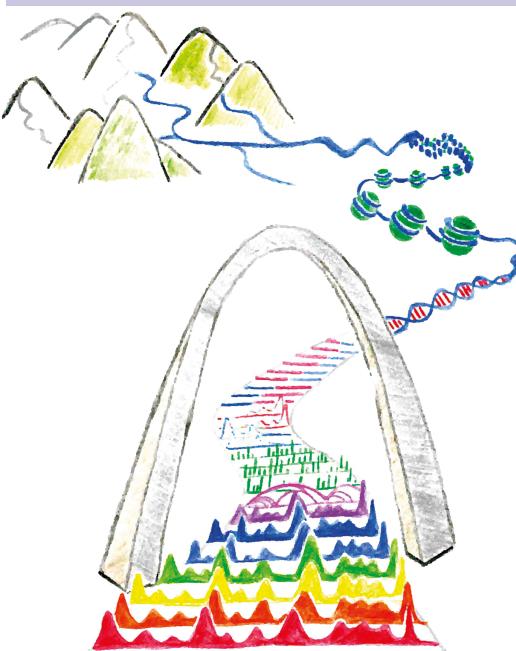


系统生物学与生物信息学
海外学者短期讲学系列课程

Current Topics in Epigenomics

表观基因组学前沿



Ting Wang
Department of Genetics
Center for Genome Sciences and Systems Biology
Washington University School of Medicine

