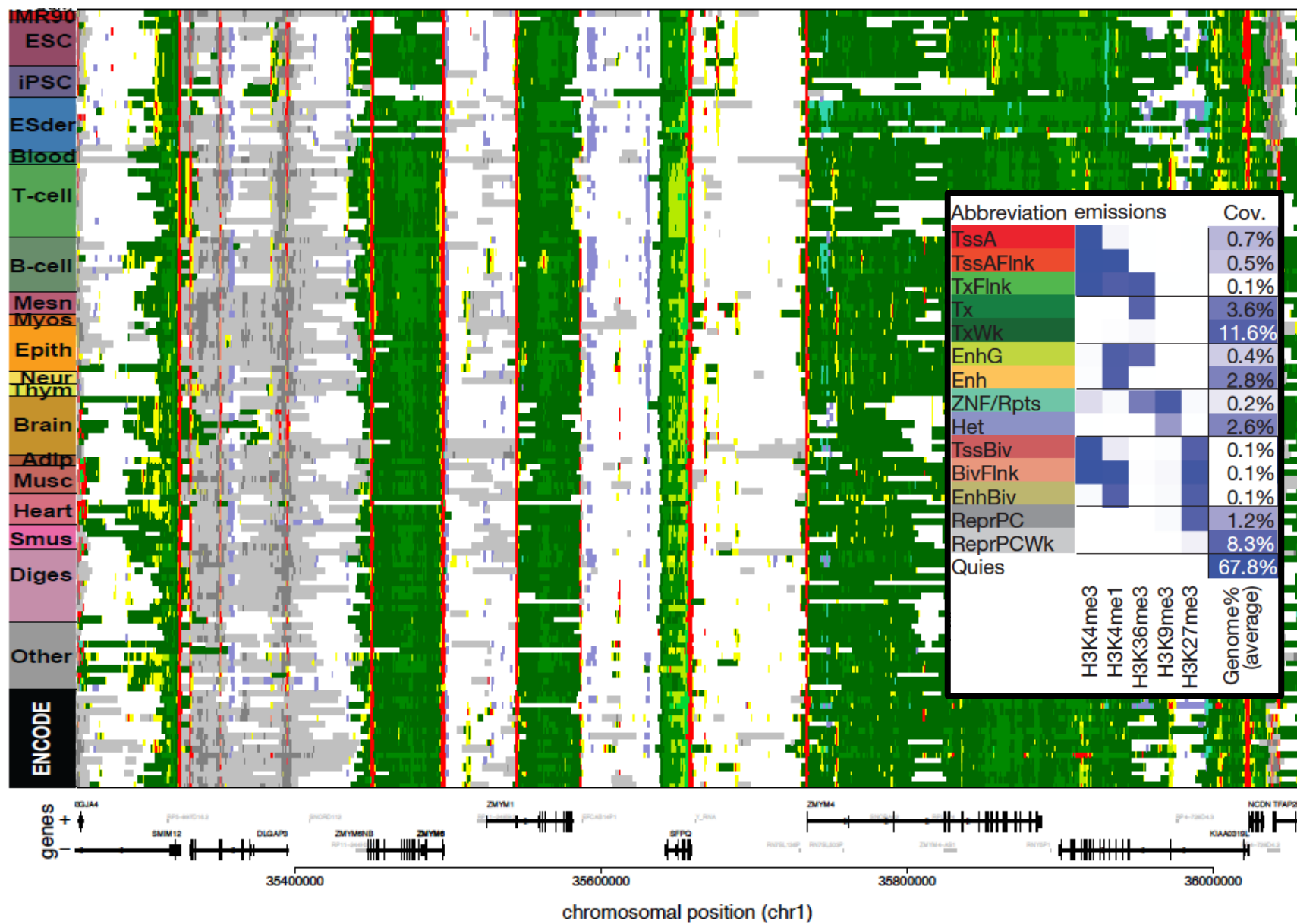
A decorative graphic consisting of a vertical cyan line and a horizontal cyan line intersecting at the center. The vertical line has a blue segment near the top and a yellow segment near the bottom. The horizontal line has a blue segment on the left, a red segment, and several yellow segments. The intersection area is filled with a dense, multi-colored pattern of green, yellow, and red lines.

# How to make (more) sense of chromatin state data

Wouter Meuleman

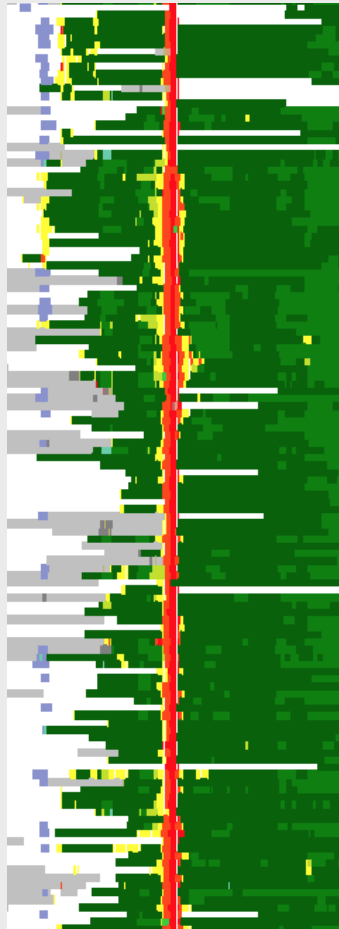
Kellis Lab (MIT/Broad)

epigenomes



Although a great summary, can still be hard to interpret / work with, especially as we're moving towards 1,000s of epigenomes.

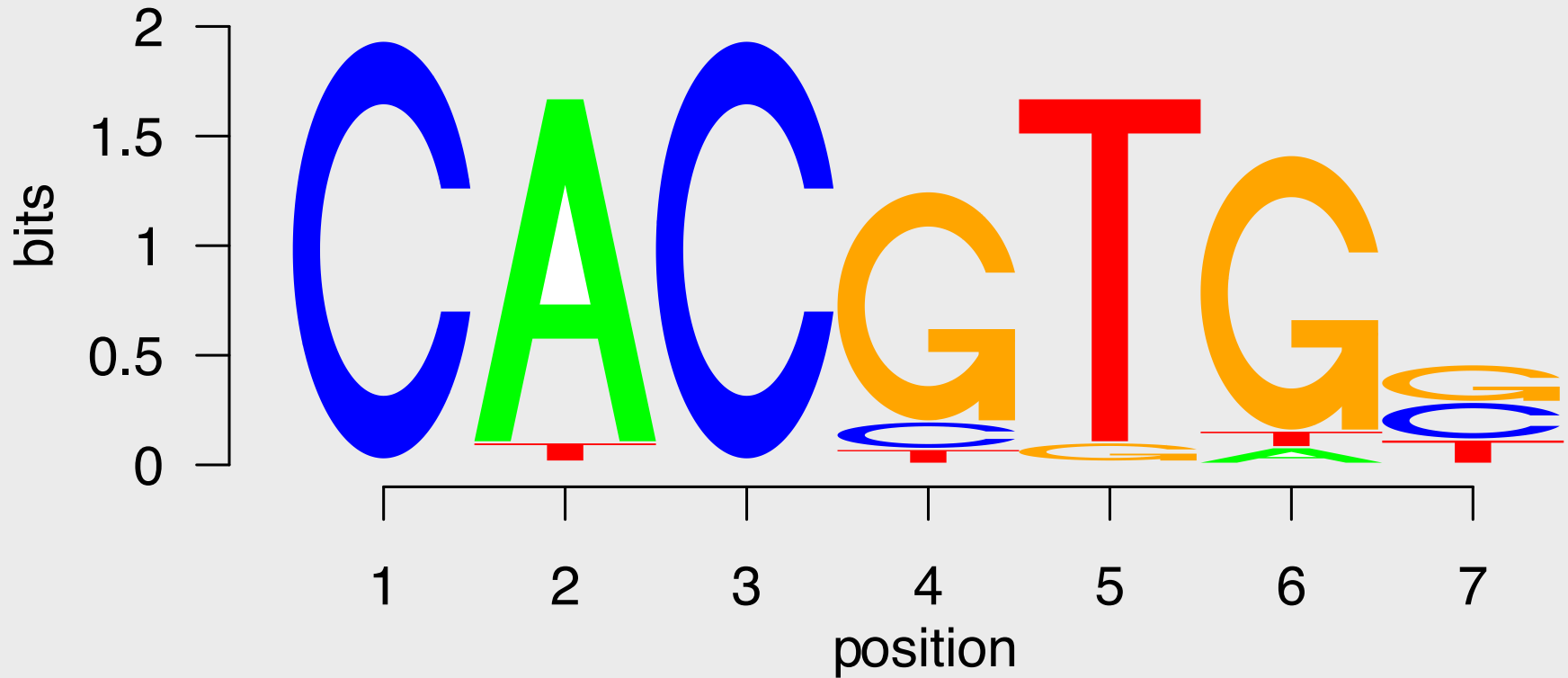
Chromatin state calls across many epigenomes can be viewed as an alignment of sequences with a finite alphabet



```
CTCTTAT  
CACGTGC  
CACGTGC  
CACGTGG  
CACGTGG  
CACGTGG  
CACGTGC  
CACGTGG  
CACGTGT  
CACGTGC  
CACGTGT  
CACGTGC  
CACCTGT  
CACCGTC  
CACGTGC  
CACGTGC  
CACGTGG  
CACGTGT  
CACGTGG  
CACGTGG  
CACGTGG
```

