

# 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

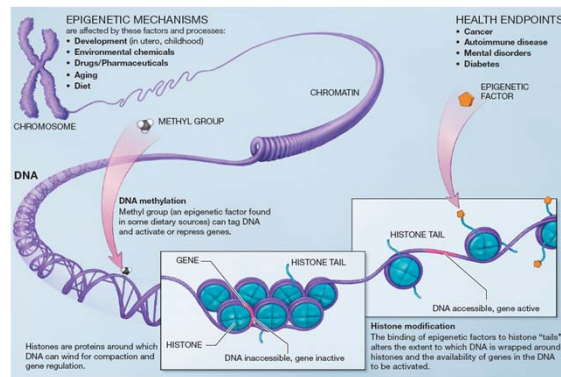
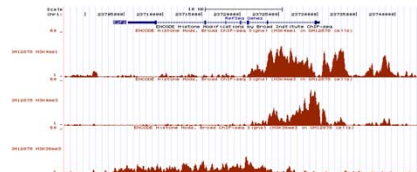
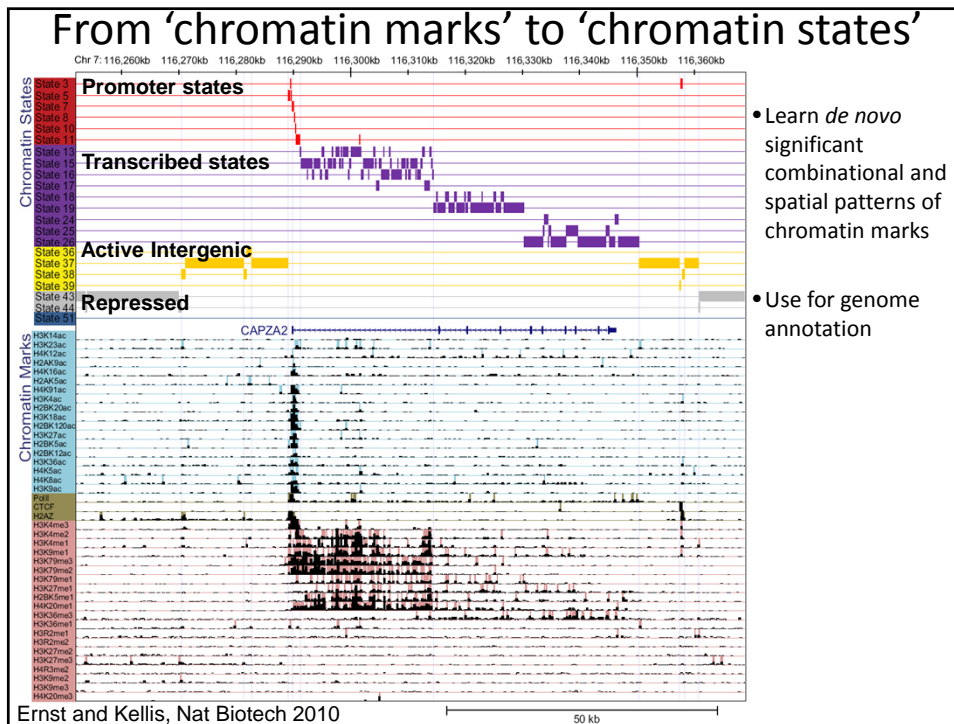


Image source: <http://nihroadmap.nih.gov/epigenomics/>

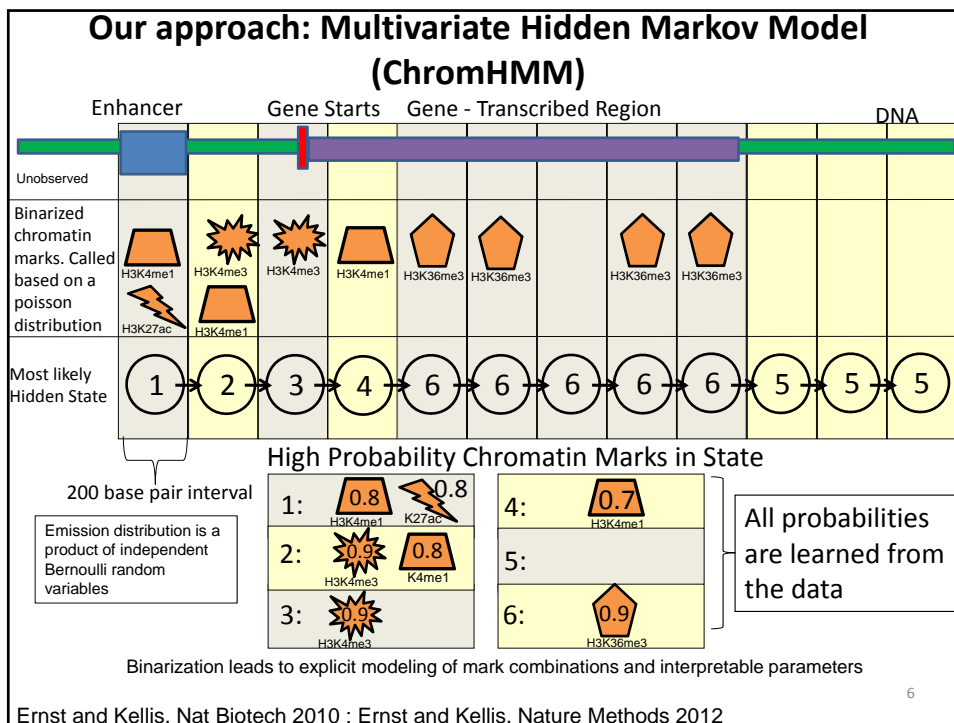
Histone Modifications can be Mapped Genome-wide with ChIP-seq



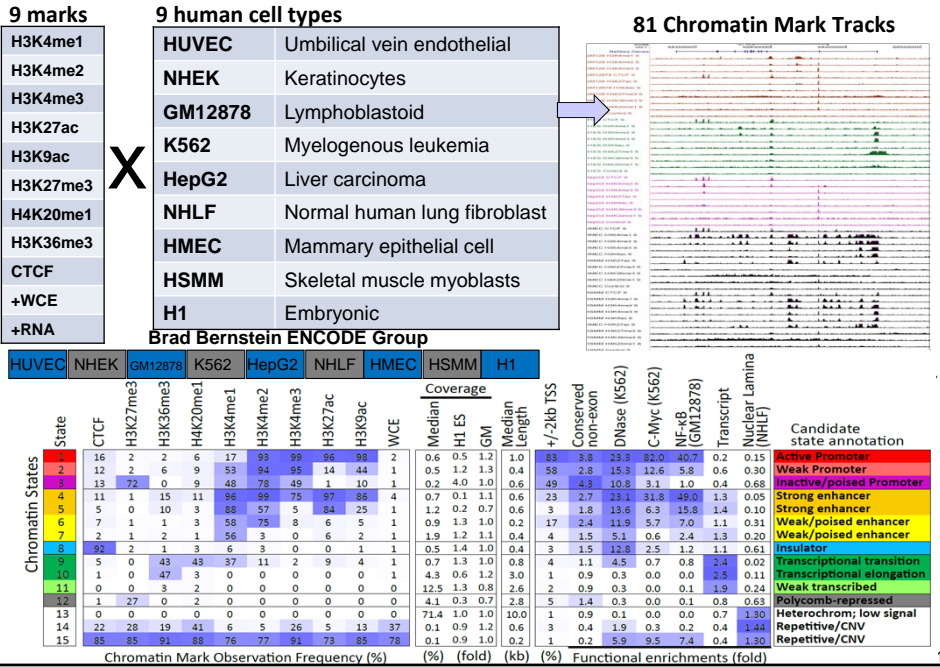


- Learn *de novo* significant combinational and spatial patterns of chromatin marks