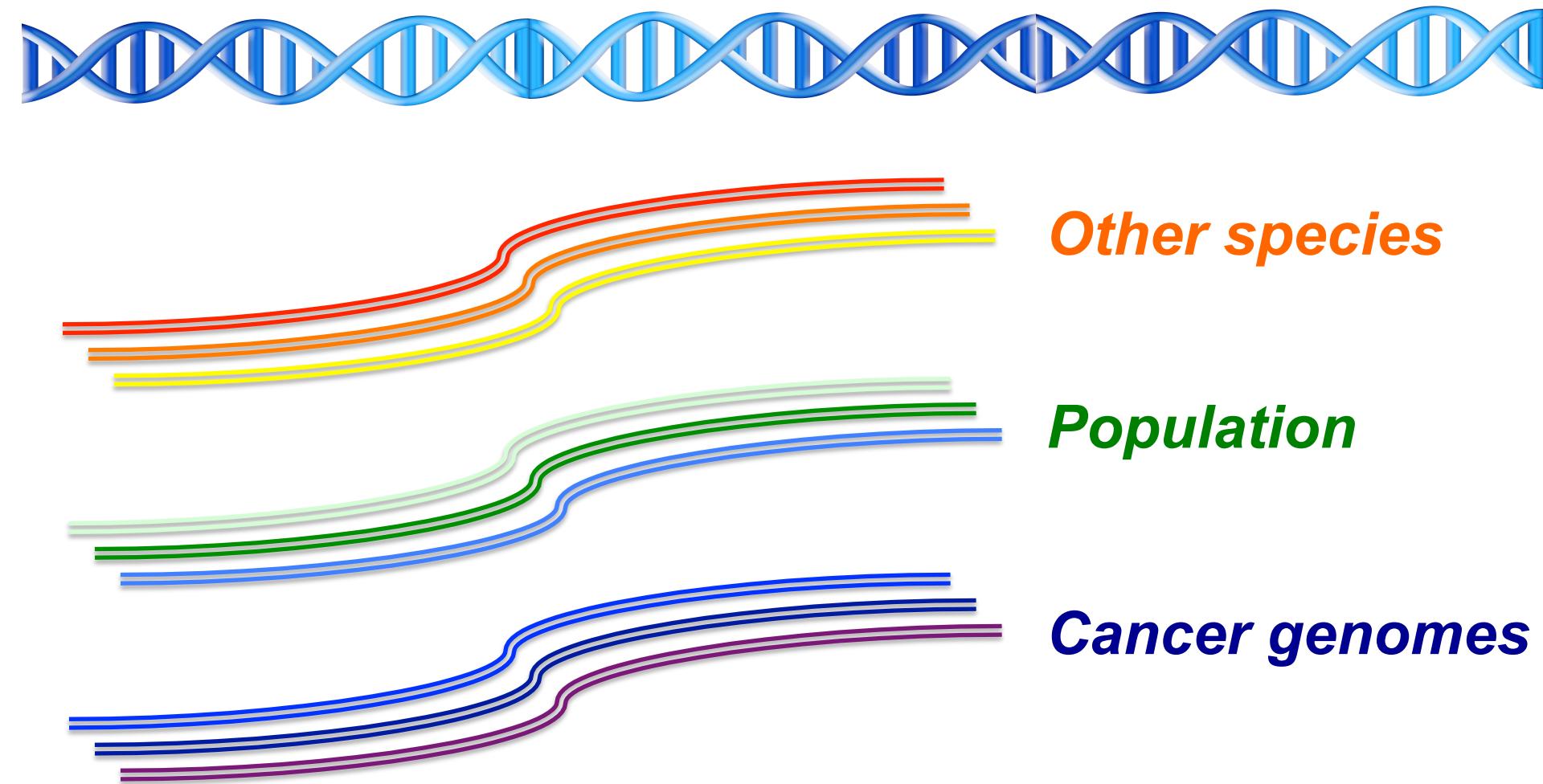
Tsinghua University
April 15-27

The era of big bio-data

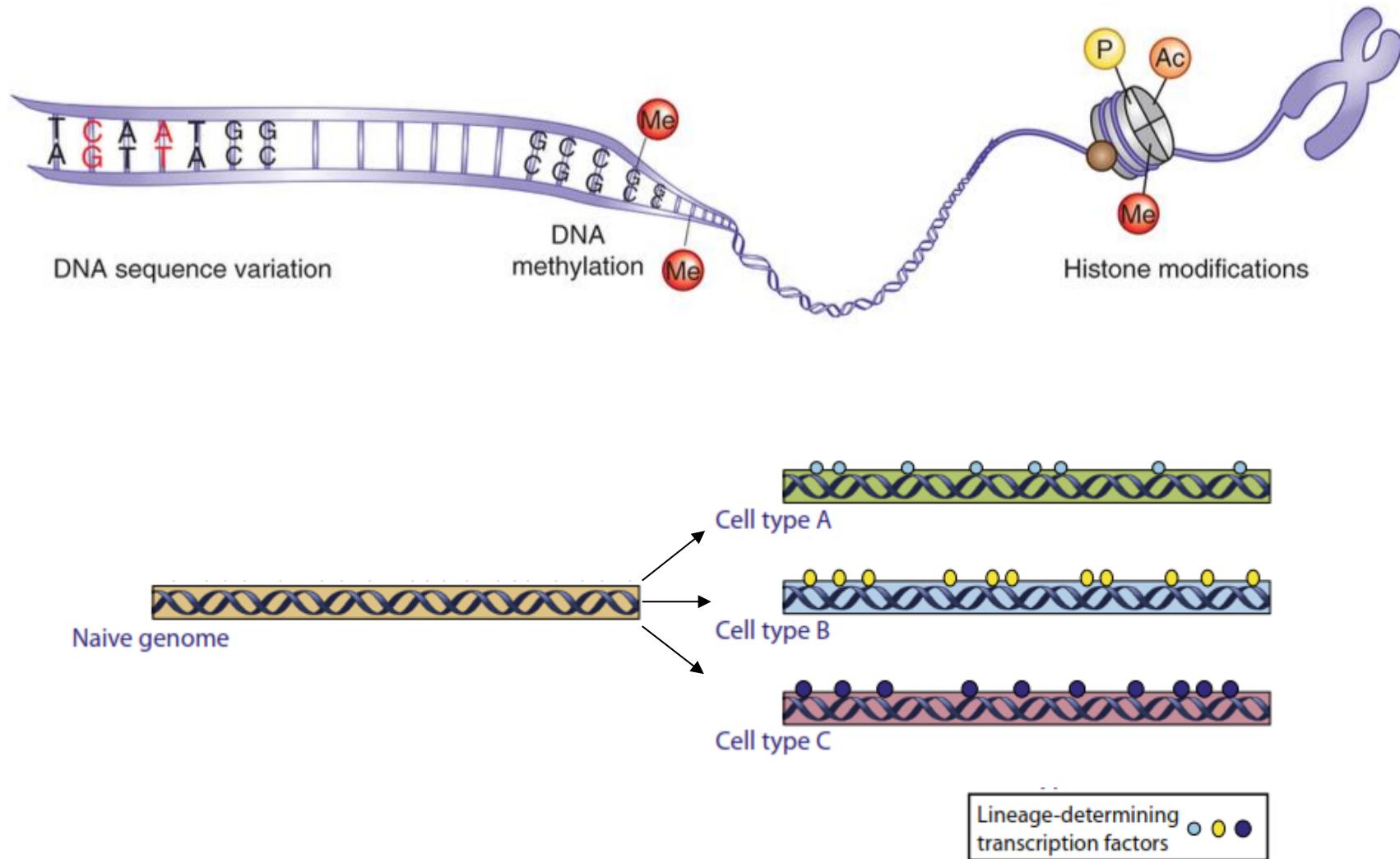


The Reference Human Genome

