


# Human Epigenome Browser

Xin Zhou, Ph.D.

Laboratory of Ting Wang  
Department of Genetics, Wash U

2010-11-10

# Background



## EPIGENETICS

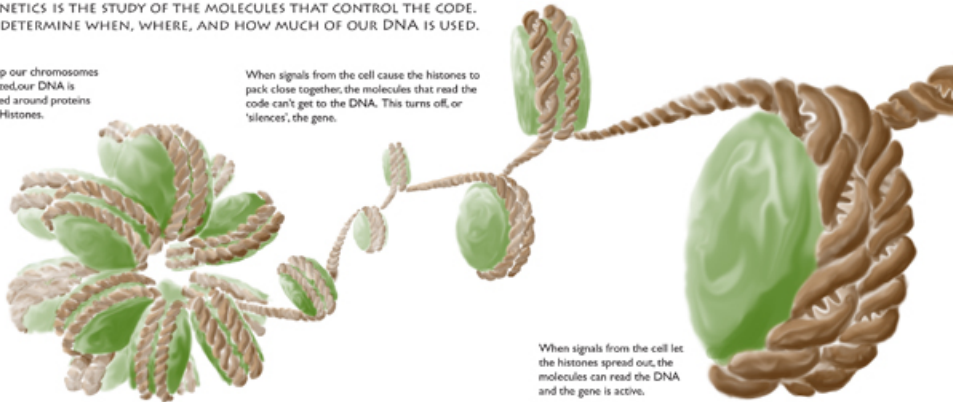
WHAT IT IS, AND WHY IT IS IMPORTANT TO YOU.

OUR DNA IS THE GENETIC CODE THAT MAKES EACH OF US. EVERY CELL IN OUR BODY HAS AN IDENTICAL COPY OF THE CODE. SO WHAT MAKES OUR SKIN CELLS DIFFERENT FROM OUR BONES?

EPIGENETICS IS THE STUDY OF THE MOLECULES THAT CONTROL THE CODE. THEY DETERMINE WHEN, WHERE, AND HOW MUCH OF OUR DNA IS USED.

To keep our chromosomes organized, our DNA is wrapped around proteins called Histones.

When signals from the cell cause the histones to pack close together, the molecules that read the code can't get to the DNA. This turns off, or 'silences', the gene.



When signals from the cell let the histones spread out, the molecules can read the DNA and the gene is active.

Epigenetics controls embryo development to form bone, muscle, and skin.

Our epigenetic controls change in response to our environment.

Only half of what we pass on is DNA. The other half are the controls we inherited and then modified during our lives.