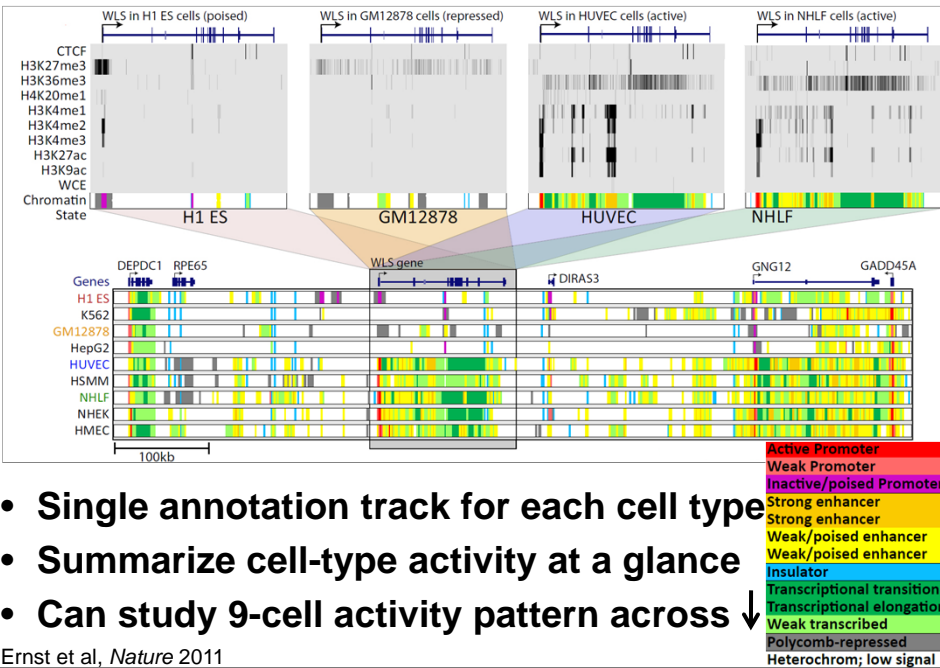
- Use for genome annotation



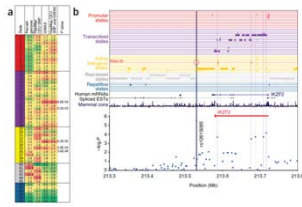
# ENCODE: Study nine marks in nine human cell lines



## Chromatin states dynamics across nine ENCODE cell types

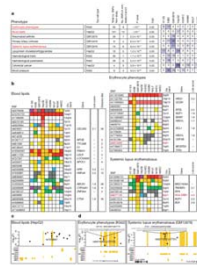


## Chromatin states to interpret disease variants



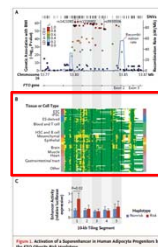
- Specific chromatin states enriched in GWAS catalog

Ernst and Kellis, *Nature Biotech* 2010



- Enhancers from different cell types enriched in different traits

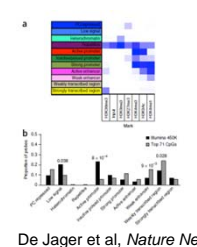
Ernst et al, *Nature* 2011



- Imputation based chromatin state used in dissection FTO loci

Clausnitzer et al, *NEJM* 2015

- Many other examples in the literature



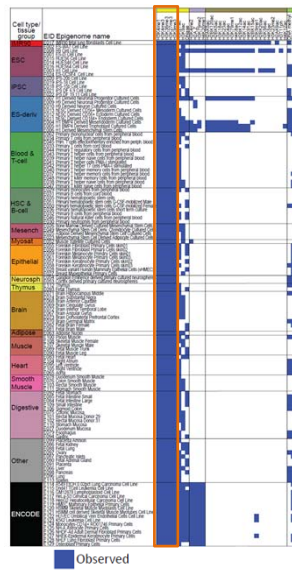
- Interpreting epigenetic disease associated variation in Alzheimer's disease