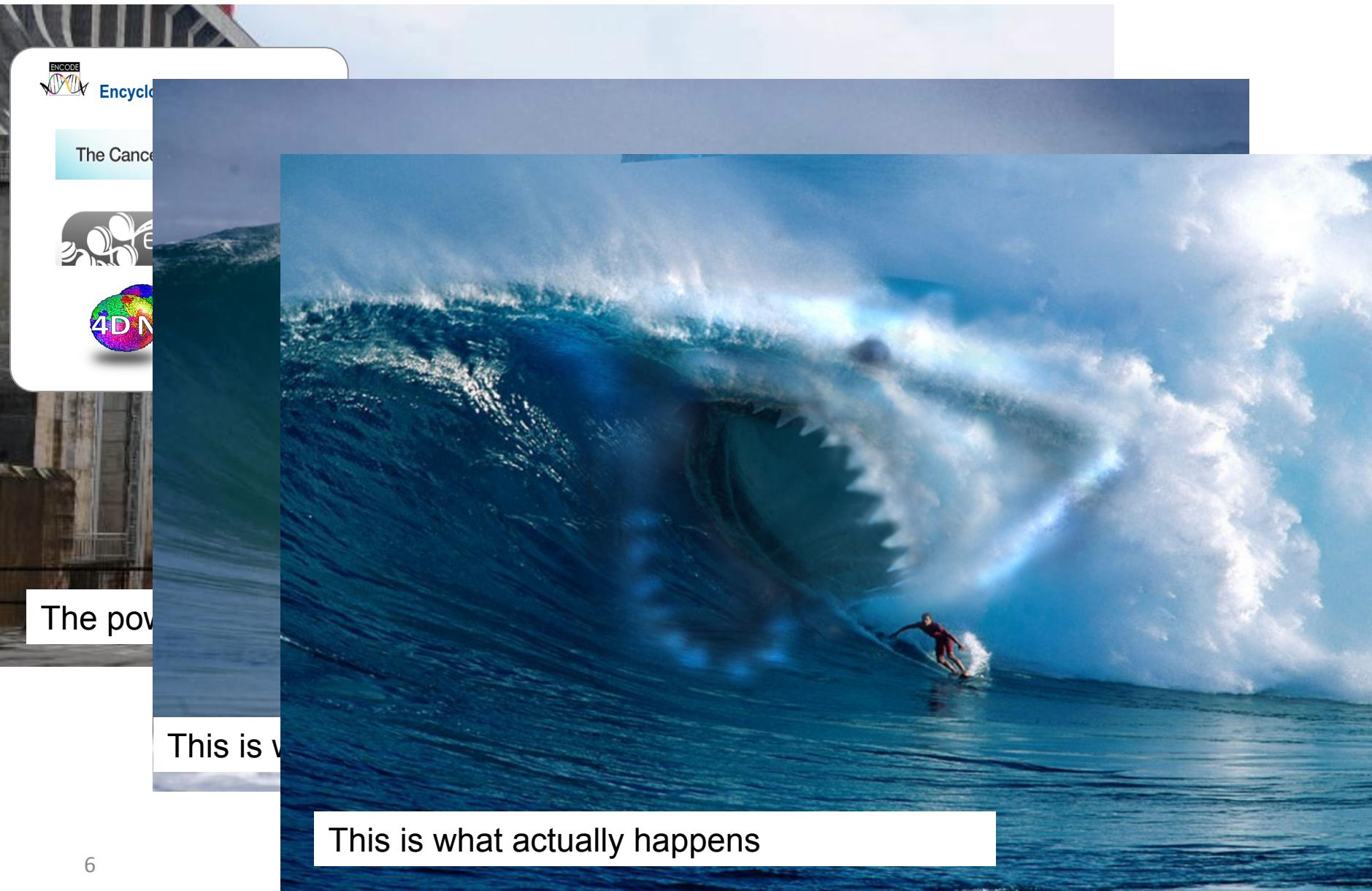
Comparing the reference human genome



Comparing the human reference epigenome



Data in the post-genome era



ENCODE