## *Agouti* mice





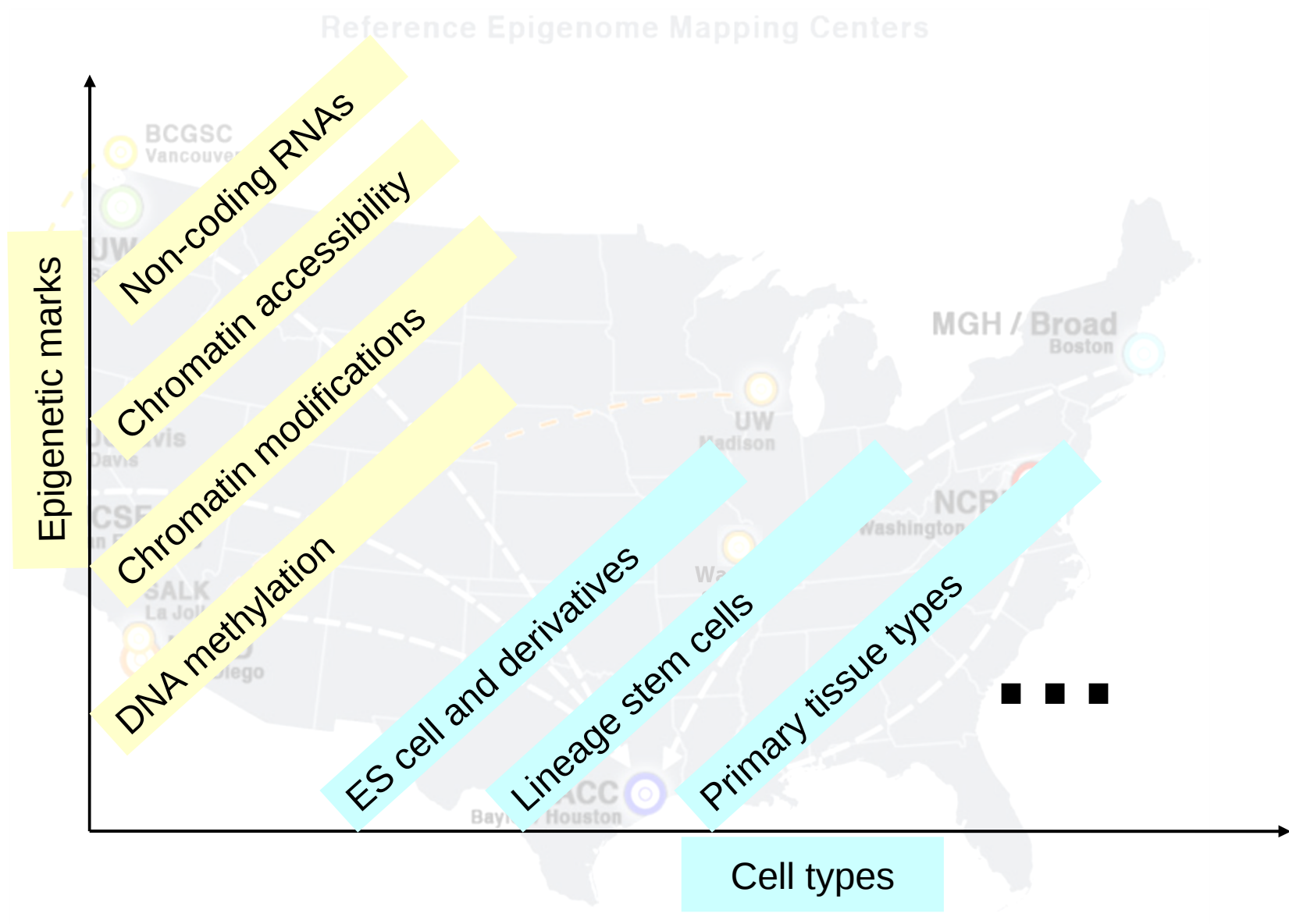
<http://www.roadmapepigenomics.org/>

### Reference Epigenome Mapping Centers





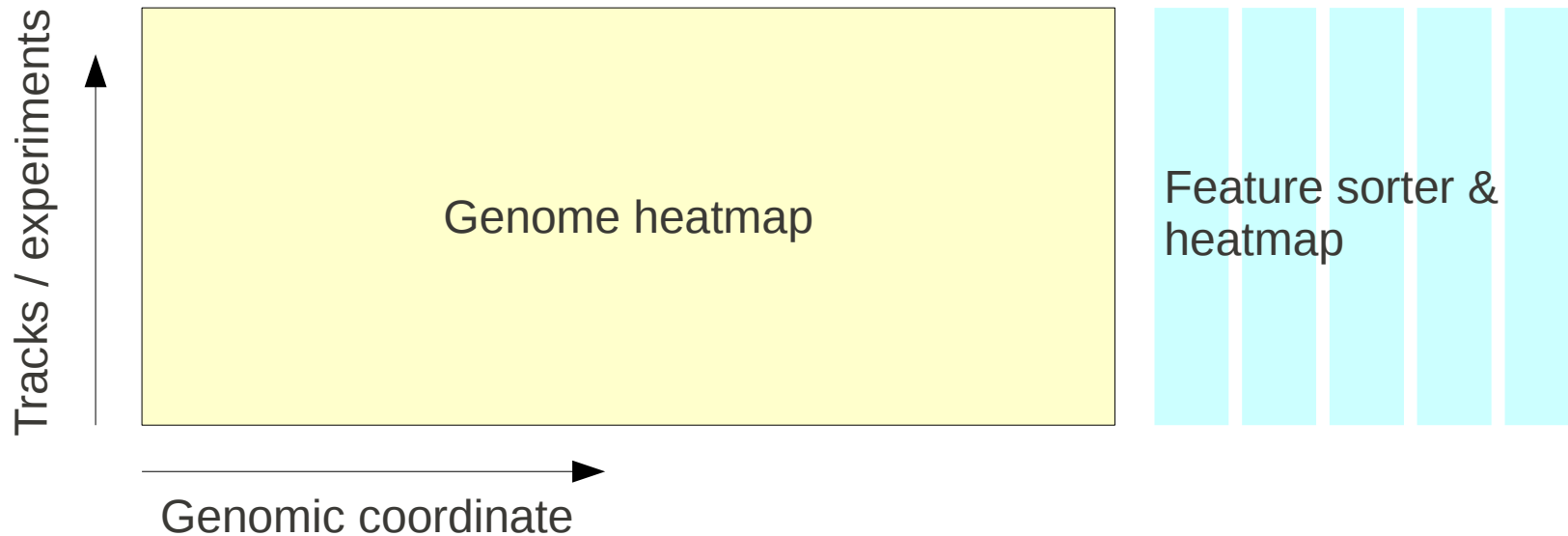
<http://www.roadmapepigenomics.org/>





# Human epigenome browser

- Adopt the UCSC Cancer Browser display style

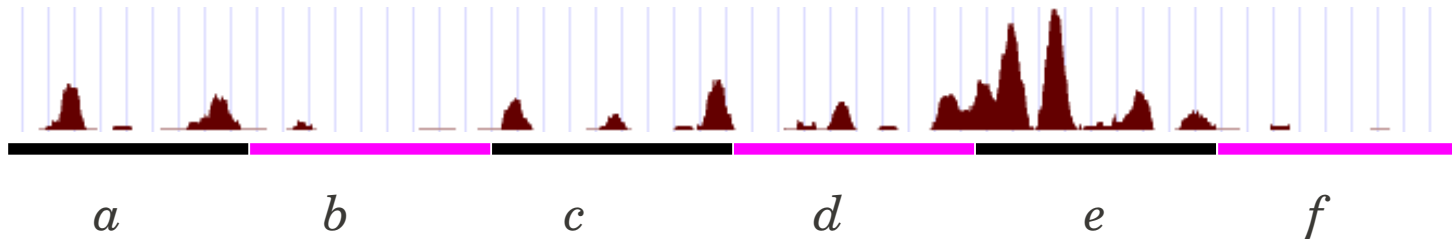


# Data transformation

Step 1: divide genome into bins

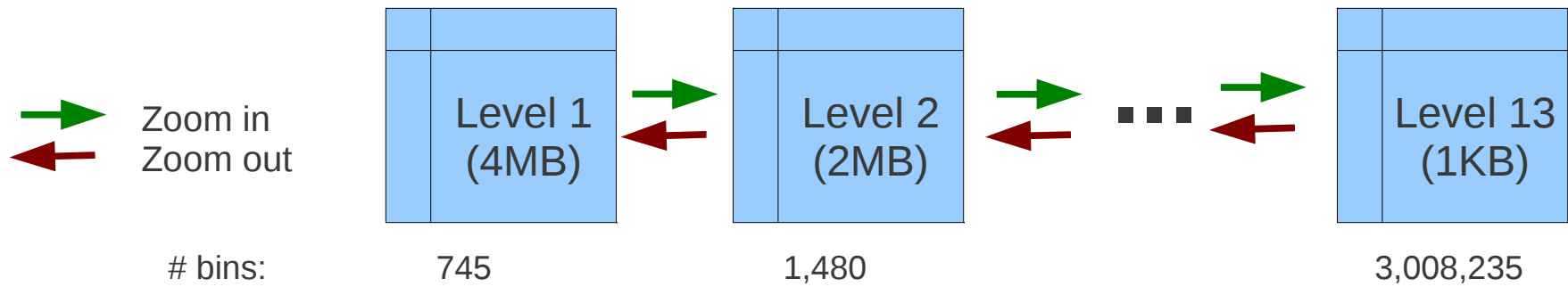


