Con. By Assay Super-track Settings

Roadmap Consolidated Analysis hub organized by assay Tracks

Display mode:

chromHMM tracks from Roadmap

DNase hypersensitivity tracks from Roadmap

H2A.Z tracks from Roadmap

H2AK5ac tracks from Roadmap

H2AK9ac tracks from Roadmap

H2BK120ac tracks from Roadmap

H2BK12ac tracks from Roadmap

H2BK15ac tracks from Roadmap

H2BK20ac tracks from Roadmap

H2BK5ac tracks from Roadmap

H3K14ac tracks from Roadmap

H3K18ac tracks from Roadmap

H3K23ac tracks from Roadmap

H3K23me2 tracks from Roadmap

H3K27ac tracks from Roadmap

H3K27me3 tracks from Roadmap

H3K36me3 tracks from Roadmap

H3K4ac tracks from Roadmap

H3K4me1 tracks from Roadmap

Accessing Roadmap ChromHMM through the UCSC Genome Browser Track Hubs

chromHMM Track Settings

genome.ucsc.edu/cgi-bin/hgTrackUi?hgSID=433369863_x2foKvh2UrcANNOra6ya2zbrTS4e&c=chr1&g=hub_24125_RoadmapConsolidatedAssaya27004

Most Visited Getting Started Latest Headlines Thumbnail image of Fi...

Genomes Genome Browser Tools Mirrors Downloads My Data Help About Us

Subtracks Description

chromHMM Track Settings

chromHMM tracks from Roadmap ([Con. By Assay](#))

Maximum display mode:

Select views ([help](#)):

[AuxiliaryHMM](#) [PrimaryHMM](#) [ImputedHMM](#)

Select subtracks by views and sample type:

Data Type: Real Imputed

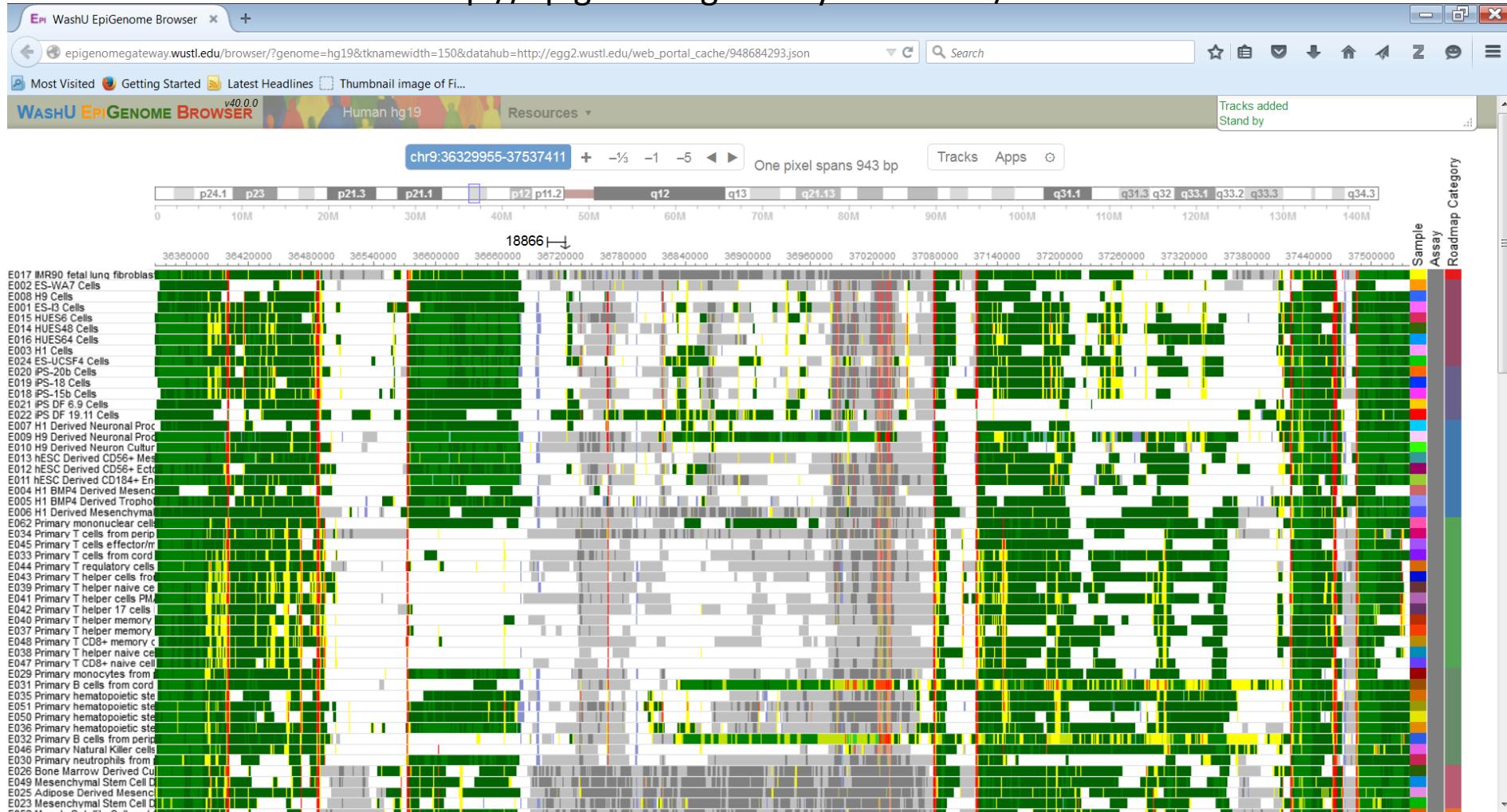
Sample Type	Views	AuxiliaryHMM	PrimaryHMM	ImputedHMM
HepG2	[+/-]	[+/-]	[+/-]	[+/-] <input checked="" type="checkbox"/>
CD4+ CD25- IL17- PMA-Ionomycin stimulated MACS purified Th Primary Cells	[+/-]	[+/-]	[+/-]	[+/-] <input checked="" type="checkbox"/>
Lung	[+/-]	[+/-]	[+/-]	[+/-] <input checked="" type="checkbox"/>
Ovary	[+/-]	[+/-]	[+/-]	[+/-] <input checked="" type="checkbox"/>
Thymus	[+/-]	[+/-]	[+/-]	[+/-] <input checked="" type="checkbox"/>
HMEC	[+/-]	[+/-]	[+/-]	[+/-] <input checked="" type="checkbox"/>
CD4 Naive Primary cells	[+/-]	[+/-]	[+/-]	[+/-] <input checked="" type="checkbox"/>
CD4+ CD25- CD45RA+ Naive Primary Cells	[+/-]	[+/-]	[+/-]	[+/-] <input checked="" type="checkbox"/>
CD8 Naive Primary cells	[+/-]	[+/-]	[+/-]	[+/-] <input checked="" type="checkbox"/>
CD4 Memory Primary cells	[+/-]	[+/-]	[+/-]	[+/-] <input checked="" type="checkbox"/>
Mobilized CD34 Primary cells	[+/-]	[+/-]	[+/-]	[+/-] <input checked="" type="checkbox"/>
CD3 Primary cells	[+/-]	[+/-]	[+/-]	[+/-] <input checked="" type="checkbox"/>
Stomach Mucosa	[+/-]	[+/-]	[+/-]	[+/-] <input checked="" type="checkbox"/>
CD19 Primary cells	[+/-]	[+/-]	[+/-]	[+/-] <input checked="" type="checkbox"/>
Penis Foreskin Fibroblast Primary Cells	[+/-]	[+/-]	[+/-]	[+/-] <input checked="" type="checkbox"/>