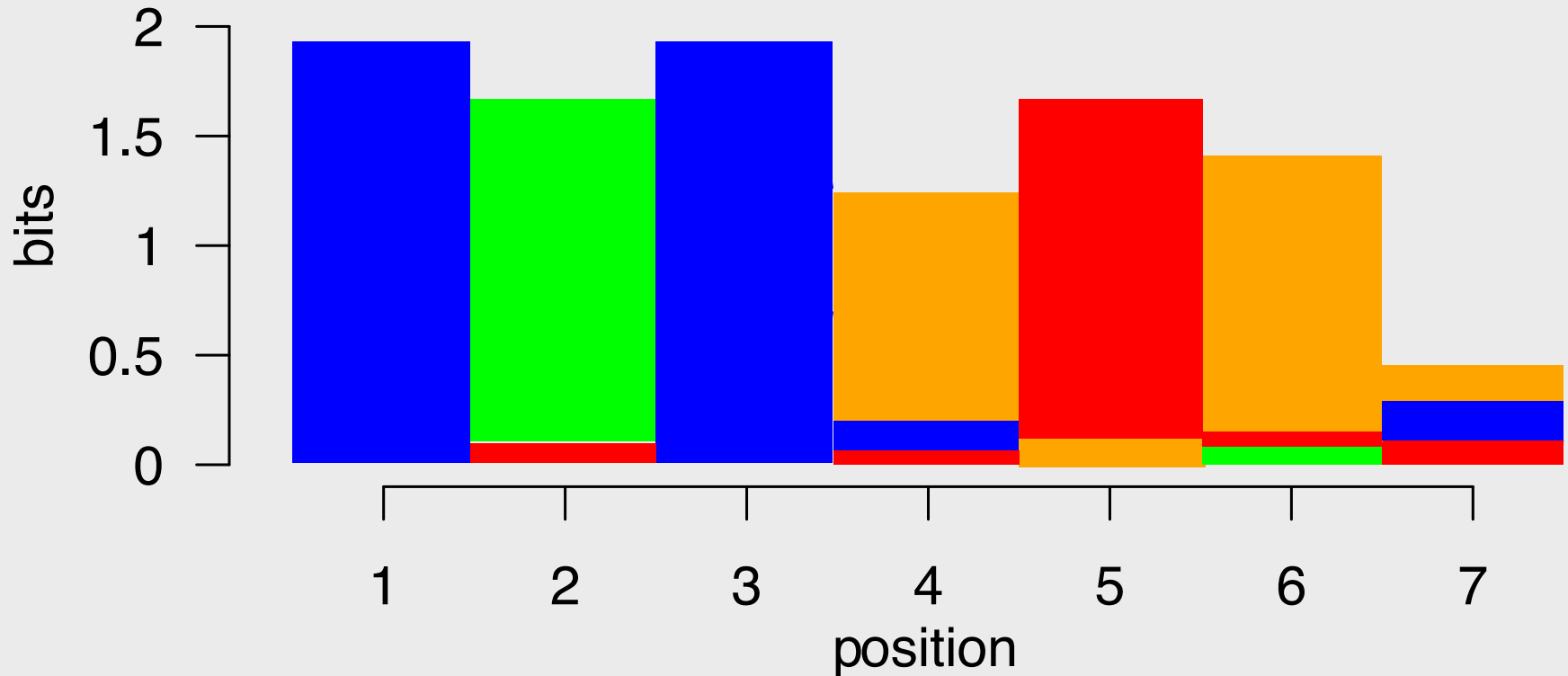
There are good ways of modeling such alignments

# Sequence logos!



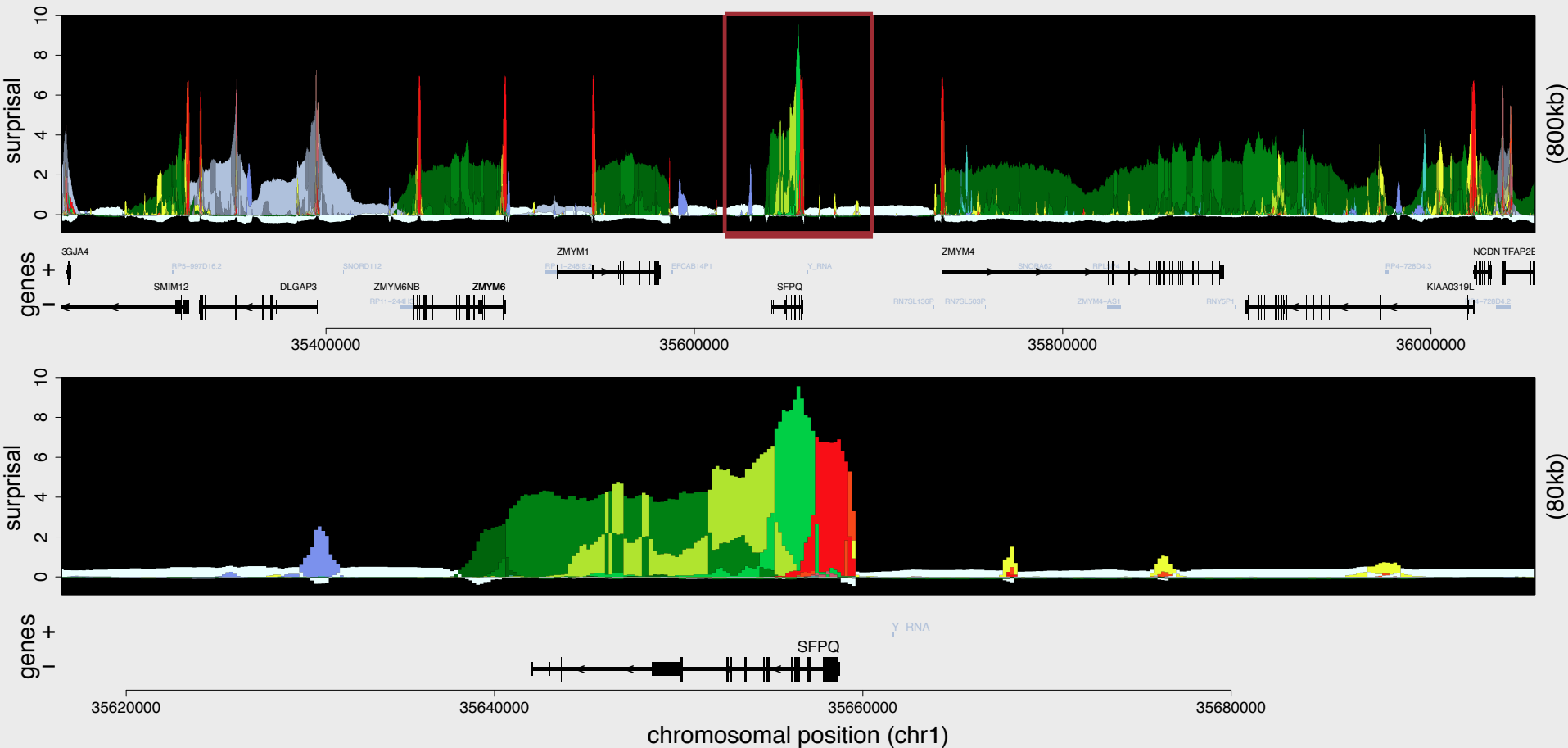
Instead of letters, let's consider an alphabet of colors only.