Step 2: compute score for each bin using sequencing data



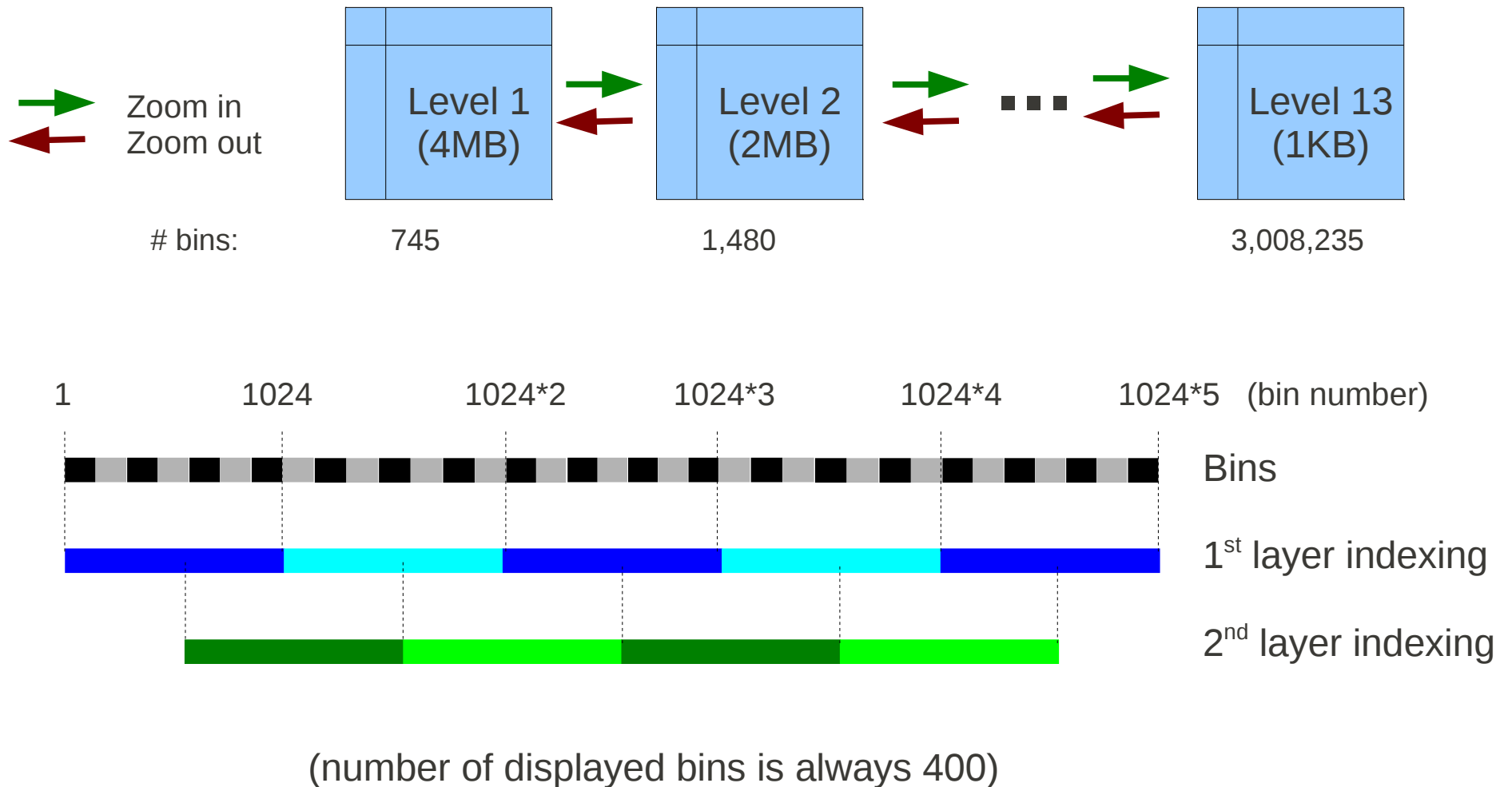
Step 3: apply steps 1-2 to multiple level of bins

# Database design





# Database design



# Software implementation

- Server-side
  - MySQL
  - C, Kent source tree
    - CGI, MySQL communication, image rendering
  - R
- Front-end
  - JavaScript, Ajax

# Human Epigenome Browser

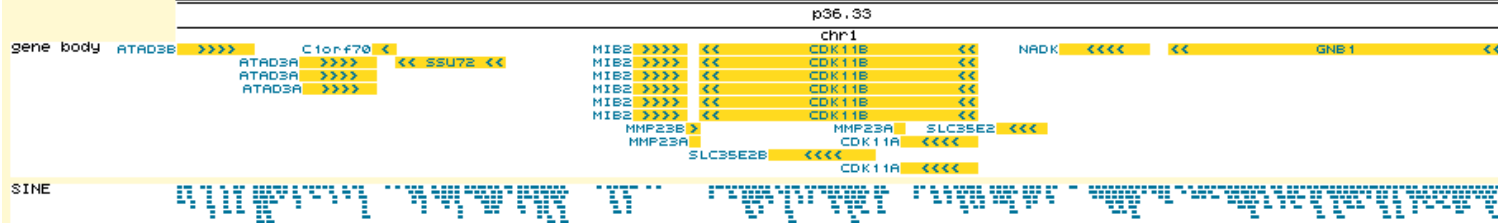
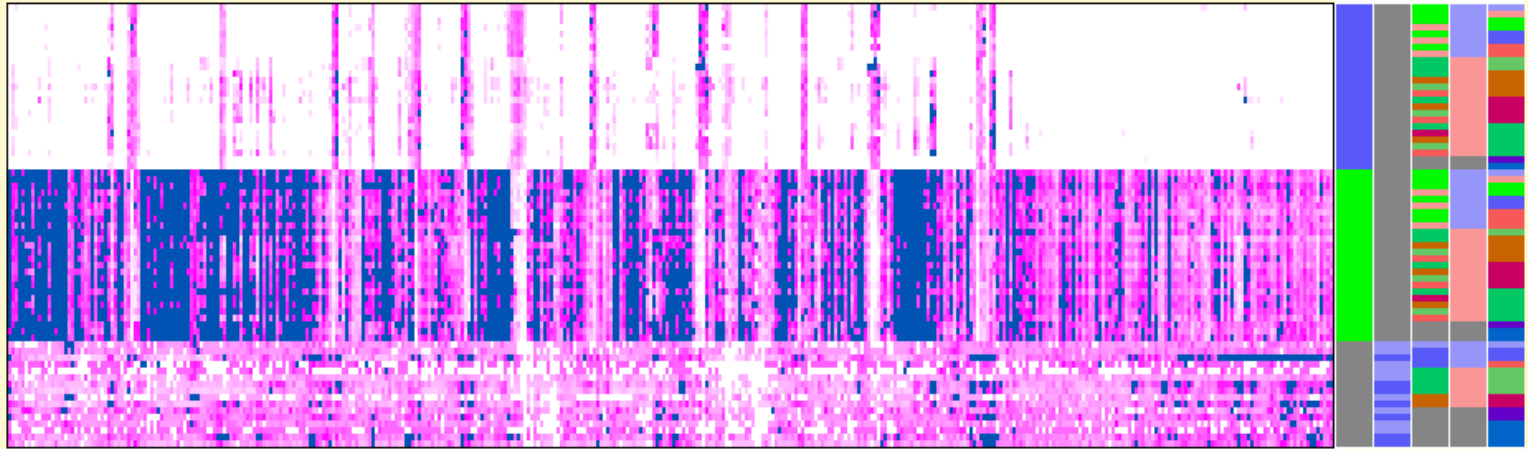
\*\*\* Displaying genome \*\*\*