De Jager et al, *Nature Neuroscience* 2014

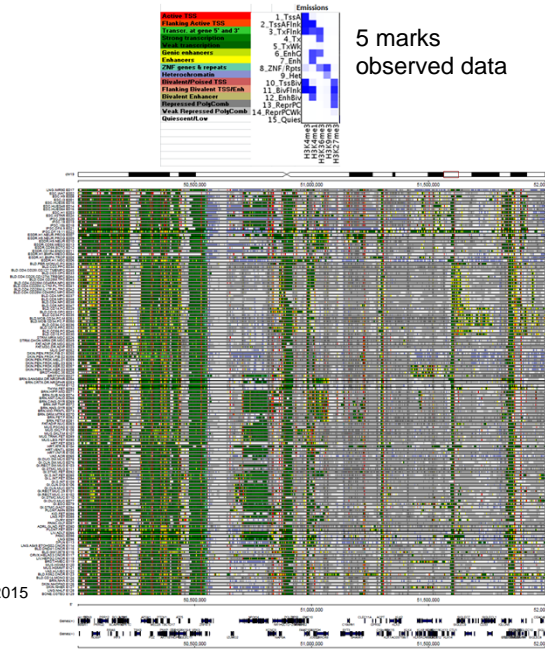
## Talk Outline

- 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

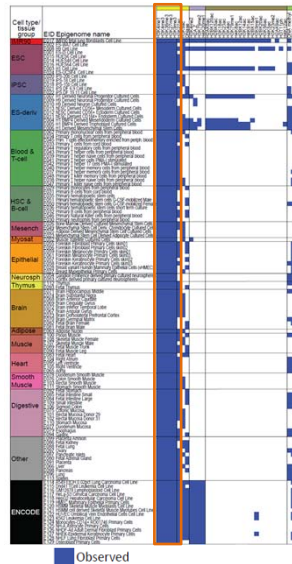


Roadmap Epigenomics Consortium et al, *Nature* 2015  
16 cell types from ENCODE2

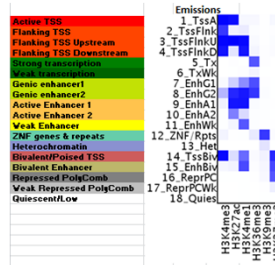


5 marks  
observed data

## Extended Chromatin State Model with H3K27ac



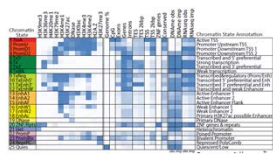
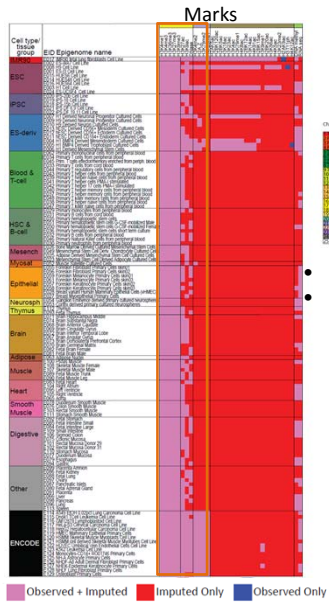
16 cell types from ENCODE2  
Roadmap Epigenomics Consortium et al, *Nature* 2015



- 6 marks observed data
- Defined on 98 cell/tissue types only



# Chromatin States Defined on Imputed Data

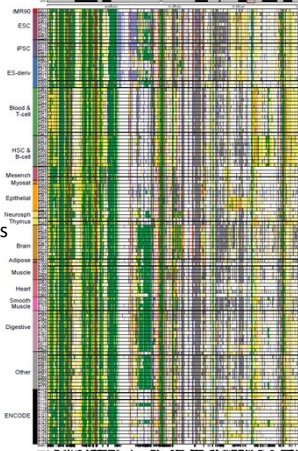


- 12 marks imputed data
- Defined on all cell/tissue types

ChromImpute method



Ernst and Kellis, *Nature Biotech* 2015



# ChromHMM Models across Many Roadmap/ENCODE2 Cell and Tissue Types

## Primary Core Marks segmentation

- State 1 - **Red** - TssA (Active\_TSS)
- State 2 - **OrangeRed** - TssAFlnk (Flanking\_Active\_TSS)
- State 3 - **OrangeRed** - TxFlnk (Transcr\_at\_gene\_5\_and\_3primer)
- State 4 - **Green** - Tx (Strong\_transcription)
- State 5 - **GreenYellow** - 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 (BivalentPoised\_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 - **OrangeRed** - TssFlnk (Flanking\_TSS)
- State 3 - **OrangeRed** - TssFlnkU (Flanking\_TSS\_Upstream)
- State 4 - **OrangeRed** - TssFlnkD (Flanking\_TSS\_Downstream)
- State 5 - **Green** - Tx (Strong\_transcription)
- State 6 - **GreenYellow** - 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 - **MediumAquamarine** - ZNF/Rpts (ZNF\_genes&repeats)
- State 13 - **PaleTurquoise** - Het (Heterochromatin)
- State 14 - **IndianRed** - TssBiv (BivalentPoised\_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 - **OrangeRed** - PromU1 (Promoter Upstream TSS)
- State 3 - **OrangeRed** - PromD1 (Promoter Downstream TSS with DNase)
- State 4 - **OrangeRed** - PromD2 (Promoter Downstream TSS)
- State 5 - **Green** - Tx5' (Transcription 5')
- State 6 - **Green** - Tx (Transcription)
- State 7 - **Green** - Tx3' (Transcription 3')
- State 8 - **LightOrange** - 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 - **MediumAquamarine** - 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

Images from UCSC genome browser

Note: additional ChromHMM annotations based on older ENCODE models available from UCSC genome browser

# Accessing Roadmap/ENCODE2 ChromHMM through the UCSC Genome Browser Track Hubs

<http://genome.ucsc.edu>

The screenshot shows the UCSC Genome Browser interface for Human chr1:53,600,000-54,200,000 bp. The browser title is "UCSC Genome Browser on Human Feb. 2019 (GRCh37/hg19) Assembly". The track configuration menu is open, showing various tracks under "Mapping and Sequencing" and "Genes and Gene Predictions". The "track hubs" button is circled in red.

# Accessing Roadmap/ENCODE2 ChromHMM through the UCSC Genome Browser Track Hubs

<http://genome.ucsc.edu>

The screenshot shows the UCSC Genome Browser interface for Human chr1:53,600,000-54,200,000 bp. The browser title is "UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly". The track configuration menu is open, showing various tracks under "Mapping and Sequencing" and "Genes and Gene Predictions". The "track hubs" button is circled in red.



# Accessing Roadmap/ENCODE2 ChromHMM through the UCSC Genome Browser Track Hubs

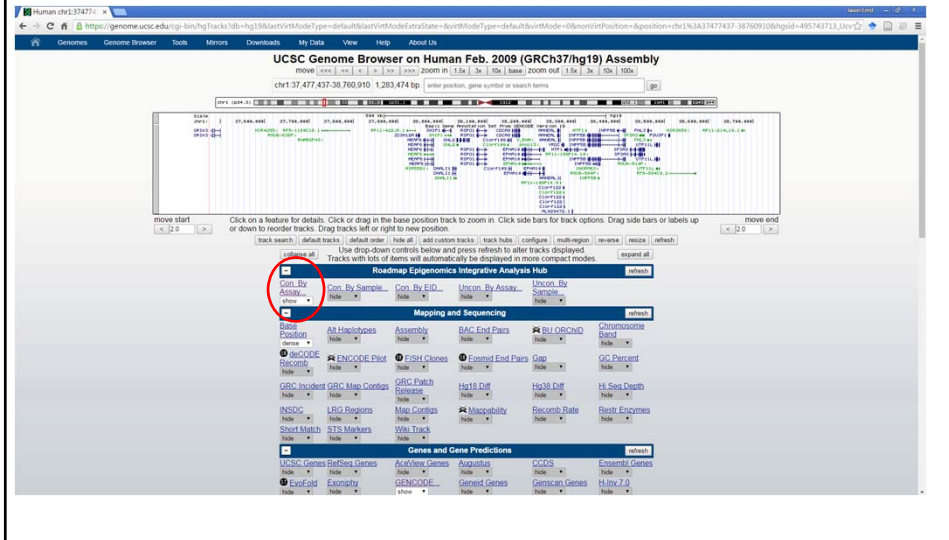
The screenshot shows the UCSC Genome Browser Track Hubs interface. A table lists various track hubs. The 'Roadmap Epigenomics Integrative Analysis Hub' is highlighted with a red circle. A yellow callout box with a blue border contains the text: "Note: Different than track hub Roadmap Epigenomics Data Complete Collection at Wash U VizHub".