See where this is going?



Logos allow for the inclusion of background frequencies:  
the expected occurrence of each color can be different

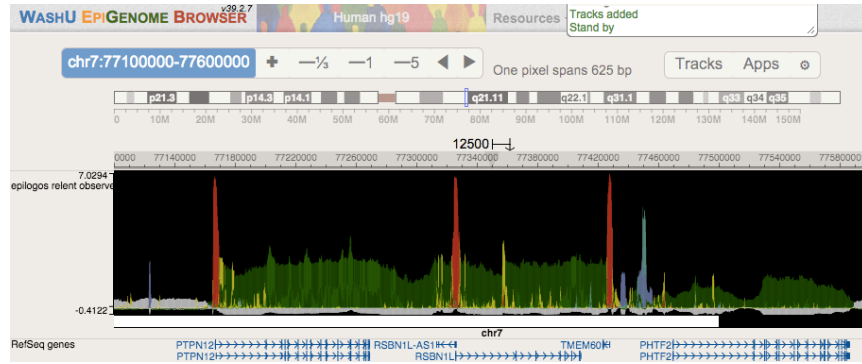
# epilogos



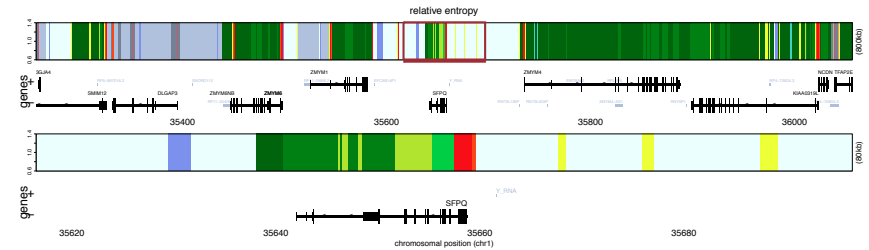
- *surprisal* indicates for each state/color how unlikely/enriched
- Sum at each position == relative entropy (Kullback-Leibler div.)

# Some applications of *epi*logos

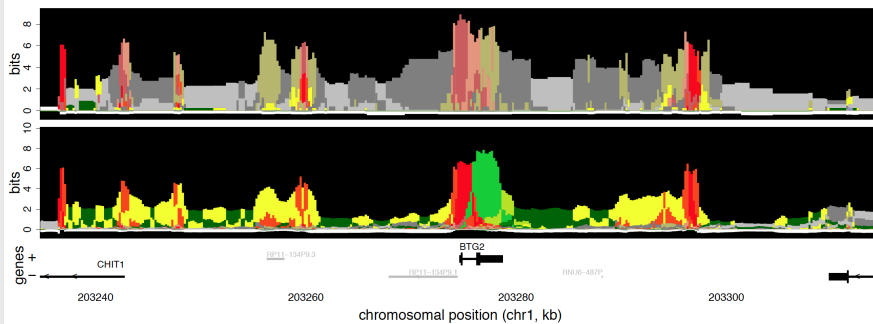
## interactive visualization



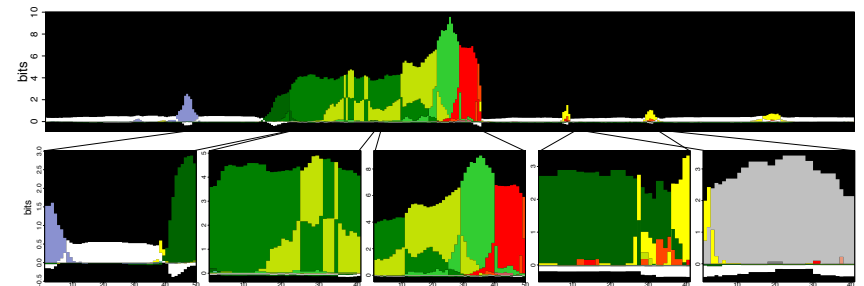
## consensus epigenomes



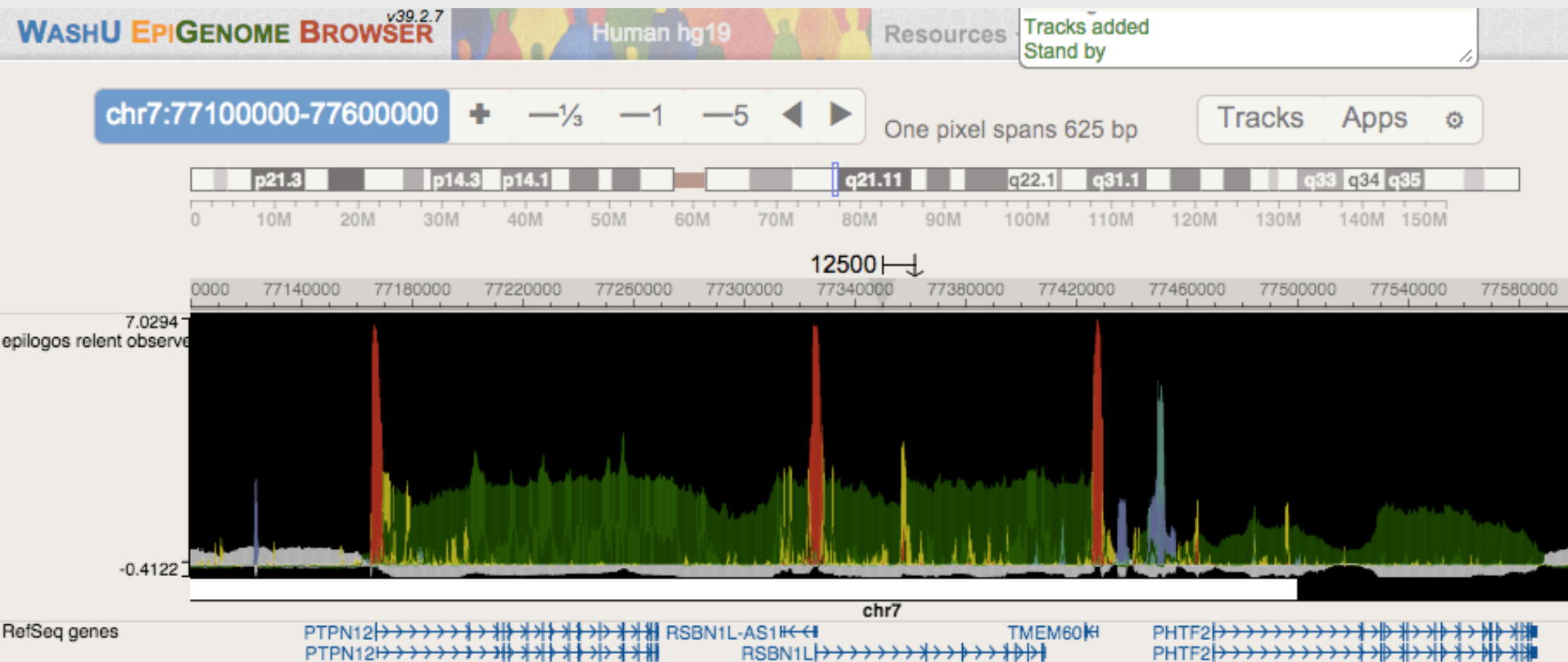
## comparative epigenomics



## pattern discovery & search



# Interactive visualization



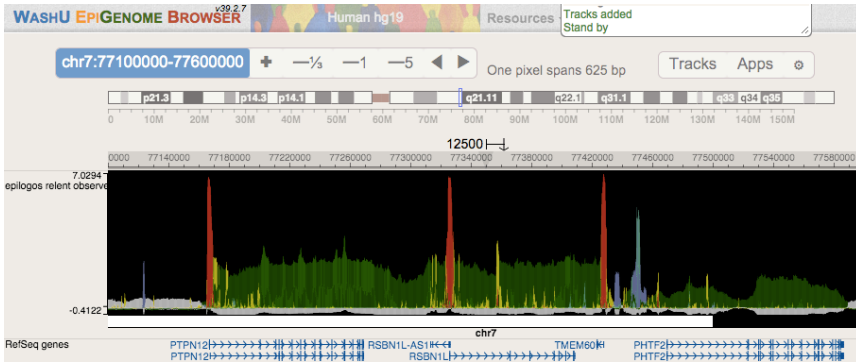
Visualization tracks for the WashU epigenome browser