- start: chr1 1398784
- stop: chr1 1809408
- spanning 401 kb

Normalize data by

[View current region in UCSC genome browser](#)

Track height:  pixels



Position  gene

e.g. chr1:1397760-1808384  
separator could be space, tab, dash, colon  
or single chromosome name

e.g. CDK11B, IL1RAPL1  
case insensitive

- Choose genomic features (horizontal axis)
- Draw decorative tracks
- Select data tracks (vertical axis)
- Track score & genomic feature density correlation analysis
- Hypothesis test
- Get data in tabular format

# Human Epigenome Browser

\*\*\* Displaying genome \*\*\*

- start: chr1 1398784
- stop: chr1 1809408
- spanning 401 kb

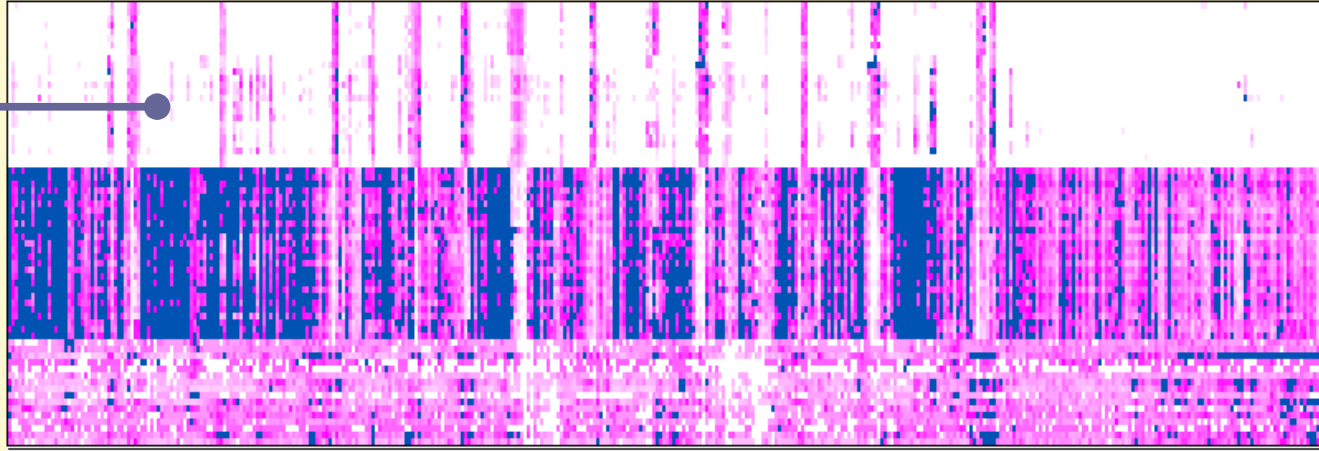
Normalize data by

[View current region in UCSC genome browser](#)

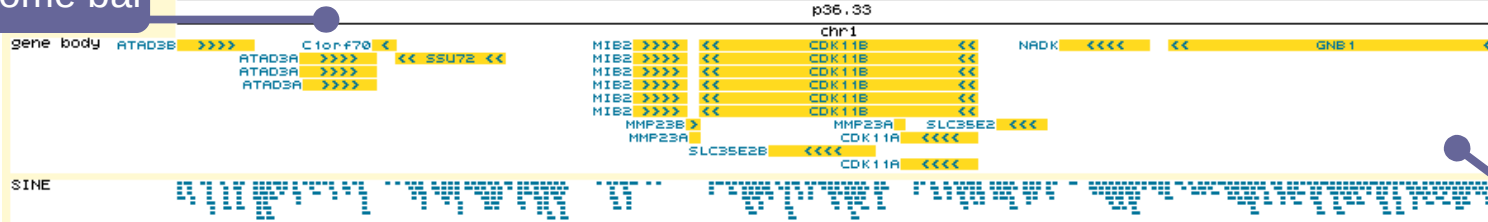
Track height:  pixels

DNA methylation  
Histone modification  
Individual  
Tissue / organ  
Cell type

Genome heatmap



Chromosome bar



Track attribute & sorting

Jump to region

Position  gene

e.g. chr1:1397760-1808384  
separator could be space, tab, dash, colon  
or single chromosome name

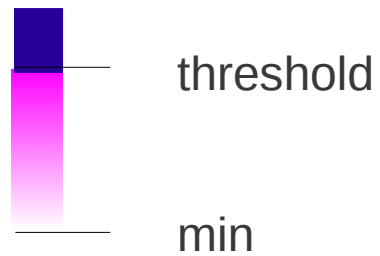
e.g. CDK11B, IL1RAPL1  
case insensitive