Connect	UCD Methylation	UCD DNA methylation data	hg19, hg18, mm10, mm9, mm8, hg19, mm10
Connect	Ensembl Regulatory Build	Evidence summaries and provisional results for the new Ensembl Regulatory Build	hg38, hg19, mm10
Connect	FANTOMS	RIKEN FANTOMS Phase1 and Phase2 data	hg38, mm10, hg19, mm9
Connect	facebase hub	genomeweb/facebase.org hub	mm9, mm10, hg18, hg19
Connect	EcolComHub	E.coli Comparative Assembly Hub	[+] EscherichiaColi042401d161985, reference...
Connect	EcolComHubWDups	E.coli Comparative Assembly Hub, With Duplications	[+] EscherichiaColi042401d161985, reference...
Connect	Croc and Bird Hub	Croc, Bird, and Archosaur Assembly Hub	[+] ANNO2, ANCO0, ANCO1, ANCO2, ANCO3...
Connect	FANTOMS CAGE RECLU DATA	RIKEN FANTOMS CAGE clusters by RECLU	hg19, hg18
Connect	LBD Human DLFC	RNAseq data across human brain development by age group from LBD	hg19
Connect	Roadmap Epigenomics Integrative Analysis Hub	Roadmap Epigenomics Integrative Analysis Hub at Washington University in St. Louis	hg19
Connect	Burgess Lab Zebrafish Genomic Resources		danRer10, danRer7
Connect	454 K562andHeLaSRNAseq	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
Connect	CPTAC Hub v1	Clinical Proteomic Tumor Analysis Consortium (CPTAC) Hub v1	hg19
Connect	rfam12_ncRNA	Rfam 12.0 non-coding RNA annotation	[+] hg38, mm10, ce10, galGAM, ce2, danRer7...
Connect	Peterhof_yeasts	Assemblies, SNI and ChIP data for Saccharomyces cerevisiae strains of the Peterhof Genetic Collection	[+] sacCer3, S1_D1606_spades, S1V_P4_danRer7...
Connect	Vista Enhancers	Vista Browser Track Hub	hg19, mm9, mm10
Connect	Pericme DNA methylation and gene transcription	DNA methylation and gene expression levels for eight tissues from	susScr3
Connect	UMassMed Zhub	UMassMed HiChame3 ChIP-seq data for Autism brains	hg19
Connect	mm9.SMC1.CNAFET	Cohesin(SMC1)-associated chromatin interactions in murine embryonic stem cells	mm9
Connect	lncRNA in Breast Cancer	long noncoding RNAs in Breast Cancer	hg19
Connect	Cancer Genomics Tracks	TCGA and ICGC Cancer Mutations, TCGA Expression, Immune Epitopes Database (IEDB), Cancer Immunity Peptides Database and Dienstmann Variant/Cancer database	hg19

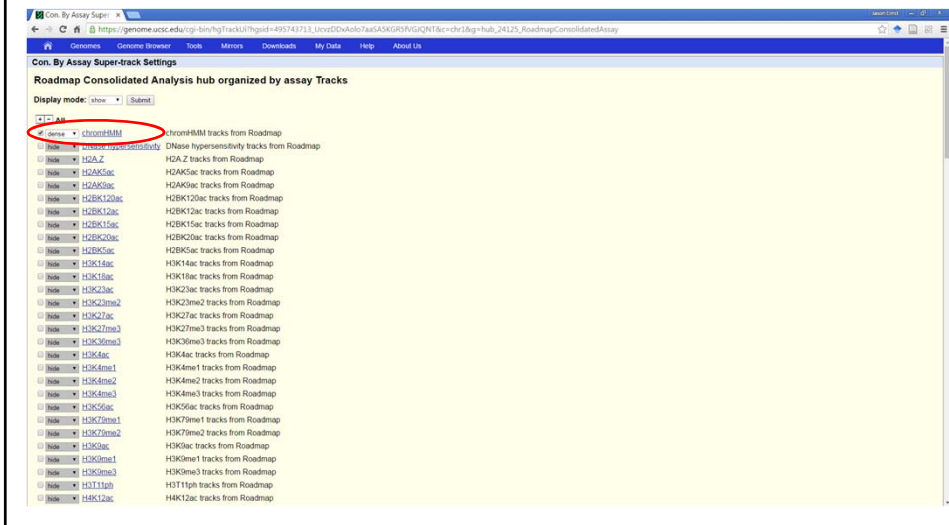
# Accessing Roadmap/ENCODE2 ChromHMM through the UCSC Genome Browser Track Hubs

The screenshot shows the UCSC Genome Browser Gateway interface. The 'Human Assembly' dropdown menu is circled in red. The 'Find Position' section shows the 'Human Assembly' selected. The page displays a list of species and a search bar for finding genomic positions.

# Accessing Roadmap/ENCODE2 ChromHMM through the UCSC Genome Browser Track Hubs



# Accessing Roadmap/ENCODE2 ChromHMM through the UCSC Genome Browser Track Hubs



## Accessing Roadmap/ENCODE2 ChromHMM through the UCSC Genome Browser Track Hubs

The screenshot shows the 'chromHMM Track Settings' page in the UCSC Genome Browser. The page title is 'chromHMM tracks from Roadmap (+Gen. By Assay)'. Below the title, there are several dropdown menus for 'Maximum display mode', 'Select views', 'AuxiliaryHMM', and 'PrimaryHMM'. A red circle highlights the 'Data Type' dropdown menu, which is currently set to 'Real' and has 'Imputed' as an alternative option. Below these settings is a table with columns for 'Sample Type', 'Views', 'AuxiliaryHMM', 'PrimaryHMM', and 'ImputedHMM'. The table lists various cell lines and tissues, such as HepG2, CD4+ CD25- IL17- PMA-IONOMYCIN stimulated MAC3 purified Th Primary Cells, Lungs, Ovary, Thymus, IMEC, CD4 Naive Primary cells, CD4+ CD25- CD45RA+ Naive Primary Cells, CD4 Naive Primary cells, CD4 Memory Primary cells, Mobilized CD4 Primary cells, CD3 Primary cells, Stomach Mucosa, CD19 Primary cells, Penis Foreskin Fibroblast Primary Cells, and Penis Foreskin Melanocyte Primary Cells. Each row has checkboxes for the different track types.