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

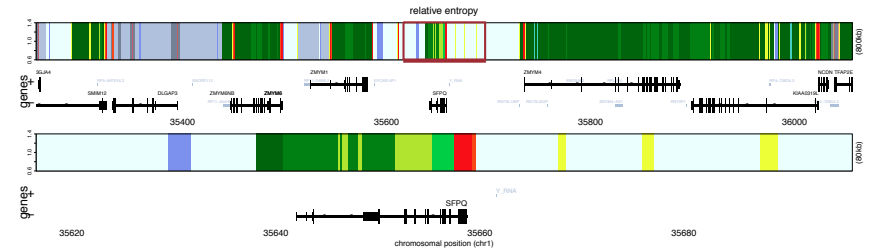
Stand-alone as well as embeddable in your own website

# Some applications of *epi*logos

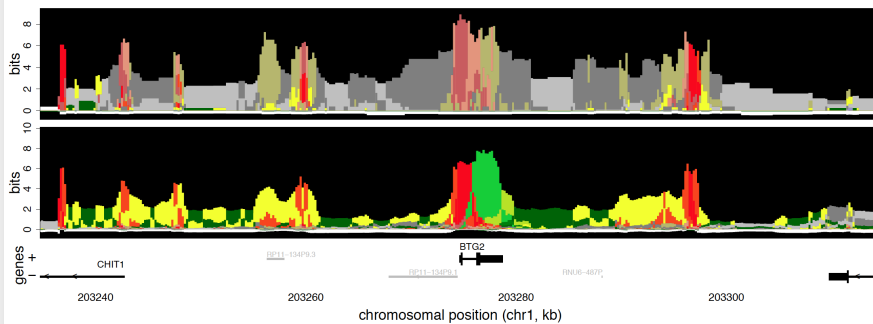
## ✓ interactive visualization



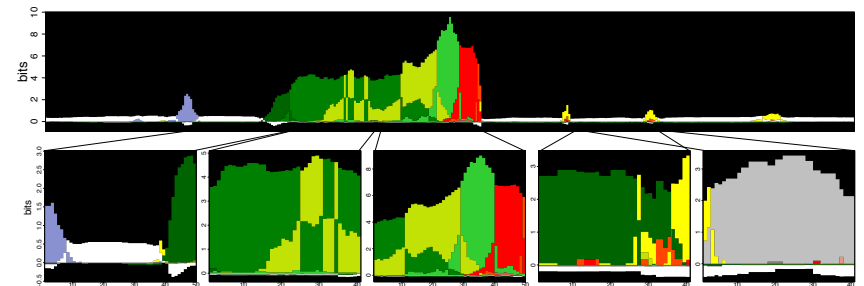
## consensus epigenomes



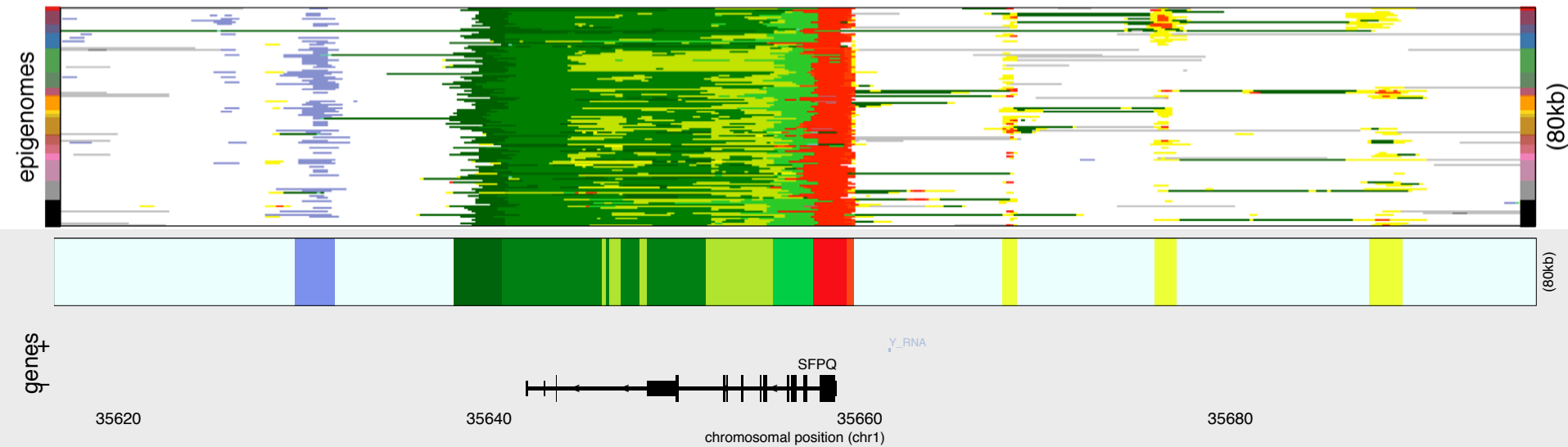
## comparative epigenomics



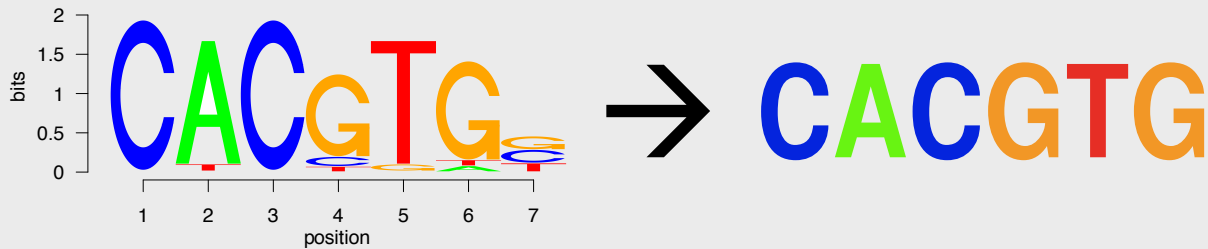
## pattern discovery & search



# Consensus epigenomes



For each position, the chromatin state with the highest information content forms the ‘consensus’.