"Decorative" track

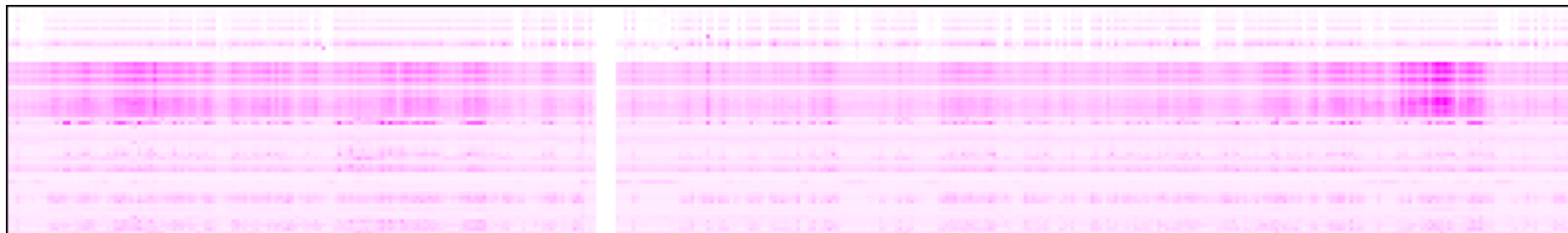
- Choose genomic features (horizontal axis)
- Draw decorative tracks
- Select data tracks (vertical axis)
- Track score & genomic feature density correlation analysis
- Hypothesis test
- Get data in tabular format

Control panel

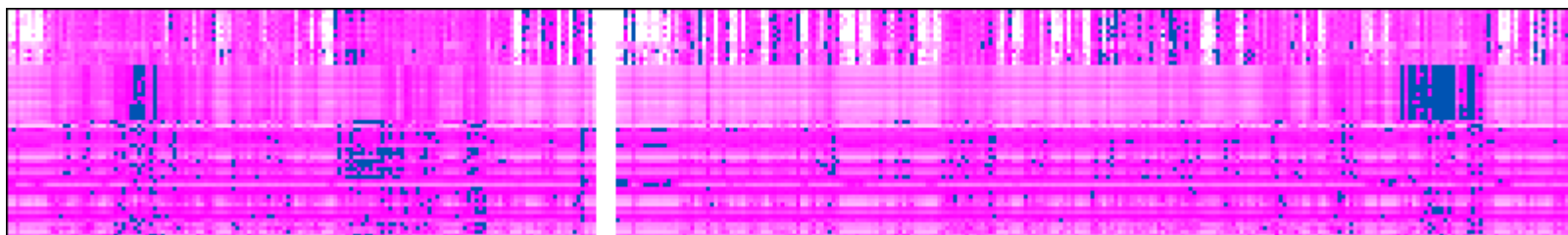
# Heatmap coloring



*Different thresholds*

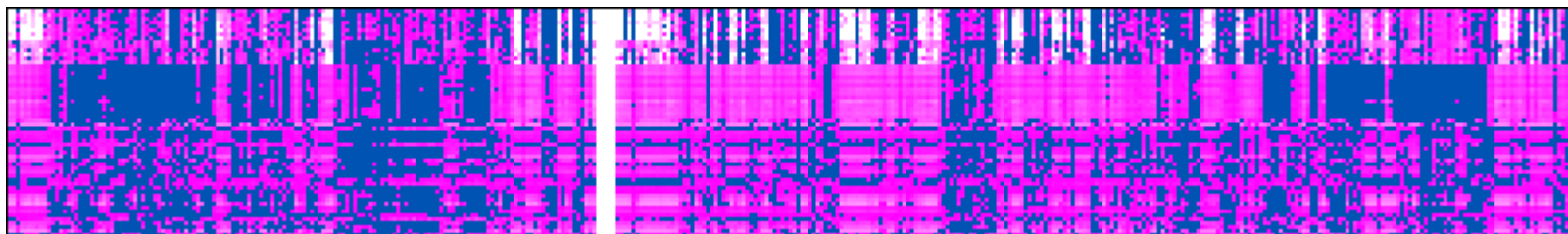


Track max



95 percentile of track

*(90, 80, 70 percentiles)*



60 percentile of track

# Track selection

Cell/tissue types

▢
**Select data tracks (vertical axis)**

	DNA methylation: RRBS	DNA methylation: MethylC-Seq	DNA methylation: MeDIP-Seq	DNA methylation: MRE-Seq	DNA methylation: MBD-Seq	Histone: H3K9me3	Histone: H3K9Ac	Histone: H3K4me3	Histone: H3K4me1	Histone: H3K36me3	Histone: H3K27me3	Histone: input
Peripheral blood cell			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>				<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
CD8 Memory cell			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>								
CD4 Memory cell			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>								
CD8 Naive cell			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>
CD4 Naive cell			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<input type="checkbox"/>			<input type="checkbox"/>	<input type="checkbox"/>	
Breast vHMEC cell			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>
Breast stem cell			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>								
Breast myoepithelial cell			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>		<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Breast luminal epithelial cell			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>				<input type="checkbox"/>			<input type="checkbox"/>	
IMR90		<input type="checkbox"/>										
H1EsB2			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
H1EsB1	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

# Hypothesis test

## on bin scores across groups of tracks

### Hypothesis test

- Howto: select a track attribute type to do test.
  - E.g. if "DNA methylation" is selected, track scores will be grouped by different type of DNA methylation experiments and tested.
  - Tracks not belonging to "DNA methylation" will not participate in test.
- Test method: Kruskal-Wallis rank sum test (R function *kruskal.test*).
- Multiple testing correction: false discovery rate ("Benjamini & Yekutieli 2001" method in R function *p.adjust*)
- Result display: vertical bar plot of p-values (log10 scaled) is placed beneath the heatmap image and aligned with it.