Accessing Roadmap ChromHMM through the UCSC Genome Browser Track Hubs



Human Epigenome Browser at Washington University

<http://epigenomegateway.wustl.edu/>

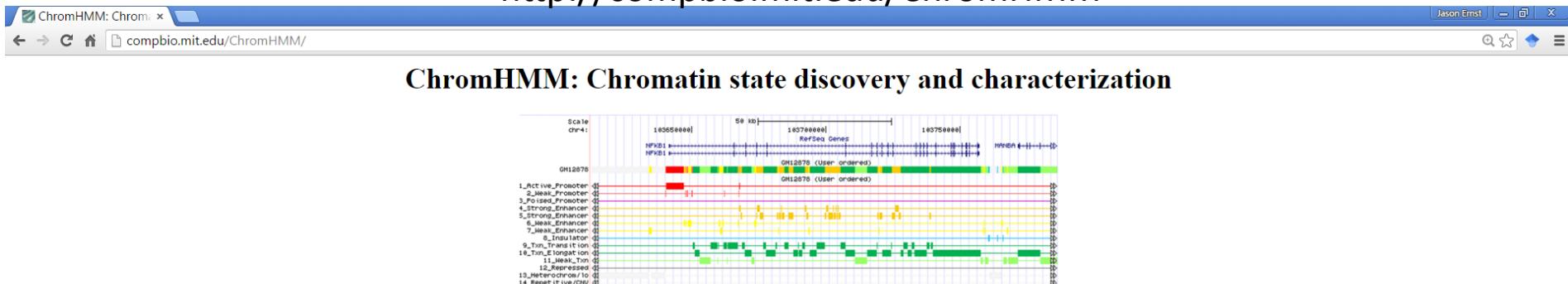


Talk Outline

- 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>



The screenshot shows a web browser window titled "ChromHMM: Chrom". The URL in the address bar is "compbio.mit.edu/ChromHMM". The main content is titled "ChromHMM: Chromatin state discovery and characterization". Below the title is a genomic track visualization for chromosome 4. The track shows genomic features like RefSeq Genes and H3K4me1, along with 16 chromatin states represented by colored bars. A legend on the left lists the states: 1_Active_Promoter, 2_Dense_Promoter, 3_Fine_Scattered, 4_Strong_Enhancer, 5_Moderate_Enhancer, 6_Weak_Enhancer, 7_Hweak_Enhancer, 8_Closed, 9_Tnn_Transition, 10_Tnn_Escapegt, 11_Junc_Tnn, 12_Repressed, 13_Interchrom, 14_Repetititve_CNV, and 15_Repetititve_CNV. A red arrow points from the "Software download" section to the "ChromHMM software v1.11 (version log)" link.

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-occurring 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.

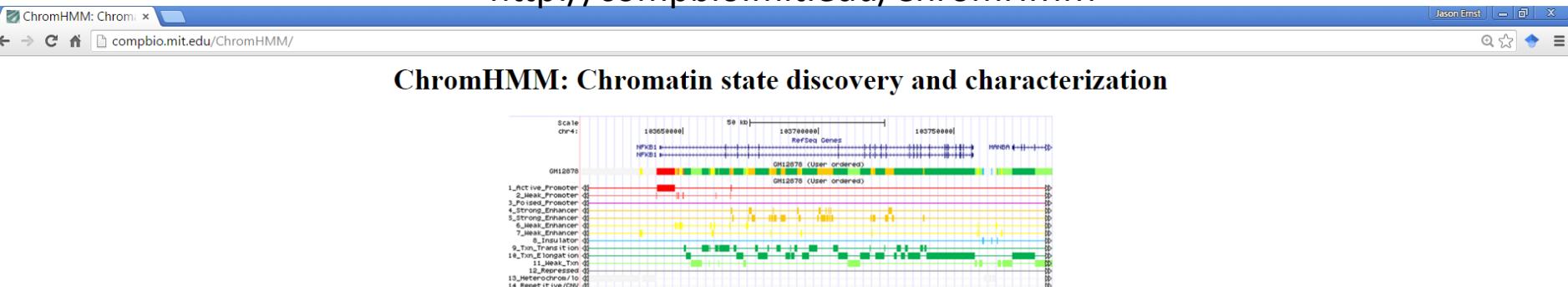
New in version 1.11: ChromHMM has a *BinarizeBam* command which allows binarizing bam files of aligned reads.

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@ucla.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-R01-HG005334 and NIH 1 U54 HG004570).