## Accessing Roadmap/ENCODE2 ChromHMM through the UCSC Genome Browser Track Hubs

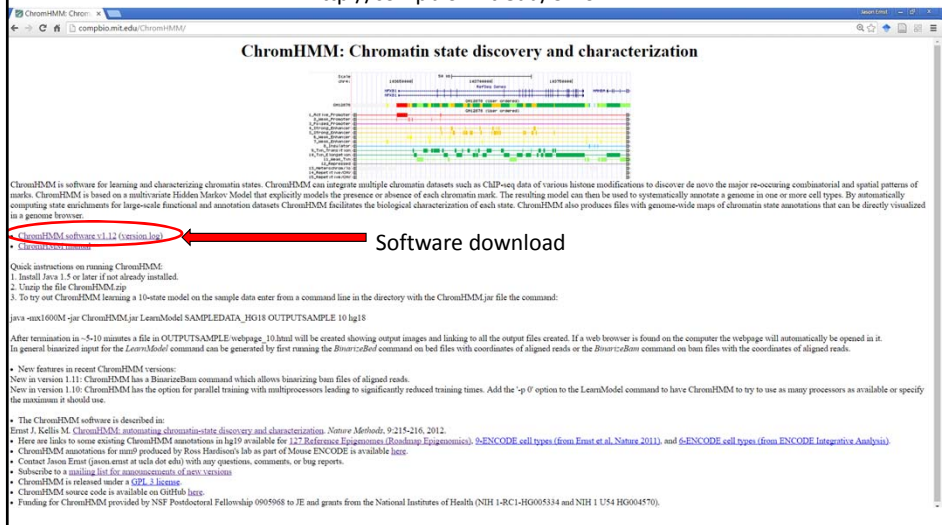
The screenshot shows a ChromHMM heatmap in the UCSC Genome Browser. The browser title is 'UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly'. The URL indicates the region is 'chr1:53,000,000-54,200,000 (600,001 bp)'. The heatmap displays signal intensity for various cell lines and tissues, with a color scale from green (low) to yellow (high). The cell lines listed on the left include HepG2, CD4+ CD25- IL17- PMA-IONOMYCIN stimulated MAC3 purified Th Primary Cells, Lungs, Ovary, Thymus, IMEC, CD4 Naive Primary cells, CD4+ CD25- CD45RA+ Naive Primary Cells, CD4 Naive Primary cells, CD4 Memory Primary cells, Mobilized CD4 Primary cells, CD3 Primary cells, Stomach Mucosa, CD19 Primary cells, Penis Foreskin Fibroblast Primary Cells, and Penis Foreskin Melanocyte Primary Cells. The heatmap shows a strong signal (yellow) for HepG2 and CD4+ CD25- IL17- PMA-IONOMYCIN stimulated MAC3 purified Th Primary Cells, and a weaker signal (green) for other cell lines.

# 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-occurring combinatorial and spatial patterns of marks. ChromHMM is based on a multi-state 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.

**Software download**

- [ChromHMM software v1.12 \(version log\)](#)
- [ChromHMM manual](#)

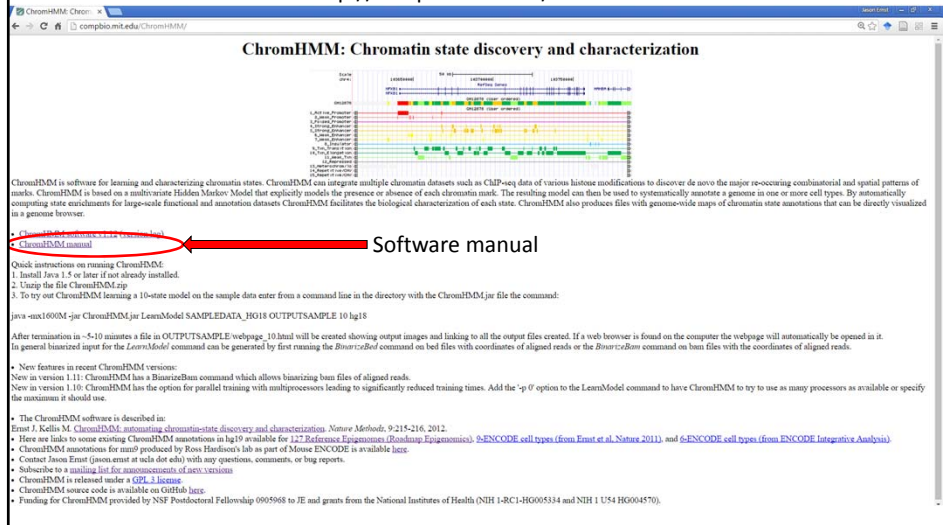
Quick instructions on running ChromHMM:  
1. Install Java 1.5 or later if not already installed.  
2. Unzip the file ChromHMM.zip  
3. To try out ChromHMM learning a 10-state model on the sample data enter from a command line in the directory with the ChromHMM.jar file the command:  
`java -mx1600M -jar 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. Its 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 features in recent ChromHMM versions:  
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` 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), 8-ENCODE cell types (from Ernst et al, Nature 2011), and 6-ENCODE cell types (from ENCODE Integrative Analysis).
- ChromHMM annotations for mm9 produced by Ross Hardison's lab as part of Mouse ENCODE is available here.
- Contact Jason Ernst (jason.ernst@tchd.tch.harvard.edu) with any questions, comments, or bug reports.
- Subscribe to a mailing list for announcements of new versions.
- ChromHMM is released under a [GPL 2 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-HG005534 and NIH 1 U54 HG004570).

# ChromHMM Website

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



The screenshot shows the ChromHMM website interface. At the top, the title is "ChromHMM: Chromatin state discovery and characterization". Below the title is a visualization of chromatin states across a genomic region. A navigation menu is located below the visualization, with "Software manual" highlighted by a red arrow. The menu items are: "ChromHMM software manual", "ChromHMM manual", "Quick instructions on running ChromHMM", "New features in recent ChromHMM versions", "New in version 1.11: ChromHMM has a BrowserizeBam command which allows browserizing 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 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. Chromatin state discovery and characterization. *Nature Methods*. 9:215-216, 2012.", "Here are links to some existing ChromHMM annotations in hg19 available for 12.2 Reference Epigenomes (Roadmap Epigenomics), 6-ENCODE cell types (from Ernst et al. Nature 2011), and 6-ENCODE cell types (from ENCODE Integrative Analysis).", "ChromHMM annotations for mmpf produced by Ross Hardison's lab as part of Mouse ENCODE is available here.", "Contact Jason Ernst (jason.ernst@mit.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 should already be installed on your computer)

1. Download

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

2. Unzip ChromHMM.zip

3. Open a command line

4. Change into the ChromHMM directory

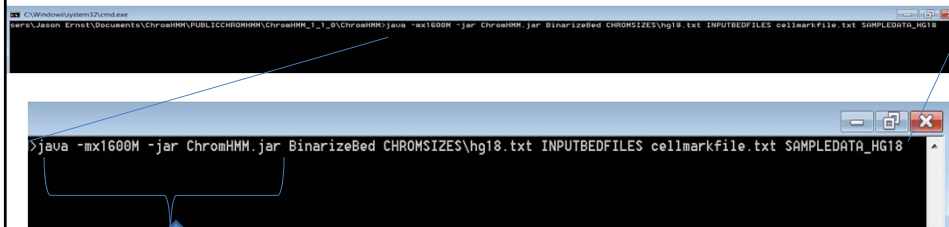
5. Enter the command:

```
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
C:\Users\Jason.Ernst\Documents\ChromHMM\PUBLIC\ChromHMM\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
sera\Jason_Ernst\Documents\ChromHMM\PUBLIC\ChromHMM\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
sera\Jason_Ernst\Documents\ChromHMM\PUBLIC\ChromHMM\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
C:\Users\Jason.Ernst\Documents\ChromHMM\PUBLIC\ChromHMM\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
C:\Users\Jason.Ernst\Documents\ChromHMM\PUBLIC\ChromHMM\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
```

cell1	mark1	cell11_mark1.bed	cell11_control.bed
cell1	mark2	cell11_mark2.bed	cell11_control.bed
cell2	mark1	cell12_mark1.bed	cell12_control.bed
cell2	mark2	cell12_mark2.bed	cell12_control.bed
cell	mark	cell-mark	