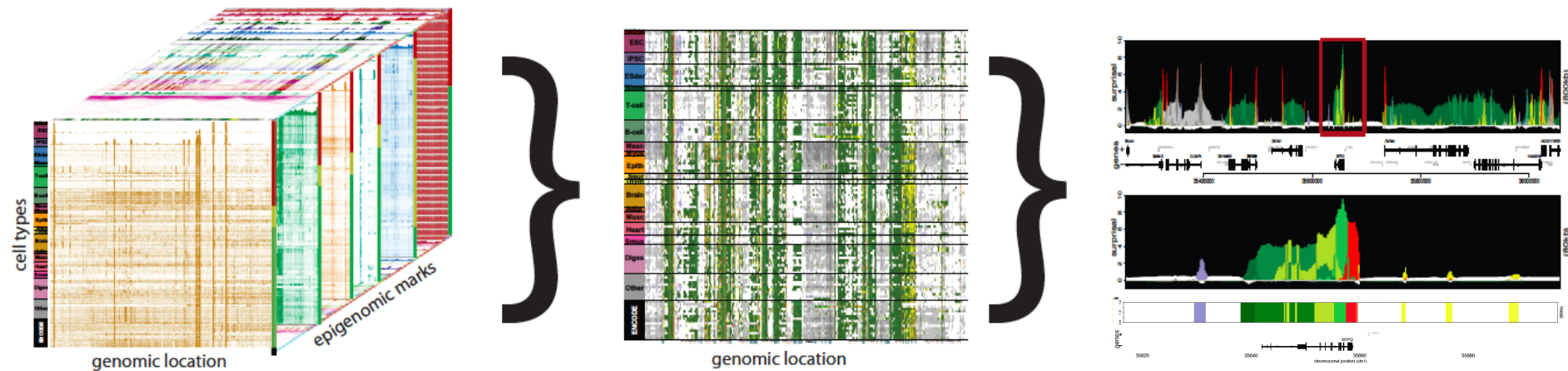


# Interpretation of large-scale (epi)genomic datasets through information-based dimensionality reduction

3D

2D

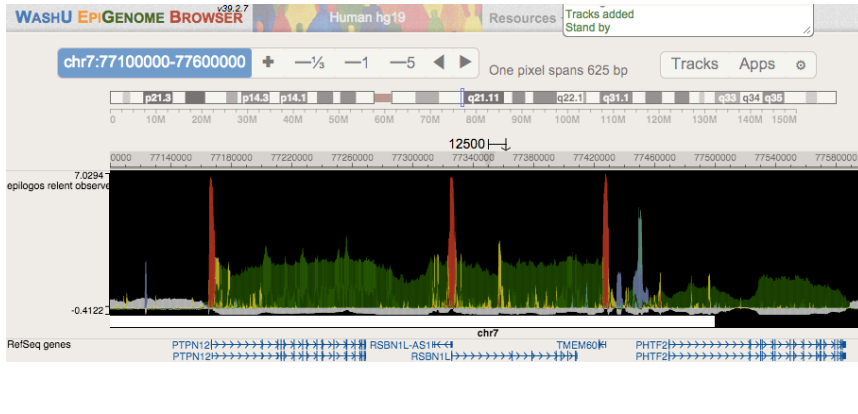
1D



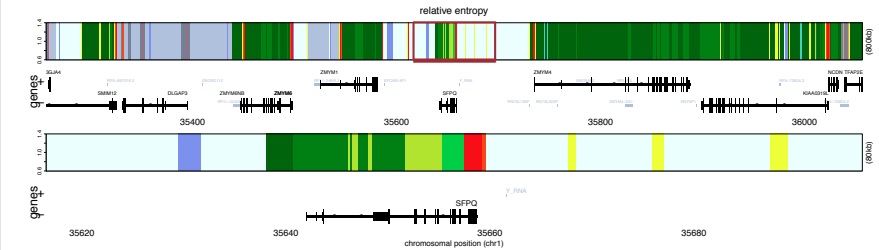
Thus far: mainly data representation/visualization.  
Can we also perform analyses using **epilogos**?