<input checked="" type="radio"/> Don't perform test	<input type="radio"/> DNA Methylation <input type="radio"/> Histone Modification <input type="radio"/> Individual <input type="radio"/> Tissue Organ <input type="radio"/> Cell Type	Multiple testing correction not applied ▼
---	--	--

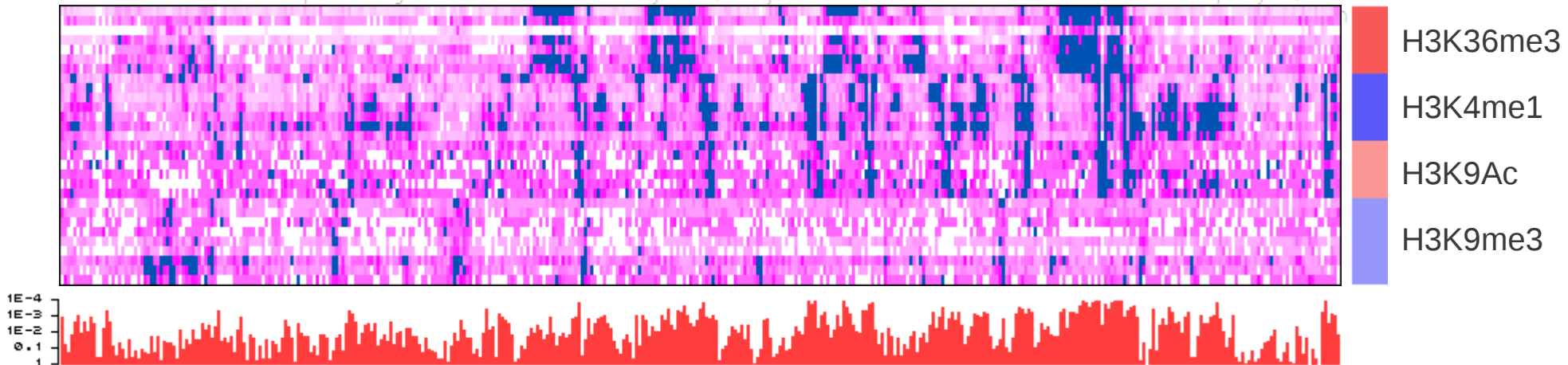
# Hypothesis test

## on bin scores across groups of tracks



### Hypothesis test

- Howto: select a track attribute type to do test.
  - E.g. if "DNA methylation" is selected, track scores will be grouped by different type of DNA methylation experiments and tested.
  - Tracks not belonging to "DNA methylation" will not participate in test.
- Test method: Kruskal-Wallis rank sum test (R function *kruskal.test*).
- Multiple testing correction: false discovery rate ("Benjamini & Yekutieli 2001" method in R function *p.adjust*)