Cell-mark-file table

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

## BinarizeBed

```
C:\Windows\system32\cmd.exe
C:\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 OUTPUTSAMPLE 10 hg18
```

```
java -mx1600M -jar ChromHMM.jar LearnModel -p 0 SAMPLEDATA_HG18 OUTPUTSAMPLE 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\PUBLICCHR0HMM\ChromHMM_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 with the Binarized Input

# LearnModel

```
C:\Windows\system32\cmd.exe
C:\Users\Jason Ernst\Documents\ChromHMM\PUBLICCHR0HMM\ChromHMM_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\PUBLIC\CHR0HMM\ChromHMM_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

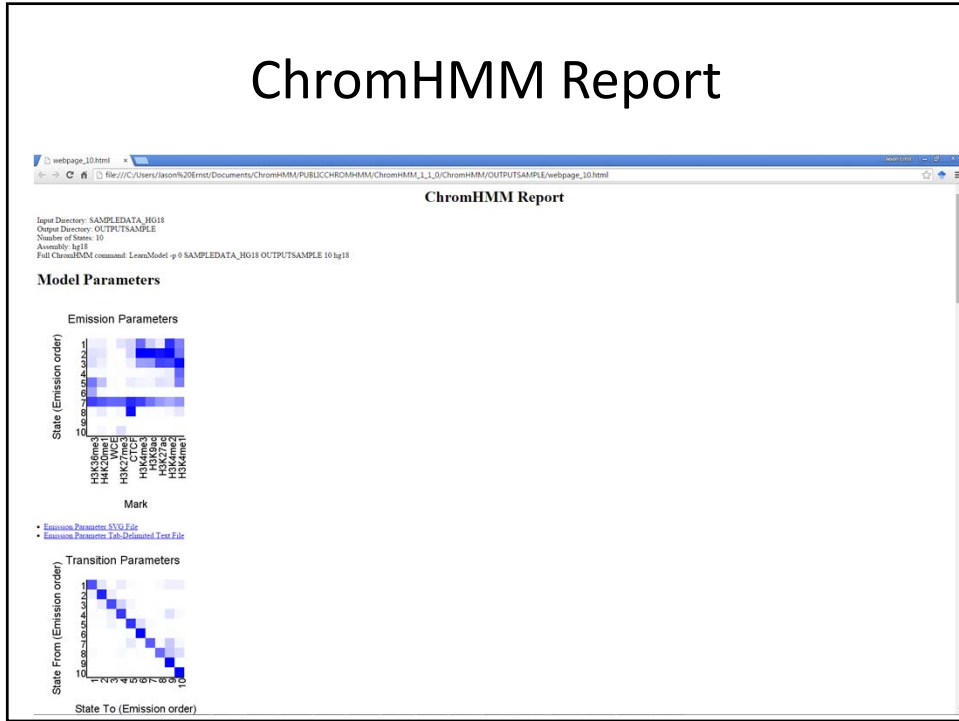
# LearnModel

```
C:\Windows\system32\cmd.exe
C:\Users\Jason Ernst\Documents\ChromHMM\PUBLIC\CHR0HMM\ChromHMM_1_0\ChromHMM>java -mx1600M -jar ChromHMM.jar LearnModel -p 0 SAMPLEDATA_HG18 OUTPUTSAMPLE 10 hg18
```

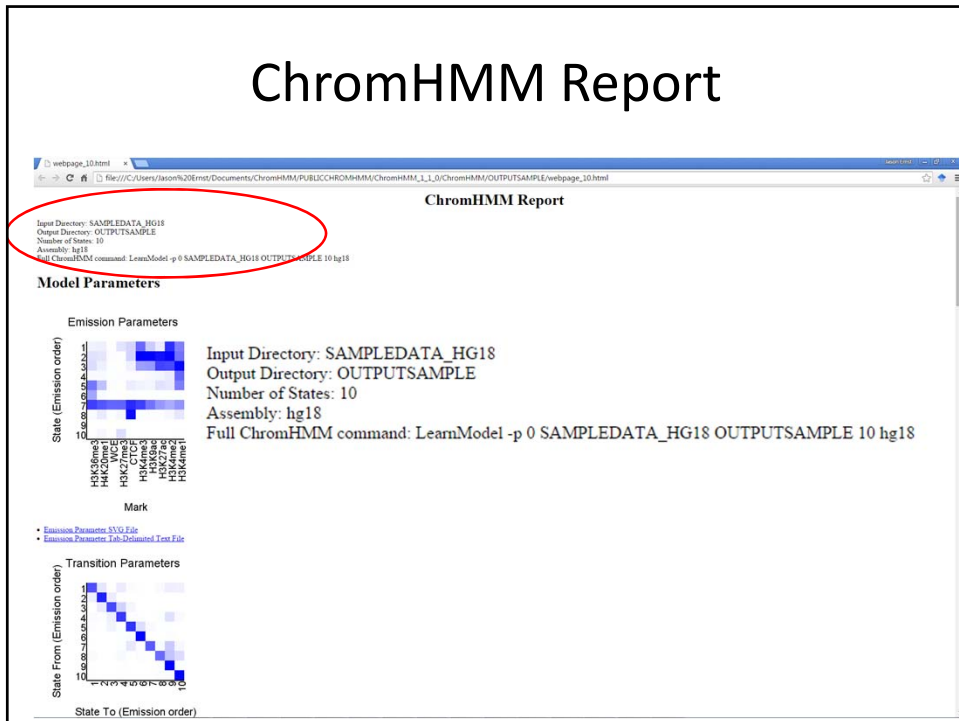
```
java -mx1600M -jar ChromHMM.jar LearnModel -p 0 SAMPLEDATA_HG18 OUTPUTSAMPLE 10 hg18
```