# Some applications of `epilogos`

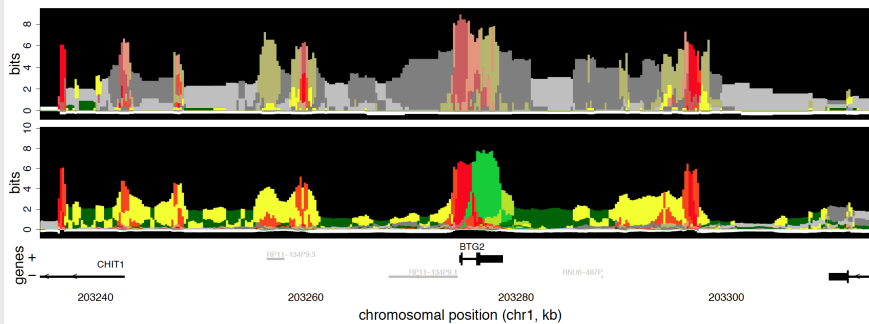
## ✓ interactive visualization



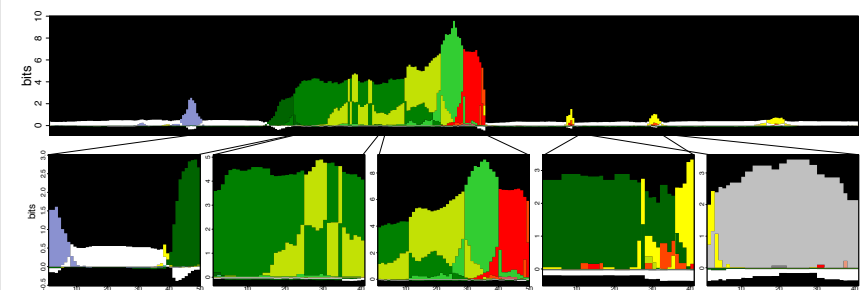
## ✓ consensus epigenomes



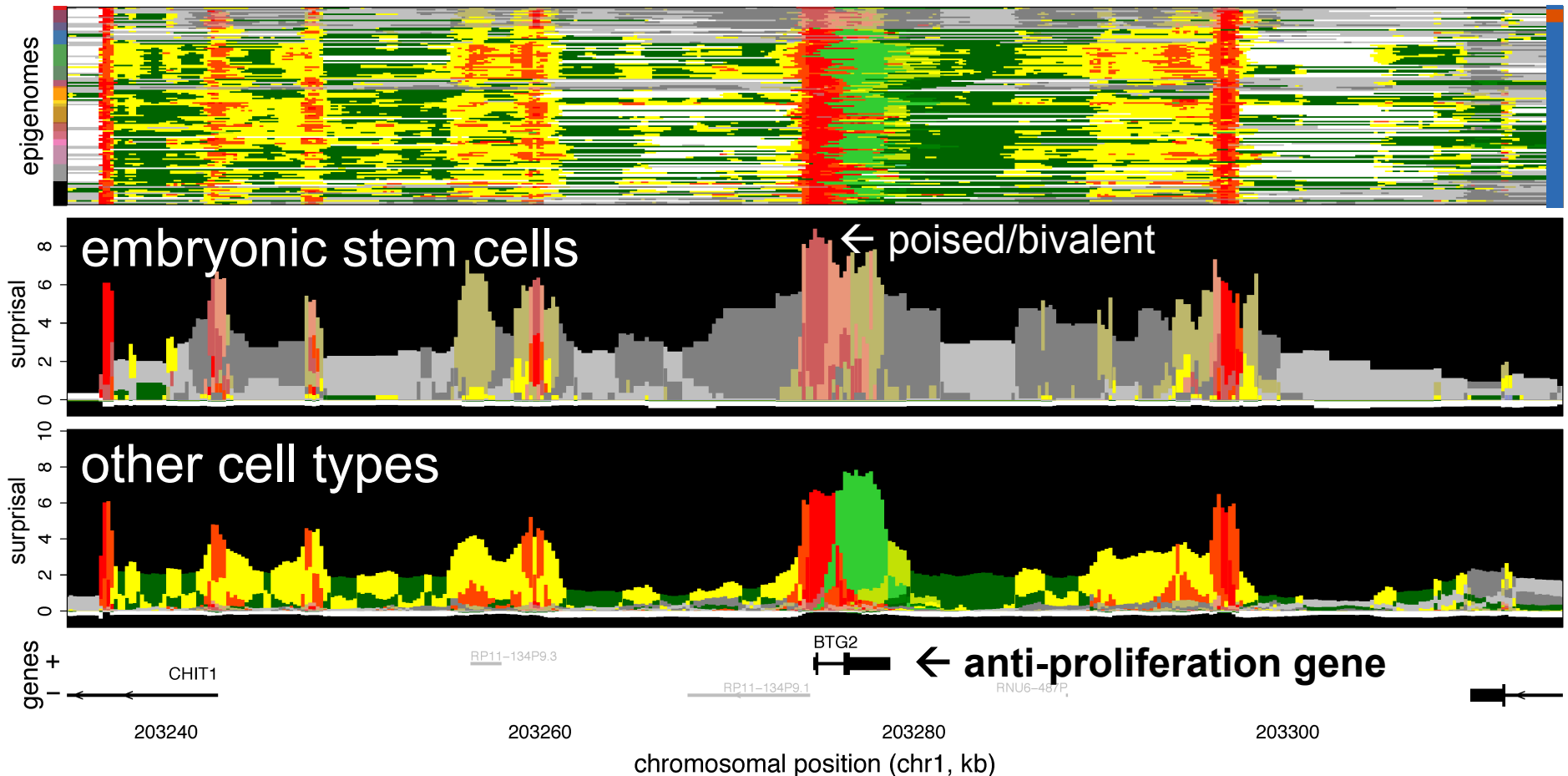
## comparative epigenomics



## pattern discovery & search



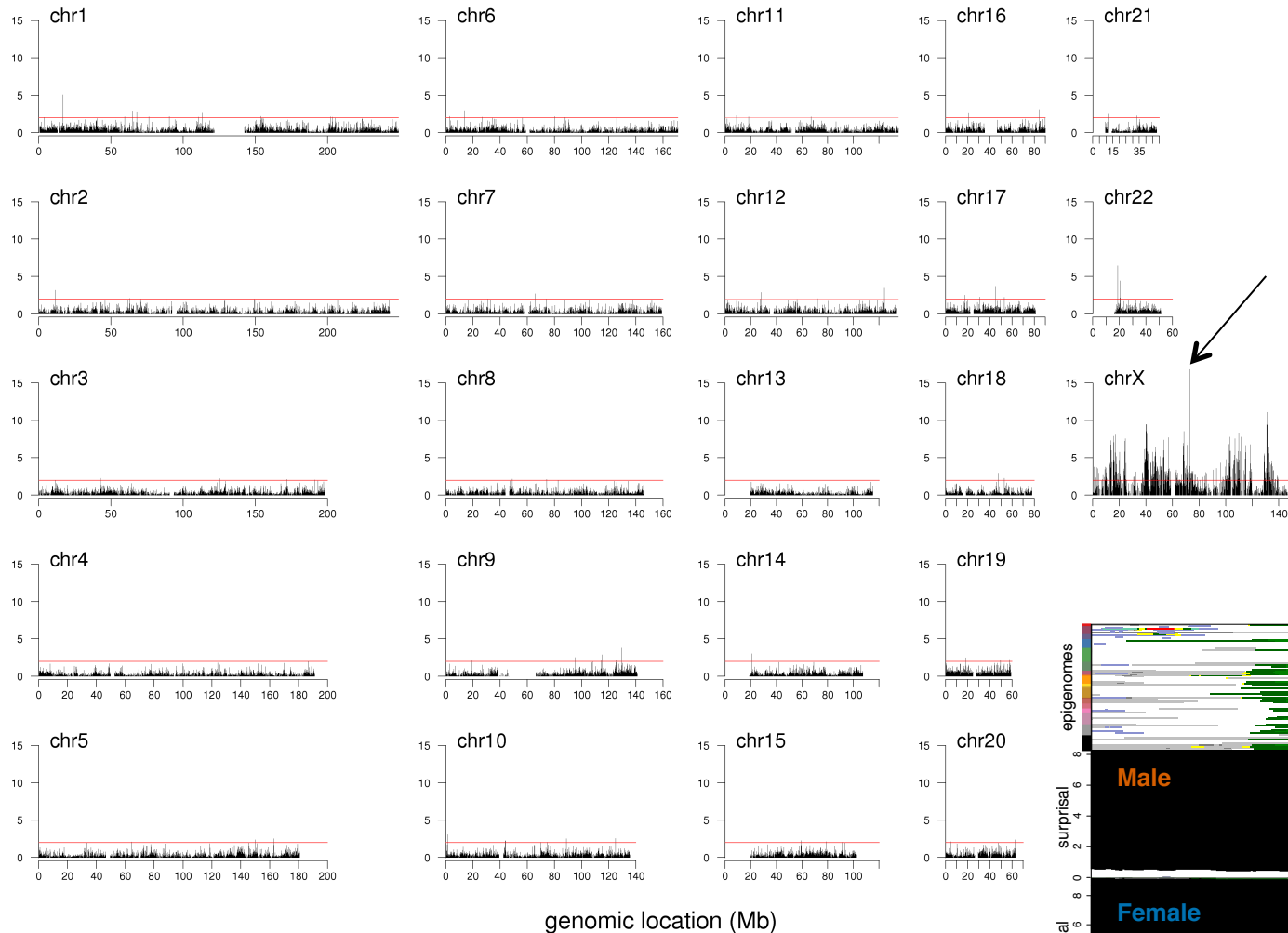
# Comparative epigenomics



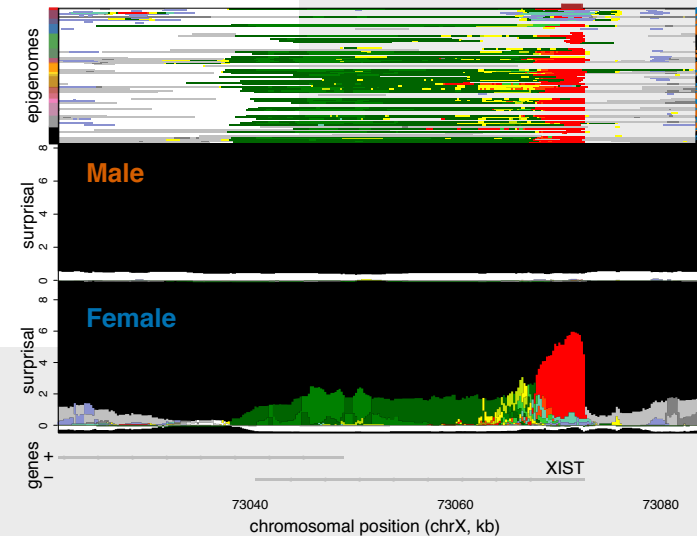
Select two subsets of epigenomes,  
build separate **epilogos** and compare them