ChromHMM Website

<http://compbio.mit.edu/ChromHMM>



The screenshot shows a web browser window titled "ChromHMM: Chrom". The address bar contains "compbio.mit.edu/ChromHMM". The main content area has a title "ChromHMM: Chromatin state discovery and characterization". Below the title is a genomic track visualization for chromosome 4. The track shows genomic features like RefSeq Genes and H3K4me1, along with chromatin state predictions. A legend on the left lists 15 states: 1_Active_Promoter, 2_Dense_Promoter, 3_Force_Promoter, 4_Strong_Enhancer, 5_Moderate_Enhancer, 6_Weak_Enhancer, 7_Hweak_Enhancer, 8_Closed, 9_Tnn_Transition, 10_Tnn_Engagement, 11_Open_Tnn, 12_Repressed, 13_Interchromosomal, 14_Repetitive_DNA, and 15_Repetititve_CNN. A red arrow points from the "Software manual" section below to the "ChromHMM manual" link in the sidebar.

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-occurring 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.

- [ChIP-seq 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@ucla.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-R01-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>

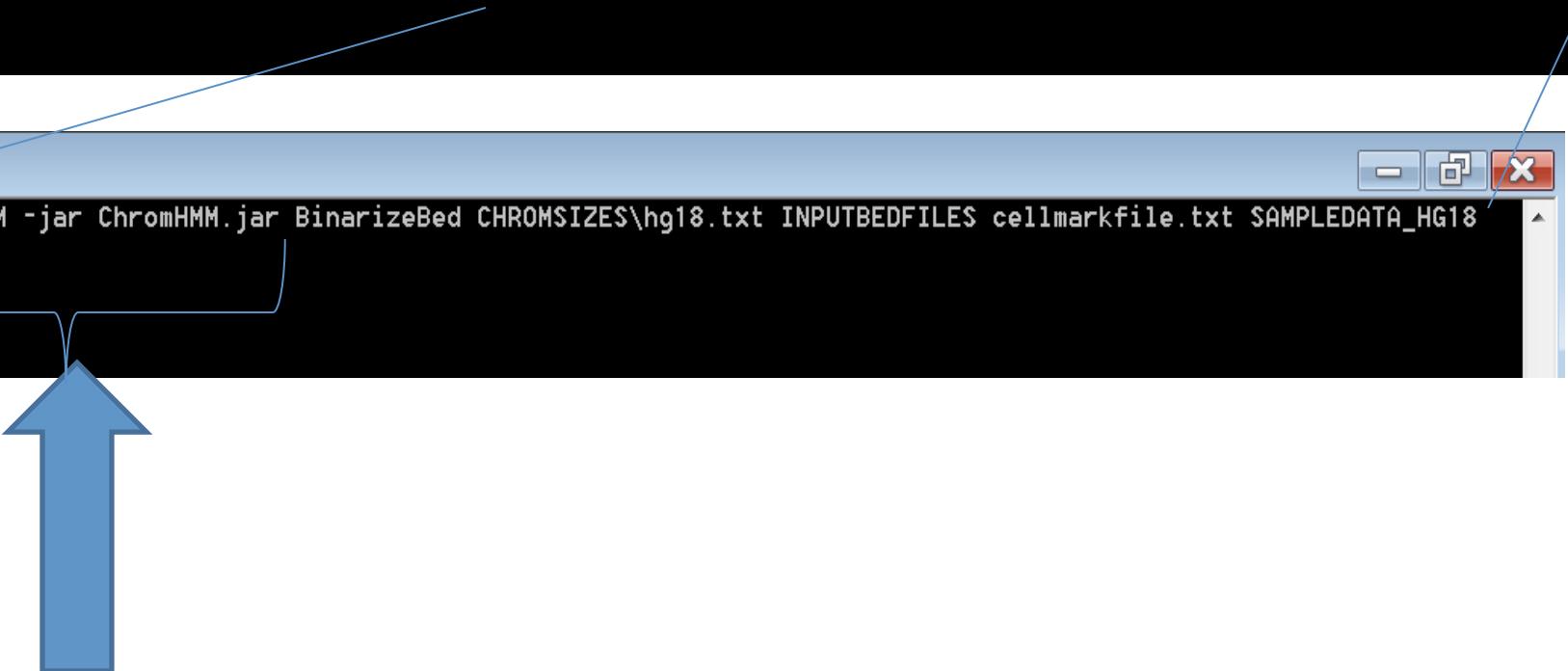
2. Unzip ChromHMM.zip
3. Open a command line
4. Change into the ChromHMM directory
5. Enter the command:

```
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



```
C:\Windows\system32\cmd.exe
users\Jason Ernst\Documents\ChromHMM\PUBLICCHROMHMM\ChromHMM_1_1_0\ChromHMM>java -mx1600M -jar ChromHMM.jar BinarizeBed CHROMSIZES\hg18.txt INPUTBEDFILES cellmarkfile.txt SAMPLEDATA_HG18
```

```
>java -mx1600M -jar ChromHMM.jar BinarizeBed CHROMSIZES\hg18.txt INPUTBEDFILES cellmarkfile.txt SAMPLEDATA_HG18
```

Java command '`-mx1600M`' specifies memory to Java

BinarizeBed

```
C:\Windows\system32\cmd.exe  
users\Jason Ernst\Documents\ChromHMM\PUBLICCHROMHMM\ChromHMM_1_1_0\ChromHMM>java -mx1600M -jar ChromHMM.jar BinarizeBed CHROMSIZES\hg18.txt INPUTBEDFILES cellmarkfile.txt SAMPLEDATA_HG18
```

```
>java -mx1600M -jar ChromHMM.jar BinarizeBed CHROMSIZES\hg18.txt INPUTBEDFILES cellmarkfile.txt SAMPLEDATA_HG18
```

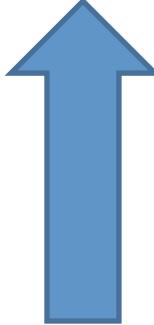


ChromHMM command

BinarizeBed

```
C:\Windows\system32\cmd.exe  
users\Jason Ernst\Documents\ChromHMM\PUBLICCHROMHMM\ChromHMM_1_1_0\ChromHMM>java -mx1600M -jar ChromHMM.jar BinarizeBed CHROMSIZES\hg18.txt INPUTBEDFILES cellmarkfile.txt SAMPLEDATA_HG18
```