Genome assembly

# ChromHMM Report

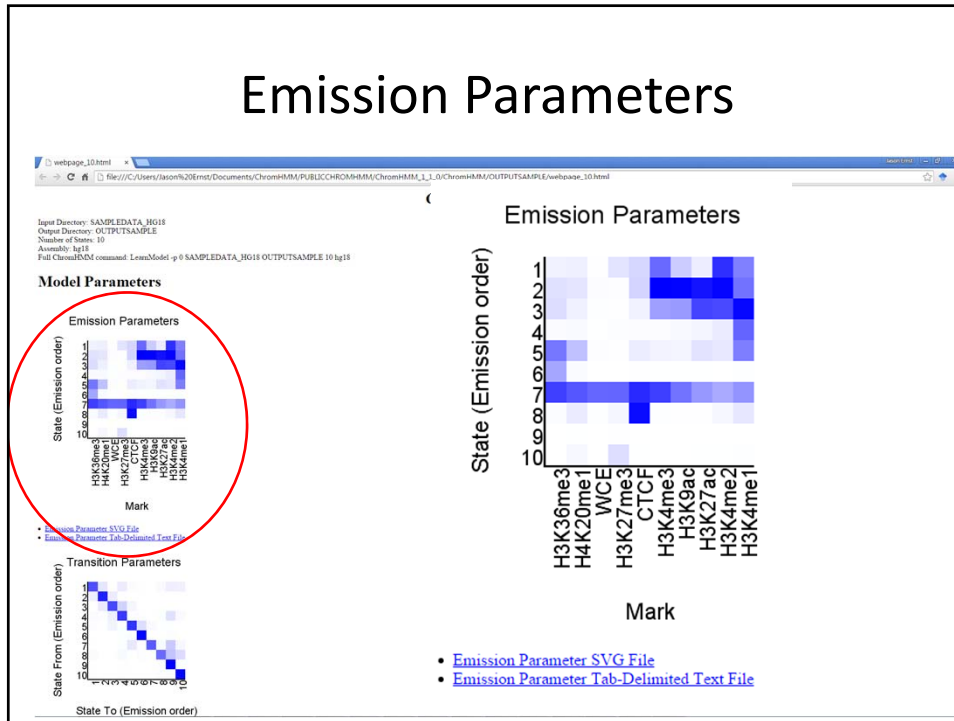


# ChromHMM Report

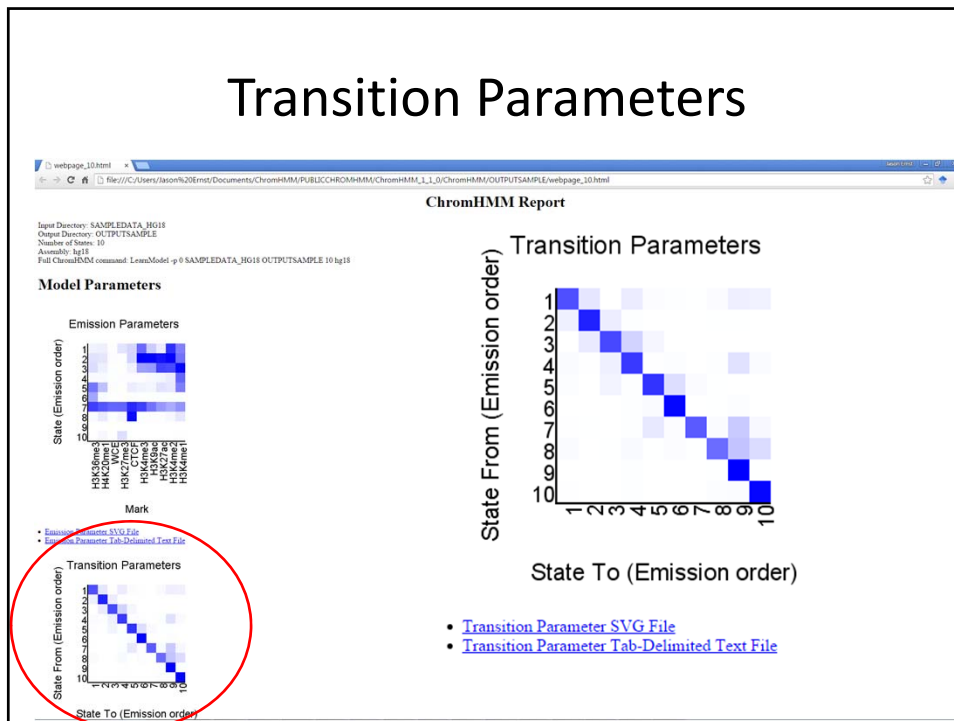




# Emission Parameters



# Transition Parameters



# Model Parameter File

webpage\_10.html

File Edit Format View Help

file:///C:/Users/jason%20Emst/Documents/ChromHM/PUBLIC/CHROMHM/ChromHM\_1\_1\_0/ChromHM/OUTPUTSAMPLE/webpage\_10.html

- [Transaction Parameters SVG File](#)
- [Transaction Parameters 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 Default File](#)
- [GM12878\\_10 Browser Custom Track Expanded File](#)
- [K562\\_10 Browser Custom Track Default File](#)
- [K562\\_10 Browser Custom Track Expanded File](#)

### State Enrichments

#### GM12878\_10 Enrichments

Fold Enrichment GM12878\_10

State (Emission order)

Category

- [GM12878\\_10 Overlay Enrichment SVG File](#)
- [GM12878\\_10 Overlay Enrichment Tab-Delimited Text File](#)

# Segmentation File

webpage\_10.html

File Edit Format View Help

file:///C:/Users/jason%20Emst/Documents/ChromHM/PUBLIC/CHROMHM/ChromHM\_1\_1\_0/ChromHM/OUTPUTSAMPLE/webpage\_10.html

- [Transaction Parameters SVG File](#)
- [Transaction Parameters 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 Default File](#)
- [GM12878\\_10 Browser Custom Track Expanded File](#)
- [K562\\_10 Browser Custom Track Default File](#)
- [K562\\_10 Browser Custom Track Expanded File](#)

### State Enrichments

#### GM12878\_10 Enrichments

Fold Enrichment GM12878\_10

State (Emission order)

Category

- [GM12878\\_10 Overlay Enrichment SVG File](#)
- [GM12878\\_10 Overlay Enrichment Tab-Delimited Text File](#)

```
GM12878_10.segment.bed - Notepad
chr11 9 58000 58000 E9
chr11 58000 58000 E9
chr11 58400 61400 E9
chr11 61400 61800 E8
chr11 61800 64800 E9
chr11 64800 67600 E4
chr11 67600 70600 E9
chr11 70600 73600 E8
chr11 73600 76600 E9
chr11 76600 79600 E8
chr11 79600 82600 E9
chr11 82600 85600 E10
chr11 85600 88600 E10
chr11 88600 91600 E9
chr11 91600 94600 E1
chr11 94600 97600 E10
chr11 97600 100600 E9
chr11 100600 103600 E8
chr11 103600 106600 E9
chr11 106600 109600 E8
chr11 109600 112600 E8
chr11 112600 115600 E9
chr11 115600 118600 E7
chr11 118600 121600 E9
chr11 121600 124600 E9
chr11 124600 127600 E9
chr11 127600 130600 E9
chr11 130600 133600 E9
chr11 133600 136600 E9
chr11 136600 139600 E9
chr11 139600 142600 E9
chr11 142600 145600 E9
chr11 145600 148600 E9
chr11 148600 151600 E9
chr11 151600 154600 E9
chr11 154600 157600 E9
chr11 157600 160600 E9
chr11 160600 163600 E9
chr11 163600 166600 E9
chr11 166600 169600 E9
chr11 169600 172600 E9
chr11 172600 175600 E9
chr11 175600 178600 E9
chr11 178600 181600 E9
chr11 181600 184600 E9
chr11 184600 187600 E9
chr11 187600 190600 E9
chr11 190600 193600 E9
chr11 193600 196600 E9
chr11 196600 199600 E9
chr11 199600 202600 E9
chr11 202600 205600 E9
chr11 205600 208600 E9
chr11 208600 211600 E9
chr11 211600 214600 E9
chr11 214600 217600 E9
chr11 217600 220600 E9
chr11 220600 223600 E9
chr11 223600 226600 E9
chr11 226600 229600 E9
chr11 229600 232600 E9
chr11 232600 235600 E9
chr11 235600 238600 E9
chr11 238600 241600 E9
```