# Correlation analysis

## Track scores & genomic feature density



### Track score & genomic feature density correlation analysis

Don't perform analysis

#### RefSeq gene

- 3kb promoter
- 5' UTR
- gene body
- 3' UTRi

#### non-coding RNA

- Non-coding RNA gene
- sno/miRNA

#### transposable elements

- DNA transposon
- SINE
- LINE
- LTR
- other repeat

#### others

- CpG island

# Correlation analysis

## Track scores & genomic feature density

Histone modification

H3K36me3

H3K4me1

H3K9Ac

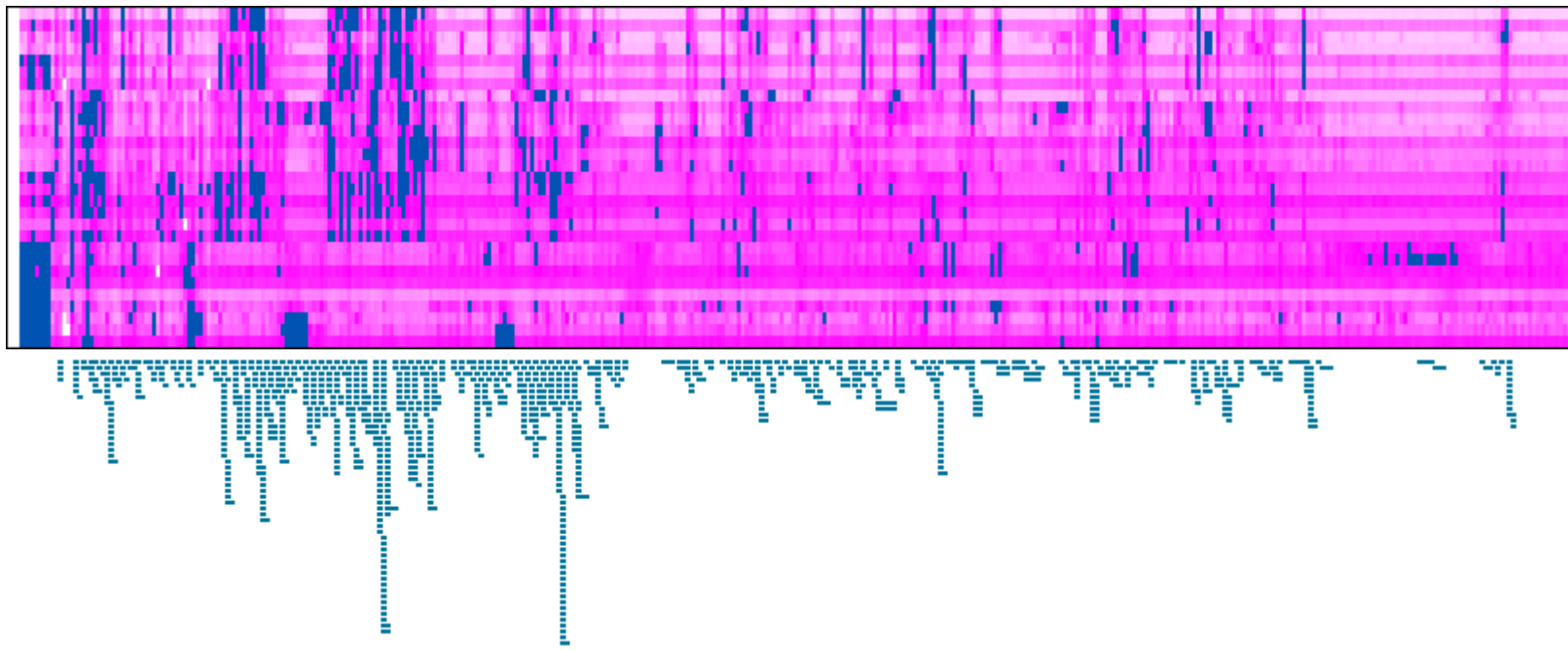
H3K9me3

Track score & genomic feature density correlation analysis

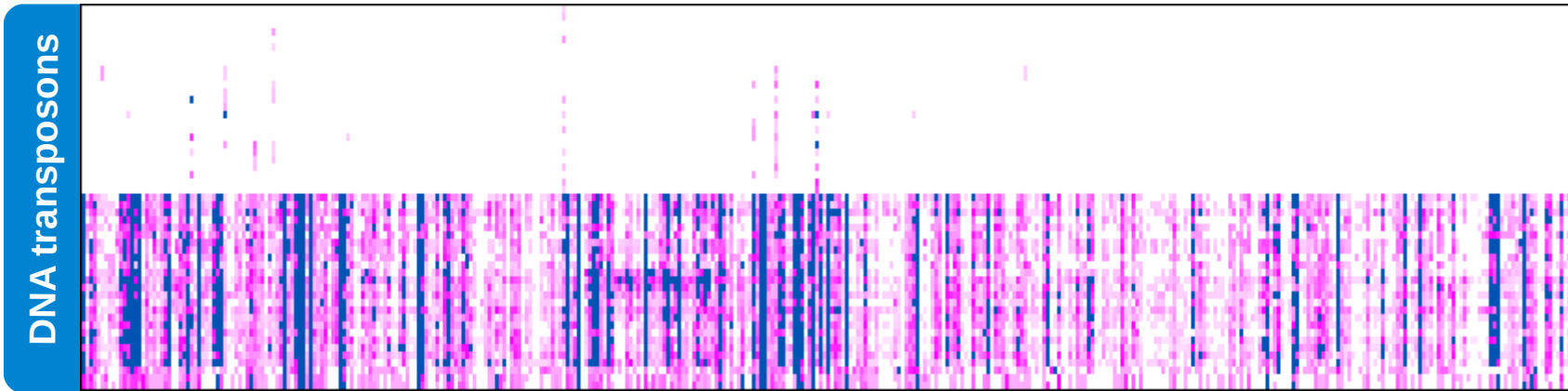
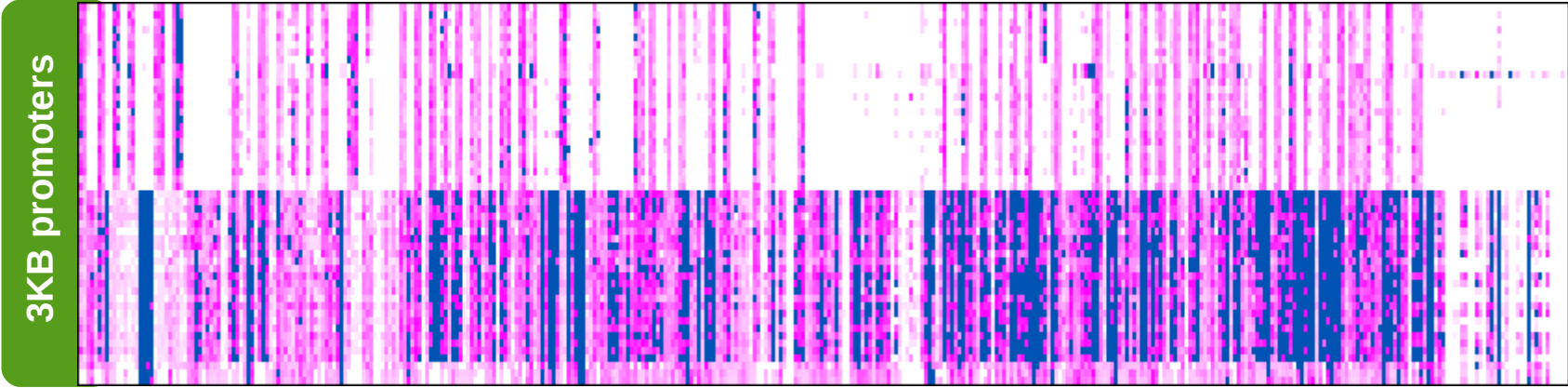
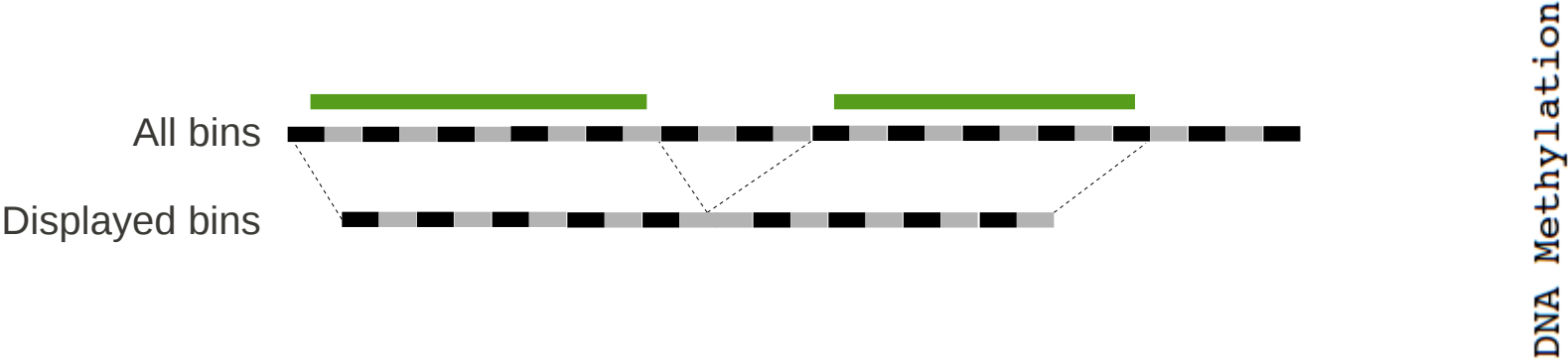
Track correlation with gene body