```
>java -mx1600M -jar ChromHMM.jar BinarizeBed CHROMSIZES\hg18.txt INPUTBEDFILES cellmarkfile.txt SAMPLEDATA_HG18
```

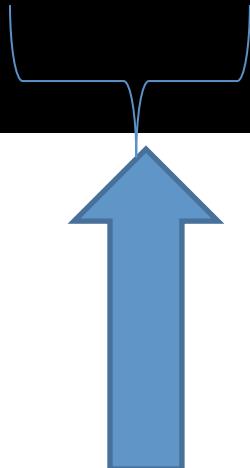


File with the chromosome lengths for the assembly

BinarizeBed

```
C:\Windows\system32\cmd.exe  
users\Jason Ernst\Documents\ChromHMM\PUBLICCHROMHMM\ChromHMM_1_1_0\ChromHMM>java -mx1600M -jar ChromHMM.jar BinarizeBed CHROMSIZES\hg18.txt INPUTBEDFILES cellmarkfile.txt SAMPLEDATA_HG18
```

```
>java -mx1600M -jar ChromHMM.jar BinarizeBed CHROMSIZES\hg18.txt INPUTBEDFILES cellmarkfile.txt SAMPLEDATA_HG18
```



DIRECTORY of BED files

BinarizeBed

```
C:\Windows\system32\cmd.exe
users\Jason Ernst\Documents\ChromHMM\PUBLICCHROMHMM\ChromHMM_1_1_0\ChromHMM>java -mx1600M -jar ChromHMM.jar BinarizeBed CHROMSIZES\hg18.txt INPUTBEDFILES cellmarkfile.txt SAMPLEDATA_HG18
```

```
>java -mx1600M -jar ChromHMM.jar BinarizeBed CHROMSIZES\hg18.txt INPUTBEDFILES cellmarkfile.txt SAMPLEDATA_HG18
```

cell	mark	cell-mark
cell1	mark1	cell1_mark1.bed
cell1	mark2	cell1_mark2.bed
cell2	mark1	cell2_mark1.bed
cell2	mark2	cell2_mark2.bed

cell1_control.bed
cell1_control.bed
cell2_control.bed
cell2_control.bed



Cell-mark –file table

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

BinarizeBed

```
C:\Windows\system32\cmd.exe
users\Jason Ernst\Documents\ChromHMM\PUBLICCHROMHMM\ChromHMM_1_1_0\ChromHMM>java -mx1600M -jar ChromHMM.jar BinarizeBed CHROMSIZES\hg18.txt INPUTBEDFILES cellmarkfile.txt SAMPLEDATA_HG18
```



```
>java -mx1600M -jar ChromHMM.jar BinarizeBed CHROMSIZES\hg18.txt INPUTBEDFILES cellmarkfile.txt SAMPLEDATA_HG18
```

Output directory

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 OUTUTSAMPLE 10 hg18
```

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



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

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 OUTUTSAMPLE 10 hg18
```

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

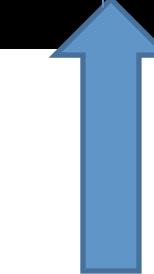


Directory with the Binarized Input

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 OUTUTSAMPLE 10 hg18
```

```
java -mx1600M -jar ChromHMM.jar LearnModel -p 0 SAMPLEDATA_HG18 OUTUTSAMPLE 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
```