Encyclo

The Cancer



The power

This is w

This is what actually happens

Big Bio Data Information Flood

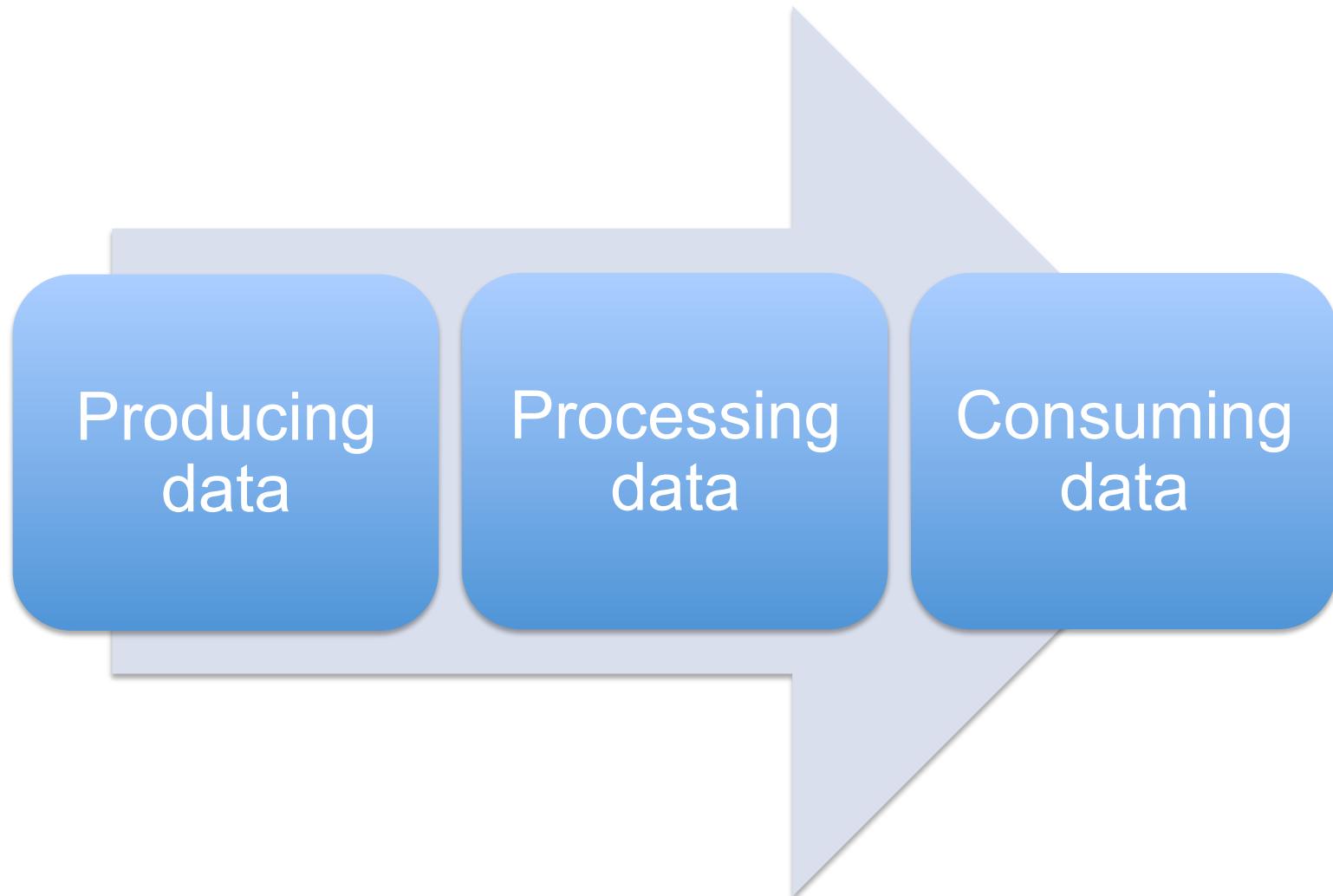


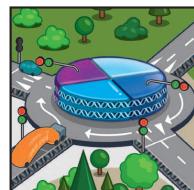
Mysterious!

Complex!

A large grid of binary code (0s and 1s) arranged in a square pattern, representing the vast amount of data being discussed.