# Application 2: Comparative epigenomics

magnitude of difference



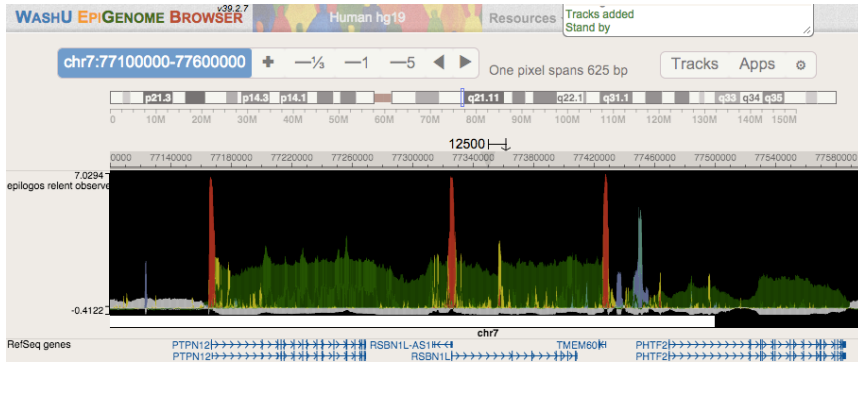
XIST



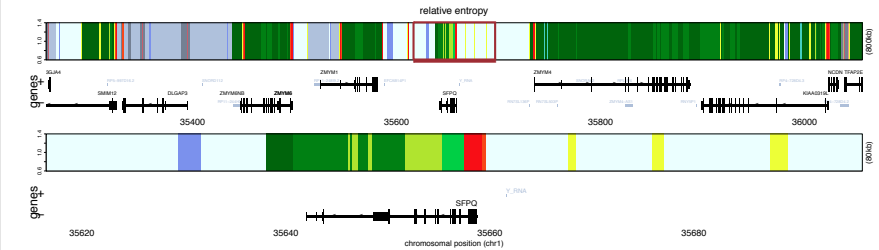
Comparison: Male vs. Female donors

# Some applications of **epilogos**

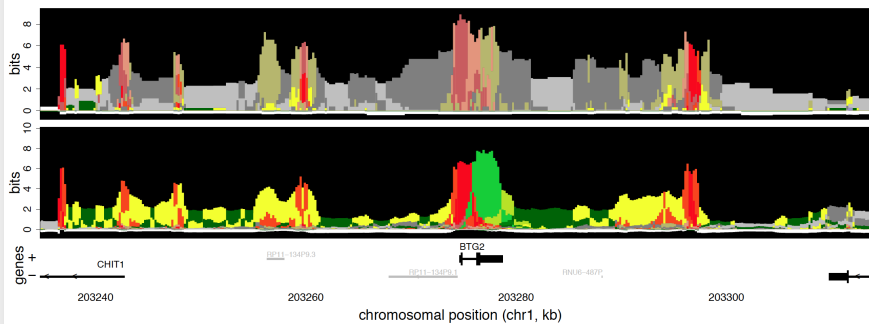
## ✓ interactive visualization



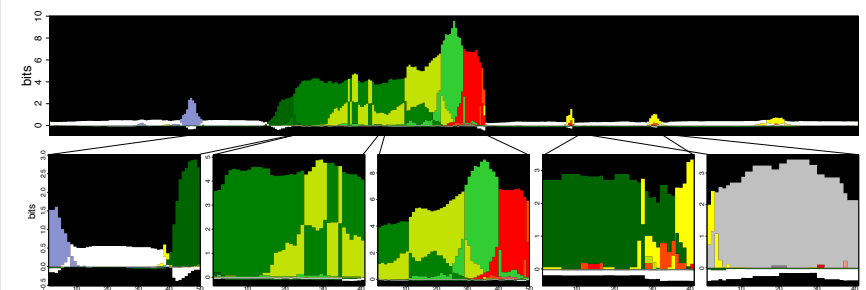
## ✓ consensus epigenomes



## ✓ comparative epigenomics

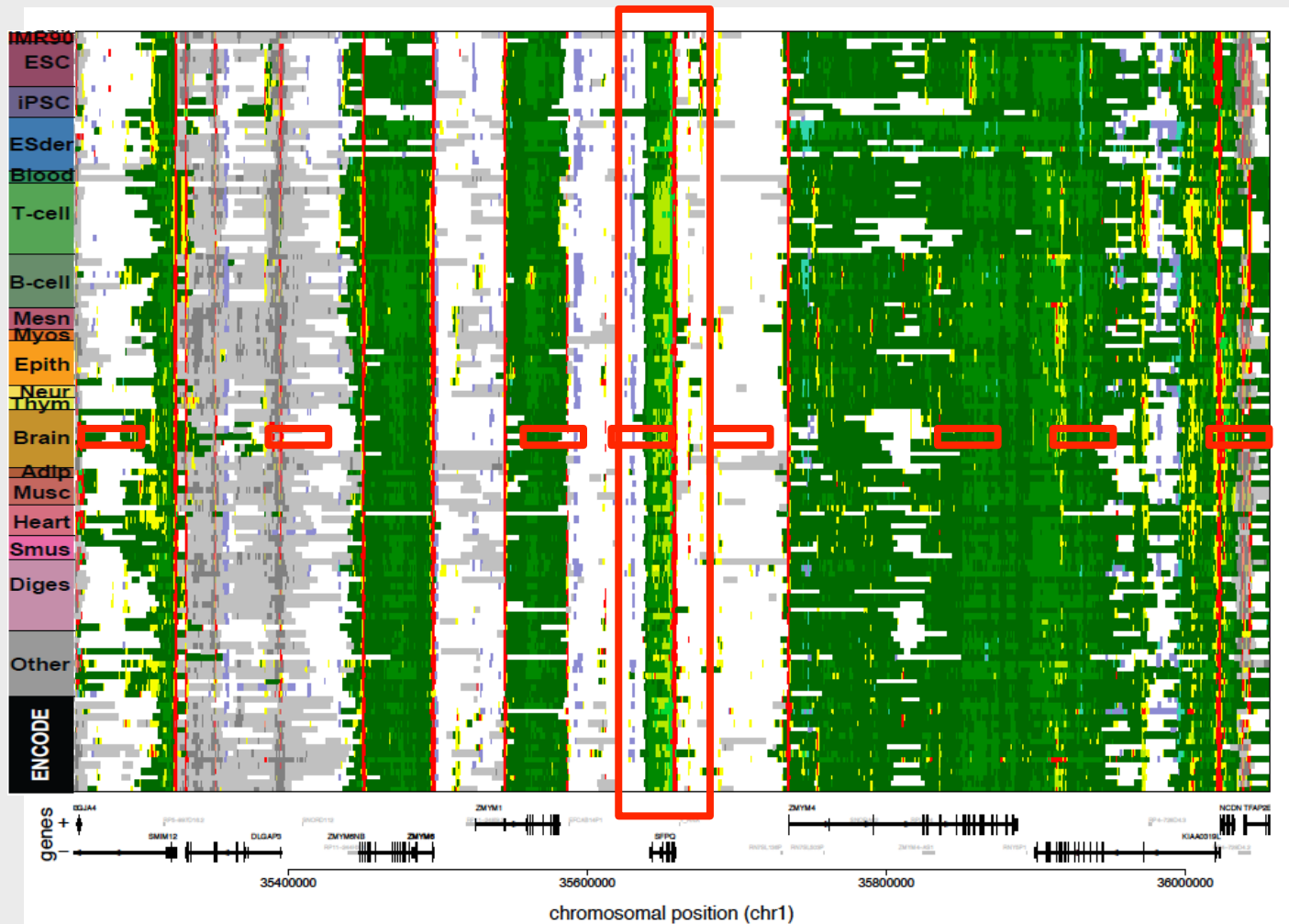


## pattern discovery & search

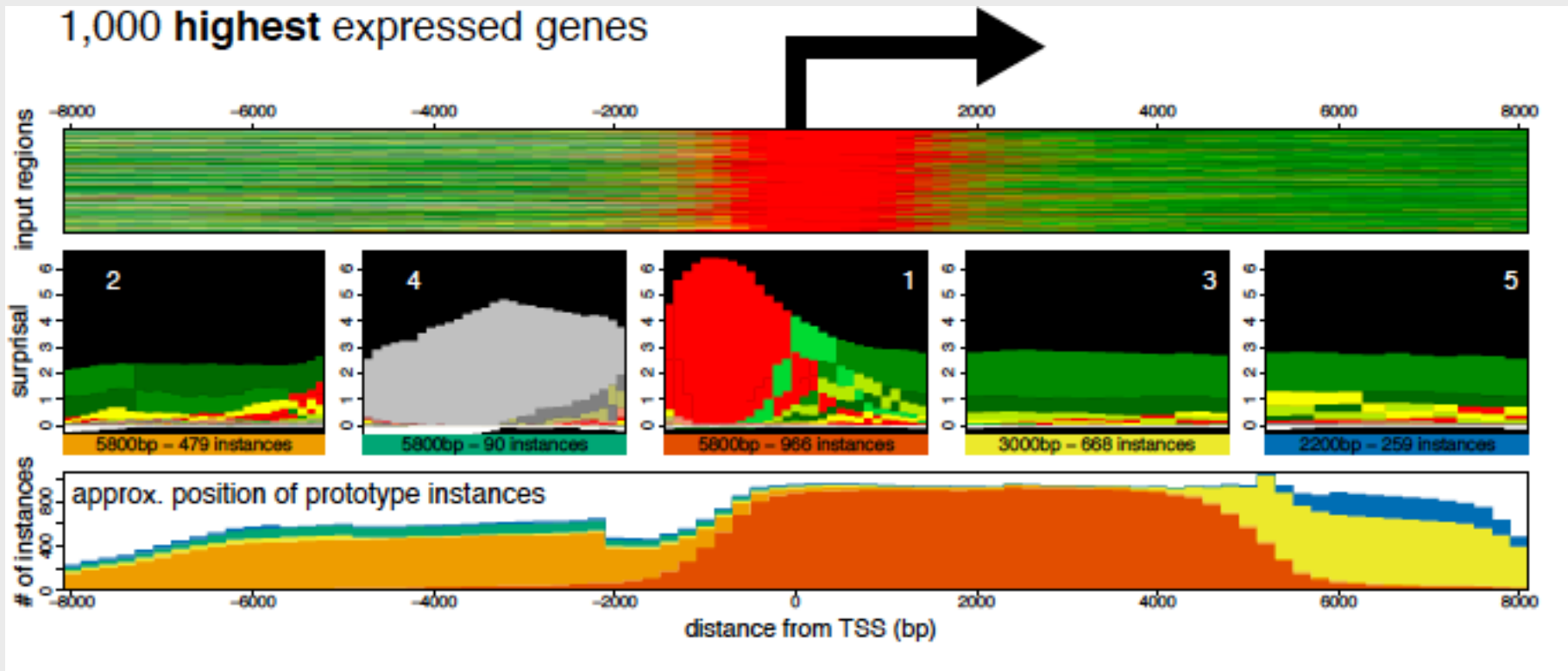


# Turning things upside down

In addition to studying single regions across many epigenomes, we can also use **epilogos** to study many regions in a single epigenome