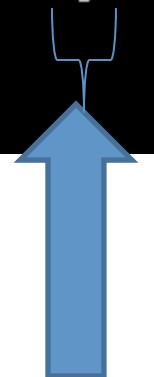


Number of states

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 OUTUTSAMPLE 10 hg18
```

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



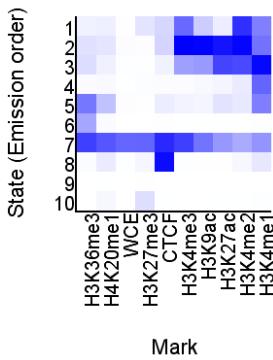
Genome assembly

ChromHMM Report



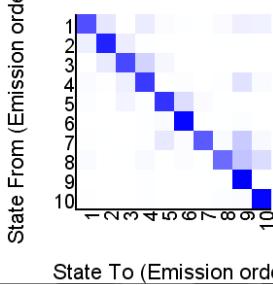
Model Parameters

Emission Parameters



- [Emission Parameter SVG File](#)
- [Emission Parameter Tab-Delimited Text File](#)

Transition Parameters



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

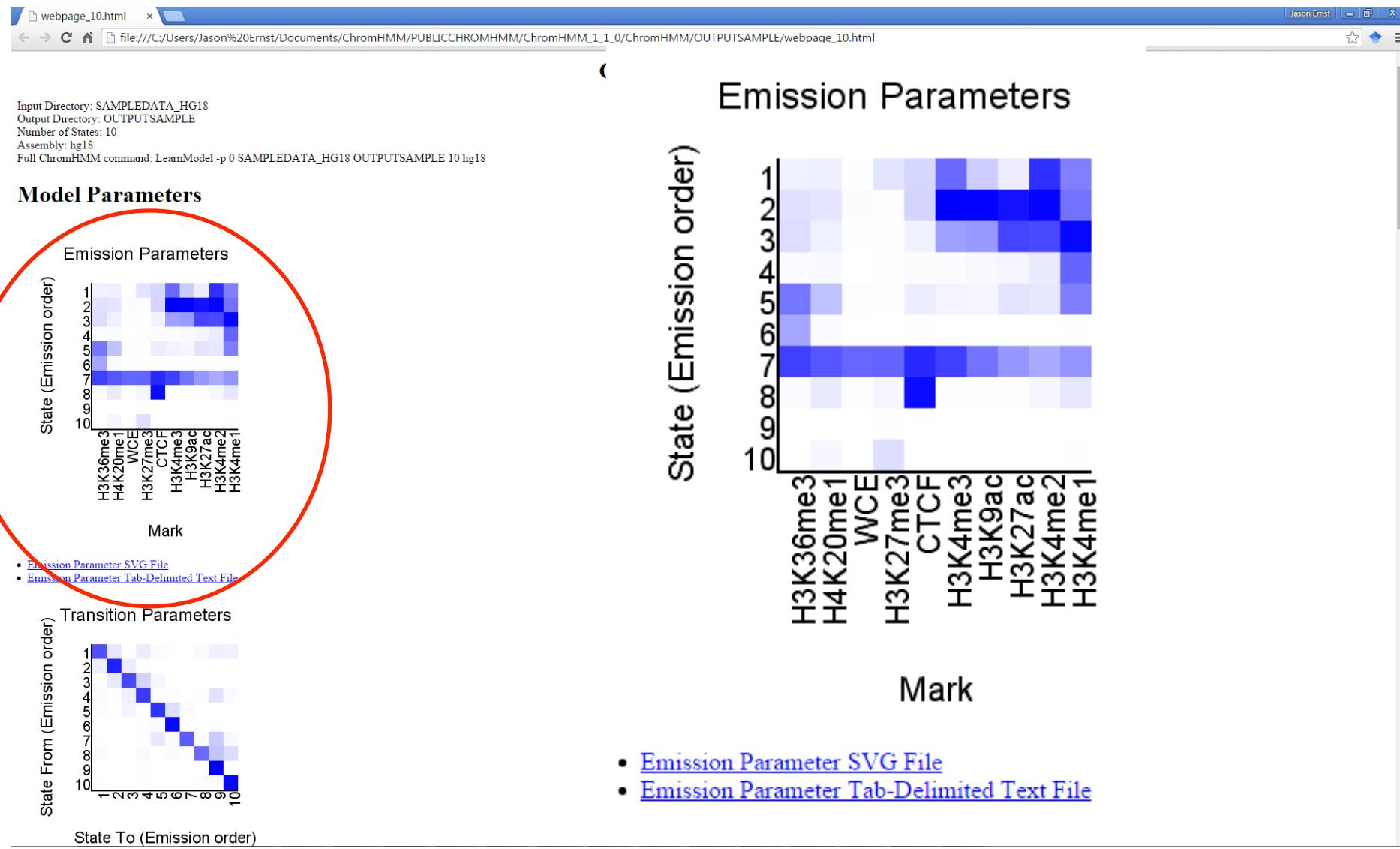
A heatmap representing the emission parameters for 10 states. The y-axis is labeled "State (Emission order)" and ranges from 1 to 10. The x-axis is labeled "Mark" and lists genomic features: H3K36me3, H4K20me1, WCE, H3K27me3, CTCF, H3K4me3, H3K9ac, H3K27ac, H3K4me2, and H3K4me1. The heatmap shows high intensity (blue) at positions (1, H3K36me3), (2, H3K27me3), (3, H3K4me3), (4, H3K9ac), (5, H3K27ac), (6, H3K4me2), (7, H3K4me1), (8, WCE), and (9, CTCF). State 10 has low intensity across all marks.

Transition Parameters

A heatmap representing the transition parameters between 10 states. The y-axis is labeled "State From (Emission order)" and the x-axis is labeled "State To (Emission order), both ranging from 1 to 10. The diagonal from (1,1) to (10,10) is dark blue, indicating high probability of staying in the same state. Off-diagonal elements show varying degrees of transition probability, with higher values generally along the top-left and bottom-right diagonals.

- [Emission Parameter SVG File](#)
- [Emission Parameter Tab-Delimited Text File](#)