# Browser Files

webpage\_10.html

- Transition Parameters SVG File
- Transition Parameters 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 Downset File
- GM12878\_10 Browser Custom Track Expanded File
- K562\_10 Browser Custom Track Downset File
- K562\_10 Browser Custom Track Expanded File

### 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

Can load into browser UCSC Genome, IGV

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

chr11:18,444,593-20,155,602 (171,350 bp)

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

# Enrichments

webpage\_10.html

### 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

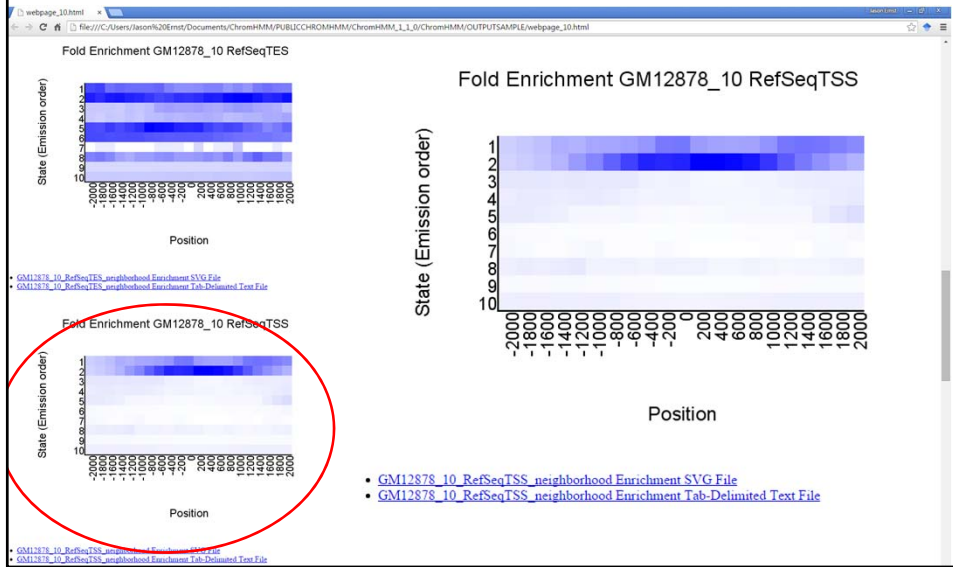
##### Fold Enrichment GM12878\_10 RefSeqTES

State (Emission order)

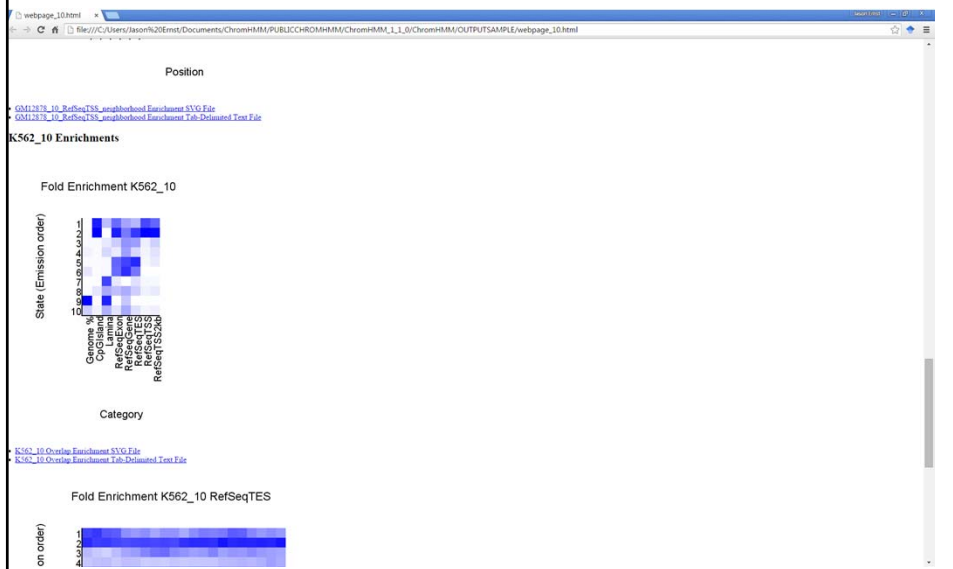
Category

- GM12878\_10 Overlap Enrichment SVG File
- GM12878\_10 Overlap Enrichment Tab-Delimited Text File

# Positional Plots

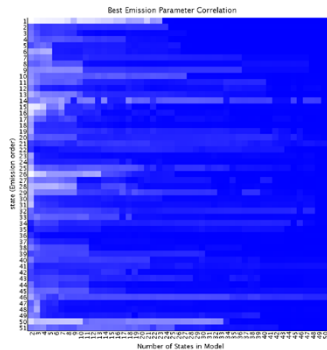


# Enrichments for Additional Cell Types



## 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.

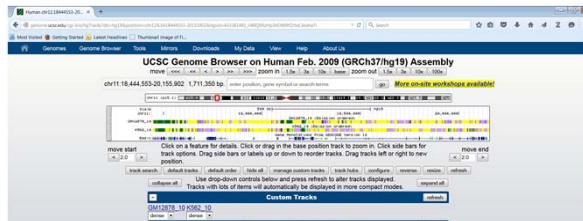


```
CompareModels [-color r,g,b] referencemodel comparedir outputprefix
```

## 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
```

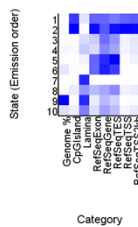


## Additional Commands

- **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
```

Fold Enrichment GM12878\_10

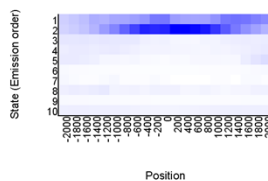


## Additional Commands

- **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
```

Fold Enrichment GM12878\_10 RefSeqTSS





## 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
```

## Summary

- Pre-computed ChromHMM chromatin state annotations available across over 100 cell/tissue types
- ChromHMM software available to run on your own data

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

## Collaborators and Acknowledgements

---

- Manolis Kellis

### ENCODE consortium

- Brad Bernstein production group



### Roadmap Epigenomics consortium



### Funding

- NHGRI, NIH, NSF, HHMI, Sloan Foundation