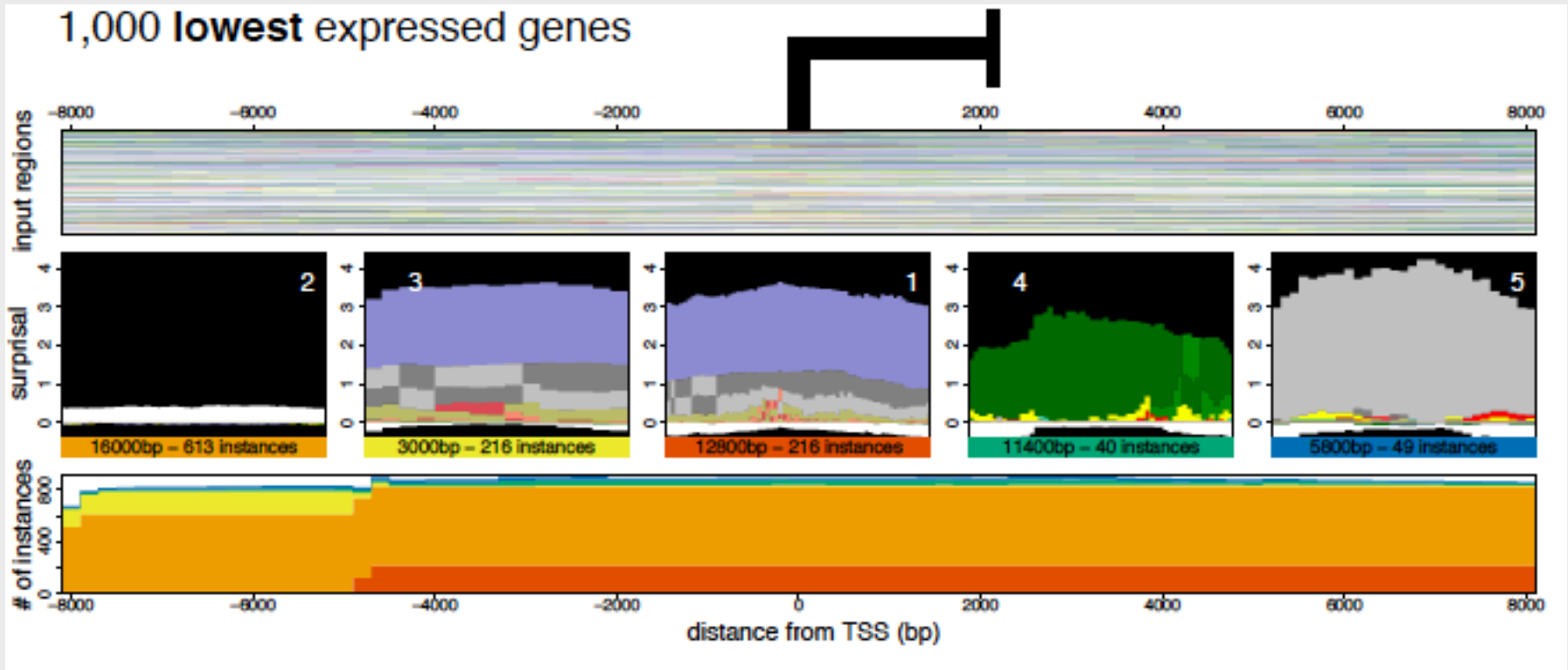


# Prototype discovery & search: active TSSs



- **Single epigenome, many regions**
- Use Gibbs sampling / EM to discover 'prototypes' for regions
- Regions do not have to be aligned around some landmark

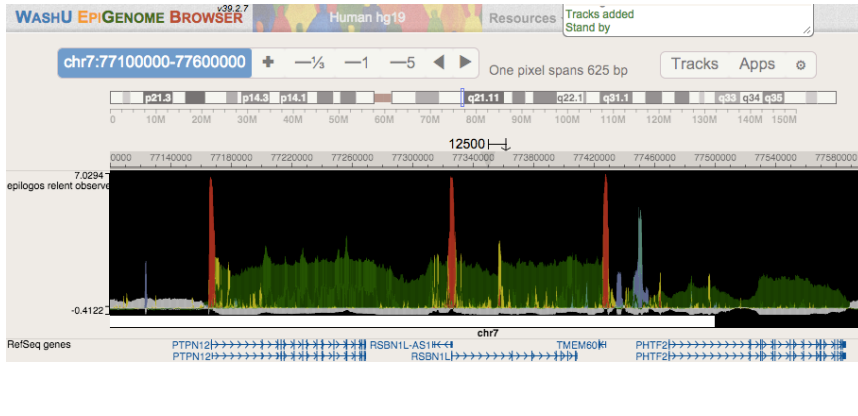
# Prototype discovery & search: inactive TSSs



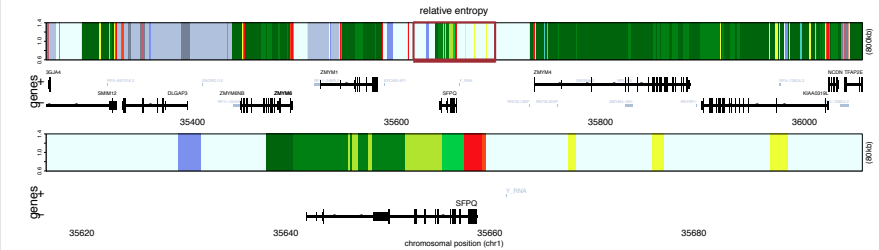
- Discovers multiple mechanisms of silencing (i.e., quiescent, heterochromatin, Polycomb)
- **Prototypes can be used to search in other epigenomes**
- Think of prototyping, e.g., enhancers in cell type X and finding instances in unrelated cell type Y

# Some applications of **epilogos**

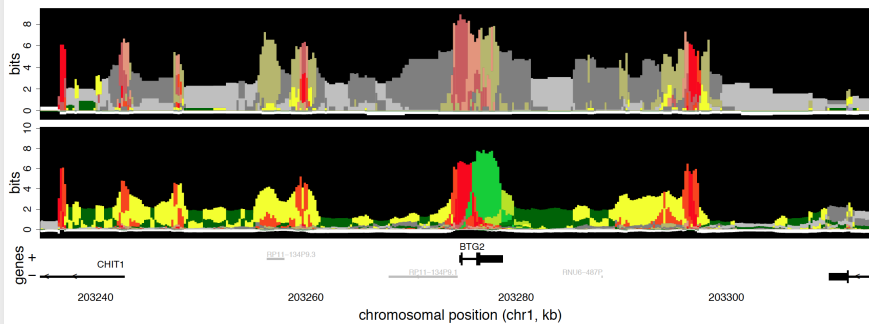
## ✓ interactive visualization



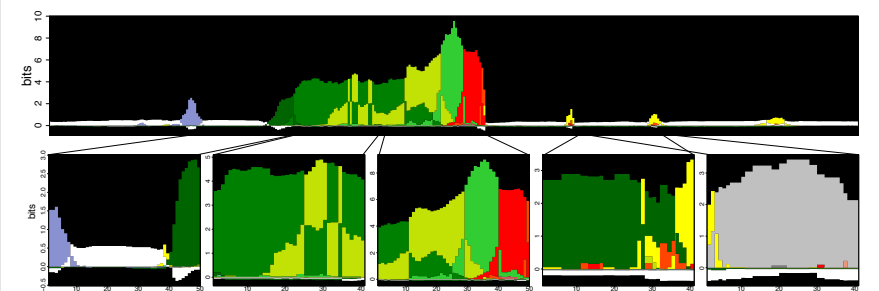
## ✓ consensus epigenomes



## ✓ comparative epigenomics



## ✓ pattern discovery & search



# epilogos

## Visualization and analysis of chromatin state models

- Information-based dimensionality reduction
- Comparative epigenomics
- Prototype discovery and search
  - Currently implementing search / scoring
- Use with custom data
  - Summer student will work on making this more accessible

**<http://compbio.mit.edu/epilogos>**



# THANKS!

- Soheil Feizi (information theory)
- Nezar Abdennur (TAD boundaries)
- Xin Zhou & Daofeng Li (WashU browser)
- Anshul Kundaje (ENCODE pipeline / Roadmap)
- Jin Wook Lee (Roadmap supplement site)
- Members of Kellis Lab (discussions / ideas)

**<http://compbio.mit.edu/epilogos>**

meuleman@mit.edu