Emission Parameters



Transition Parameters

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

State (Emission order) 1 2 3 4 5 6 7 8 9 10

Mark H3K36me3 H4K20me1 WC1 H3K27me3 CTCF H3K4me3 H3K9ac H3K27ac H3K4me2 H3K4me1

Mark

- [Emission Parameter SVG File](#)
- [Emission Parameter Tab-Delimited Text File](#)

Transition Parameters

State From (Emission order) 1 2 3 4 5 6 7 8 9 10

State To (Emission order) 1 2 3 4 5 6 7 8 9 10

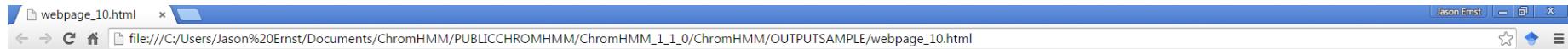
- [Transition Parameter SVG File](#)
- [Transition Parameter Tab-Delimited Text File](#)

Transition Parameters

State From (Emission order) 1 2 3 4 5 6 7 8 9 10

State To (Emission order) 1 2 3 4 5 6 7 8 9 10

Model Parameter File



Genome Segmentation Files

- [GM12878_10 Segmentation File \(Four Column Bed File\)](#)
- [K562_10 Segmentation File \(Four Column Bed File\)](#)

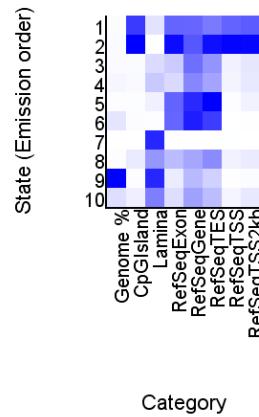
Custom Tracks for loading into the UCSC Genome Browser:

- [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

Fold Enrichment GM12878_10



- [GM12878_10 Overlap Enrichment SVG File](#)
- [GM12878_10 Overlap Enrichment Tab-Delimited Text File](#)

Segmentation File

File webpage_10.html x Jason Ernst

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

Fold Enrichment GM12878_10

State (Emission order)

Category

1 2 3 4 5 6 7 8 9 10

Genome % CpGisland Lamina RefSeqExon RefSeqGene RefSeqTSS RefSeqTS2kD RefSeqTS2kd

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	

- [GM12878_10 Overlap Enrichment SVG File](#)
- [GM12878_10 Overlap Enrichment Tab-Delimited Text File](#)

Browser Files