-1 0 1

gene body



# Focusing on genomic features



DNA Methylation

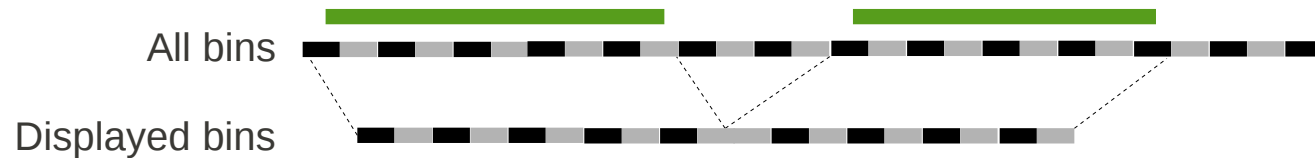
MRE-Seq

MeDIP-Seq

MRE-Seq

MeDIP-Seq

# Focusing on genomic features



Choose genomic features (horizontal axis)

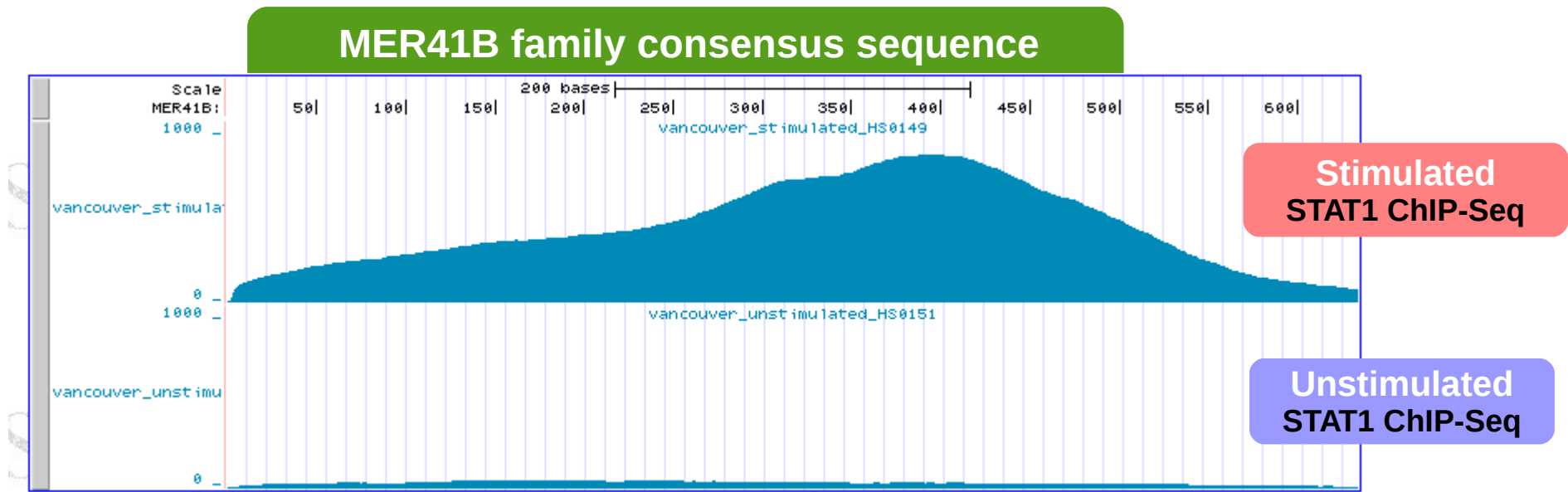
genome	RefSeq gene	non-coding RNA	transposable elements	others
<input type="radio"/> genome	<input checked="" type="radio"/> 3kb promoter <input type="radio"/> 5' UTR <input type="radio"/> gene body <input type="radio"/> exon <input type="radio"/> intron <input type="radio"/> 3' UTR	<input type="radio"/> non-coding RNA gene <input type="radio"/> sno/miRNA	<input type="radio"/> DNA transposon <input type="radio"/> SINE <input type="radio"/> LINE <input type="radio"/> LTR <input type="radio"/> other repeat	<input type="radio"/> CpG island

# Problem to solve

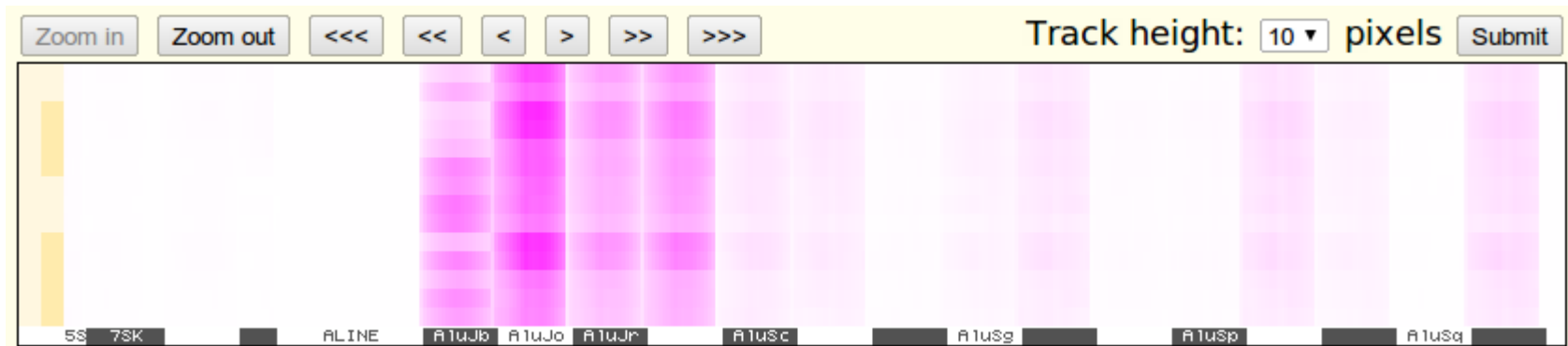
- Data transformation
- Database design
- Comments and suggestions are appreciated

# Future work

- A focus on transposable elements
  - Carry TF binding sites around
  - Shape the landscape of mammalian regulatory network



# Future work: a browser for transposable elements



- Adapt the code of epigenome browser
- Family centric view
- Reprocess REMC epigenome dataset, display on this browser

# Acknowledgement

- Ting Wang's lab



- Brett Maricque

- 

- Vasavi Sundaram

- 



- Mingchao Xie

- 

- Xiaoyun Xing