• 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

Can load into browser UCSC Genome, IGV

State Enrichments

GM12878_10 Enrichments

Fold Enrichment GM12878_10

State (Emission order)

Category

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

UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

chr11:18,444,553-20,155,902 1,711,350 bp. enter position, gene symbol or search terms

More on-site workshops available!

Scale: chr11: 19,000 kb 19,500 kb 20,000 kb

Move start: < 2.0 > Move end: < 2.0 >

Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels up or down to reorder tracks. Drag tracks left or right to new position.

track search | default tracks | default order | hide all | manage custom tracks | track hubs | configure | reverse | resize | refresh | collapse all | expand all | Use drop-down controls below and press refresh to alter tracks displayed. Tracks with lots of items will automatically be displayed in more compact modes.

Custom Tracks

GM12878_10 K562_10

dense dense

IGV

File Genomes View Tracks Regions Tools GenomeSpace Help

Human hg19 chr11 chr11:111,581,366-116,303,579 Go

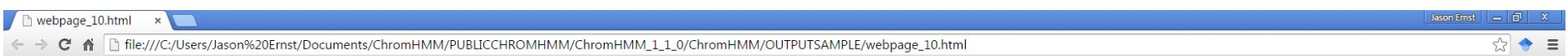
RefSeq Genes

GM12878_10

K562_10

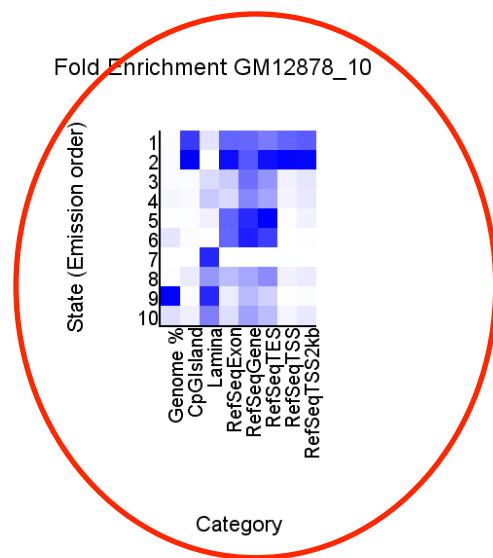
https://www.broadinstitute.org/igv/

Enrichments



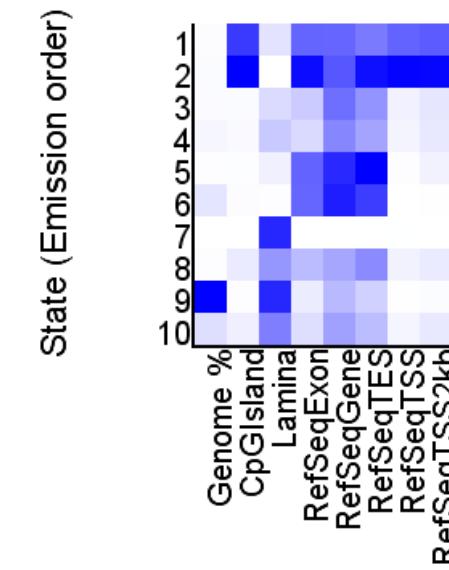
State Enrichments

GM12878_10 Enrichments

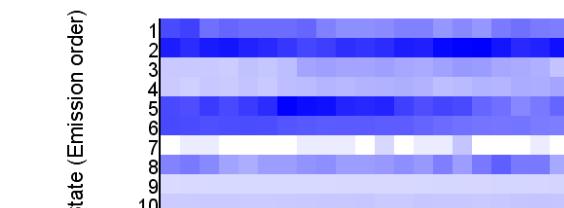


- [GM12878_10 Overlap Enrichment SVG File](#)
- [GM12878_10 Overlap Enrichment Tab-Delimited Text File](#)

Fold Enrichment GM12878_10



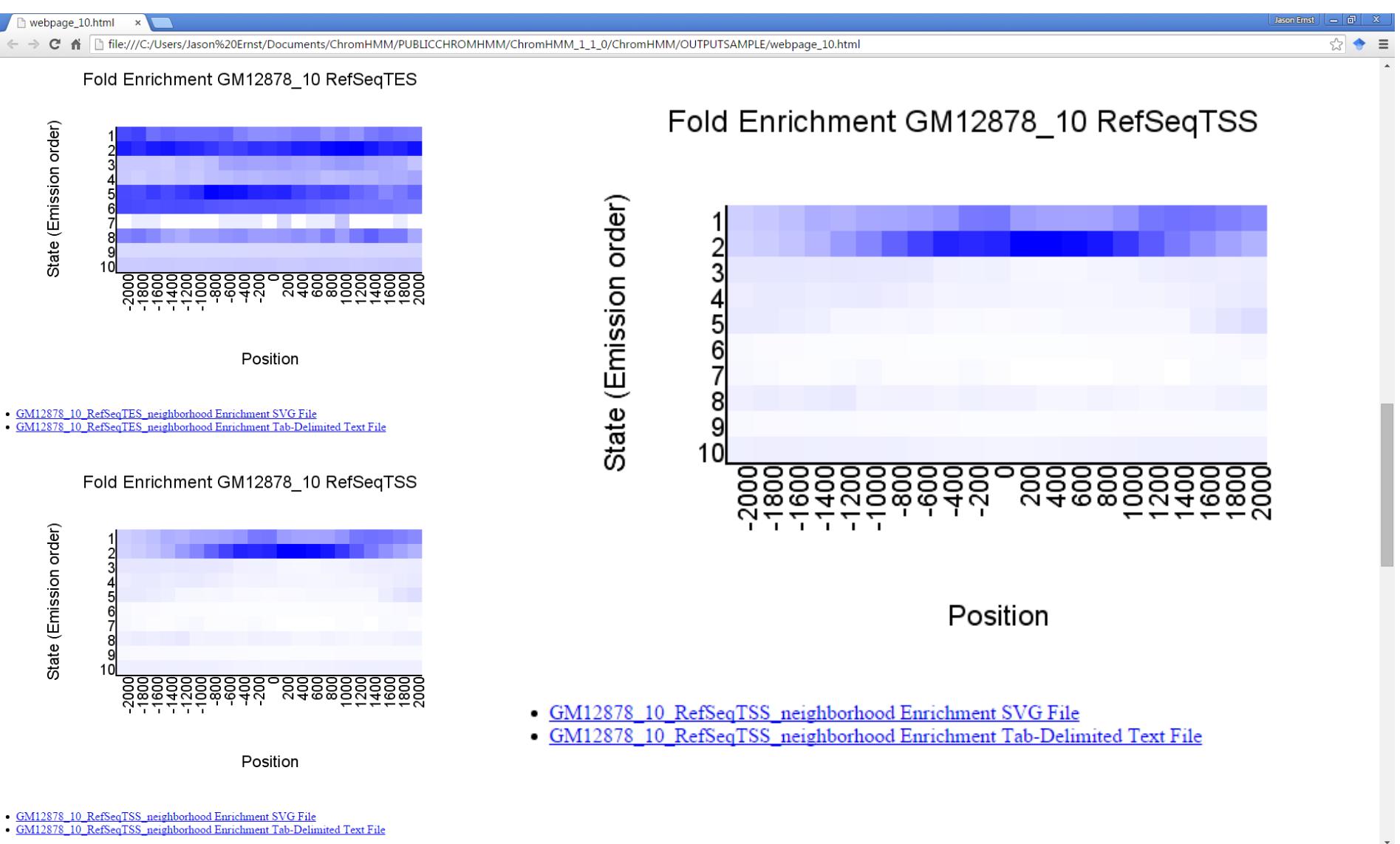
Fold Enrichment GM12878_10 RefSeqTES



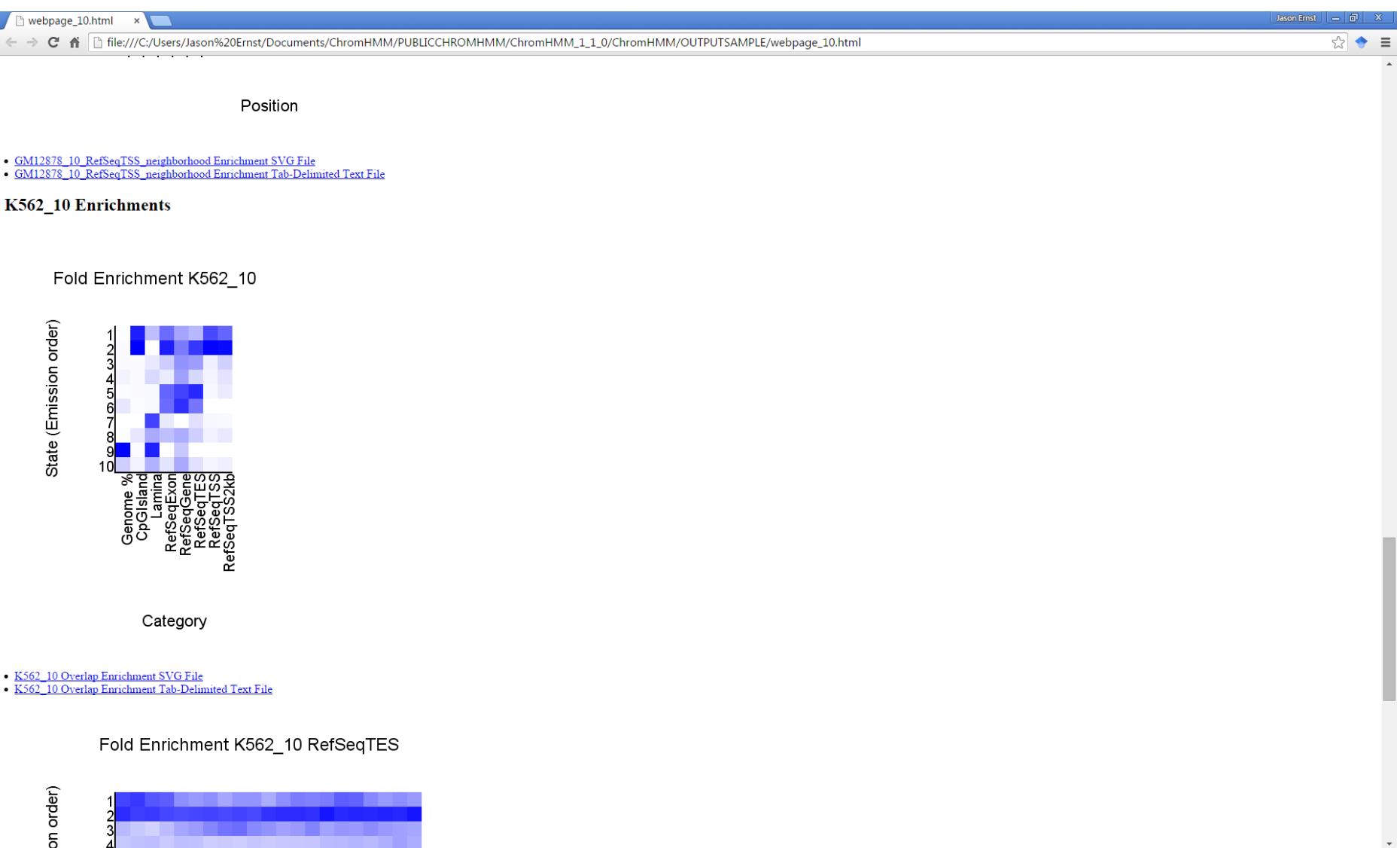
Category

- [GM12878_10 Overlap Enrichment SVG File](#)
- [GM12878_10 Overlap Enrichment Tab-Delimited Text File](#)

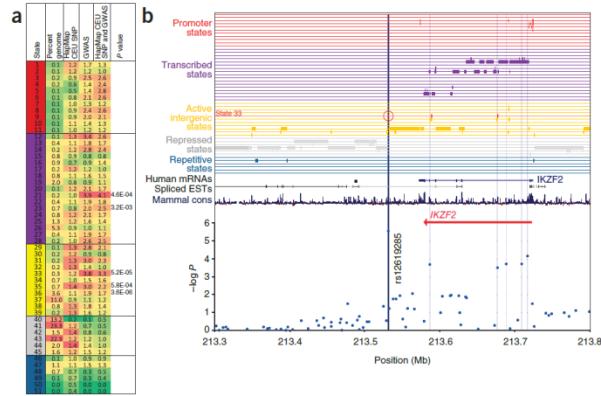
Positional Plots



Enrichments for Additional Cell Types

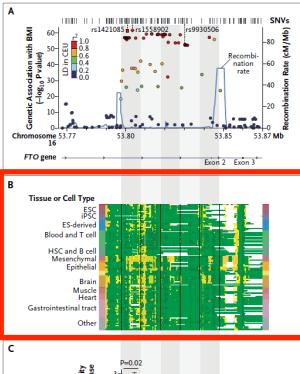


Chromatin states to interpret disease variants



- Specific chromatin states enriched in GWAS catalog

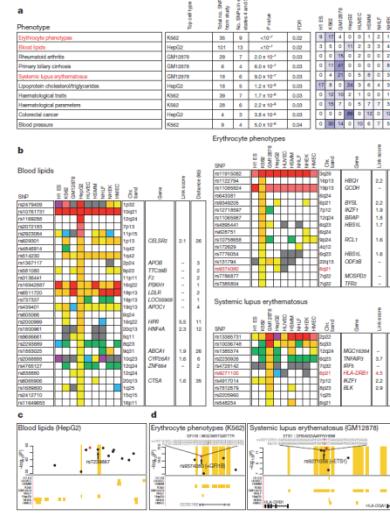
Ernst and Kellis, *Nature Biotech* 2010



- Imputation based chromatin state used in dissection FTO loci

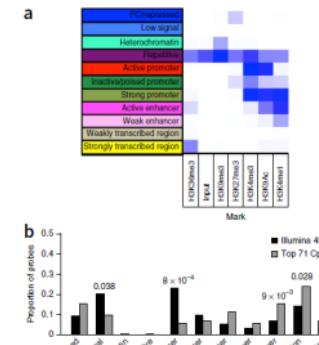
Claussnitzer et al, *NEJM* 2015

- Many other examples in the literature



- Enhancers from different cell types enriched in different traits

Ernst et al, *Nature* 2011



- Interpreting epigenetic disease associated variation in Alzheimer's disease

De Jager et al, *Nature Neuroscience* 2014

Collaborators and Acknowledgements

- Manolis Kellis

ENCODE consortium

- Brad Bernstein production group

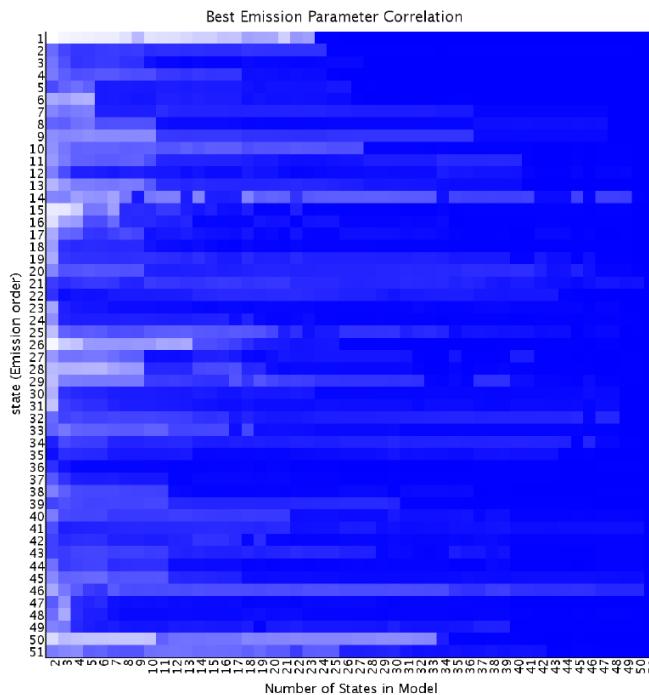
Roadmap Epigenomics consortium

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.

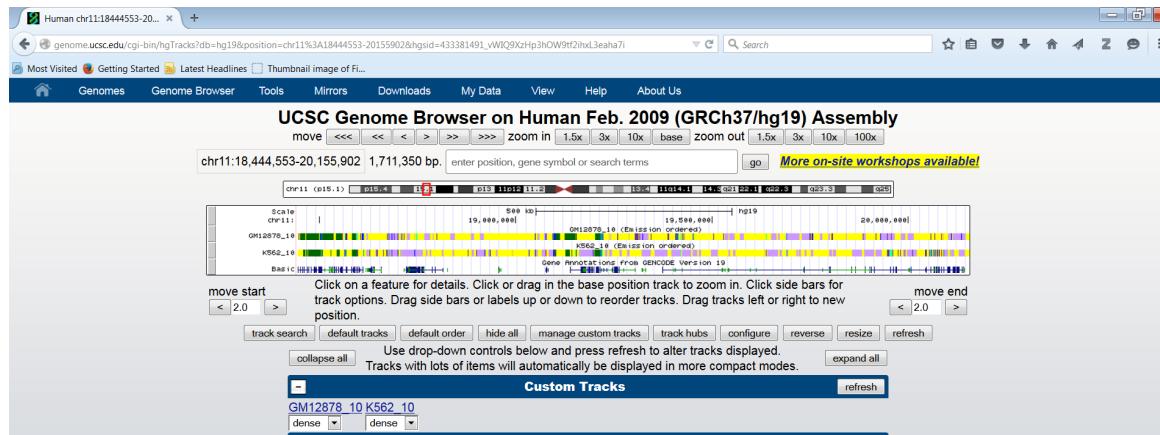


```
CompareModels [-color r,g,b] referencemodel comparedir outputprefix
```

Additional Commands

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

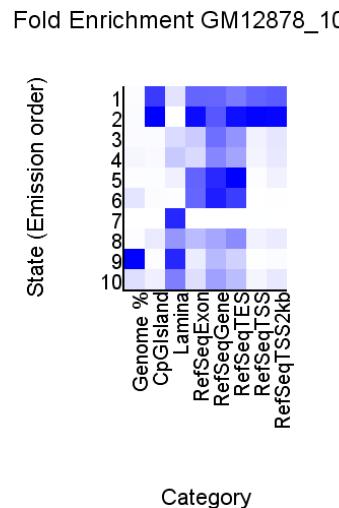
```
MakeBrowserFiles [-c colormappingfile] [-m labelmappingfile] [-n numstates]
segmentfile segmentationname outputfileprefix
```



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 inputcoaddir 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
```