Big Bio Data Information ~~Flood~~ Flow

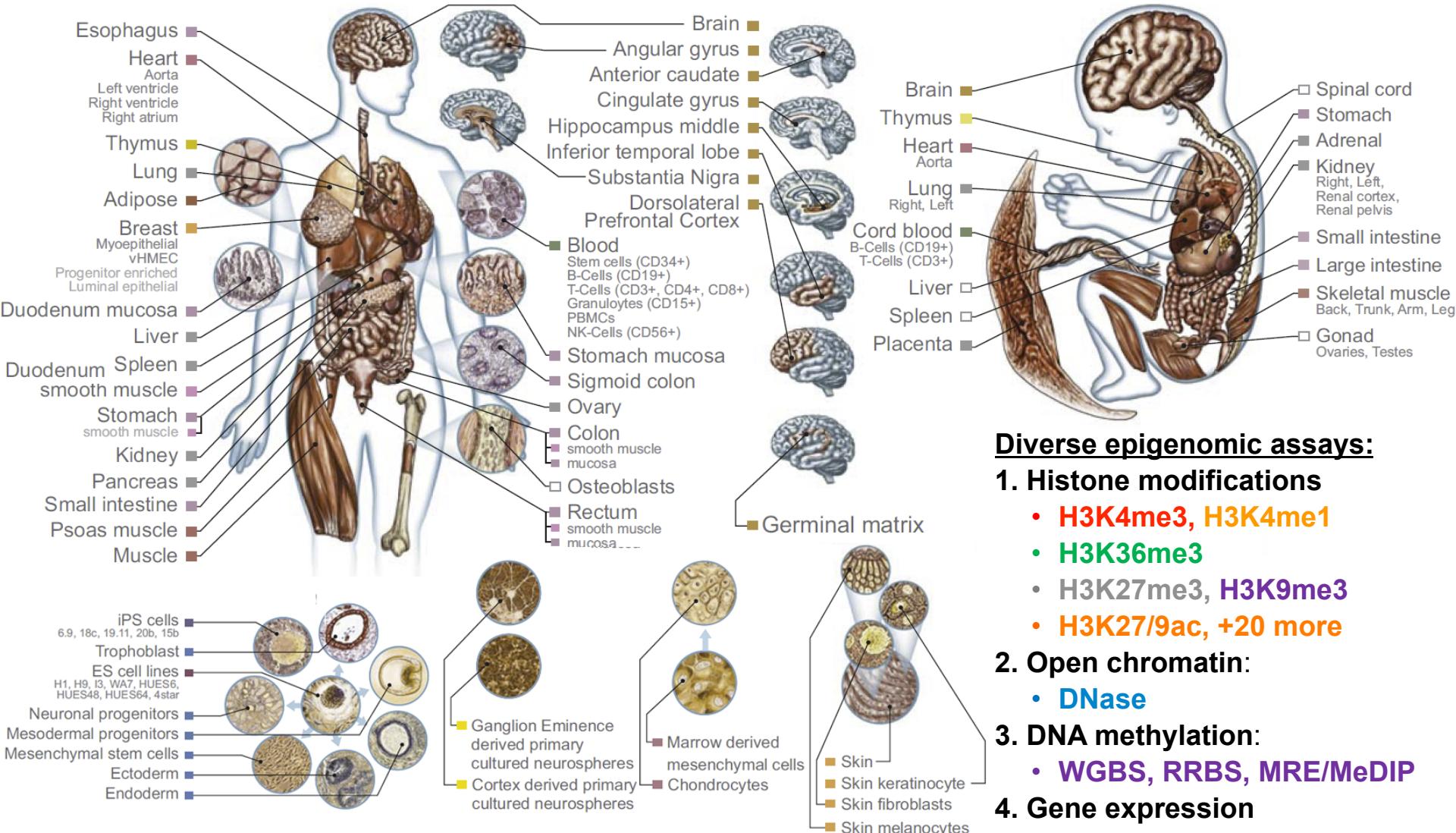




EPIGENOME ROADMAP
A *Nature* special issue
nature.com/epigenomeroadmap

NIEHS, NIDA, NHGRI, Common Fund

Epigenomics Roadmap across 100+ tissues/cell types



Diverse epigenomic assays:

- 1. Histone modifications:**
 - H3K4me3, H3K4me1
 - H3K36me3
 - H3K27me3, H3K9me3
 - H3K27/9ac, +20 more
- 2. Open chromatin:**
 - DNase
- 3. DNA methylation:**
 - WGBS, RRBS, MRE/MeDIP
- 4. Gene expression:**
 - RNA-seq, Exon Arrays

Diverse tissues and cells:

- 1. Adult tissues and cells** (brain, muscle, heart, digestive, skin, adipose, lung, blood...)
- 2. Fetal tissues** (brain, skeletal muscle, heart, digestive, lung, cord blood...)
- 3. ES cells, iPS, differentiated cells** (meso/endo/ectoderm, neural, mesench, trophobl)

Roadmap Integrative Analysis Group



Nature, Feb 18, 2015
Vol. 518, pp. 313-370



Epigenome Roadmap

Home | Research | Threads | News and Multimedia | Additional research | Sponsor



Welcome to the Epigenome Roadmap! Here, we have collected research papers describing the main findings of the NIH Roadmap Epigenomics Program, the aim of which was to systematically characterize epigenomic landscapes in primary human tissues and cells. The papers are complemented by eight threads each of which highlights a topic that runs through more than one paper. Threads are designed to help you explore the wealth of information collectively published across several Nature Publishing Group journals. Each thread consists of relevant paragraphs, figures and tables from across the papers, united around a specific theme.

We invite you to explore the research content, the News & Views, the video and other associated material.

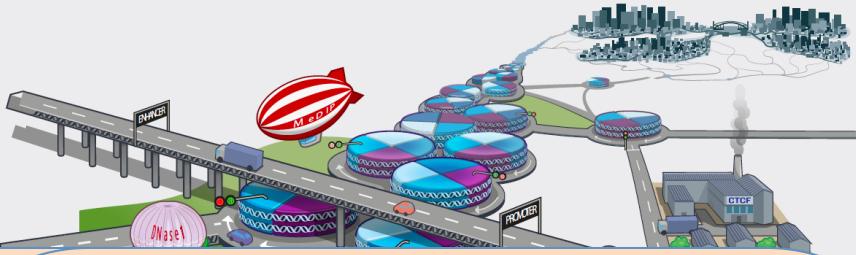
News and Multimedia

- Nature News | Editorial Beyond the genome
- Nature | News and Views Epigenomics: Roadmap for regulation
Casey E. Romanoski, Christopher K. Glass, Hendrik G. Stuennenberg, Laurence Wilson, Genevieve Almouzni
- Nature News | News Epigenome: The symphony of your cells
Kerri Smith

Thread articles

THREAD 1	THREAD 2	THREAD 3	THREAD 4	THREAD 5	THREAD 6
1. Annotation of the non-coding genome Highlight referenced papers ▶	2. Relationship between different epigenomic marks: DNA accessibility and methylation, histone marks, and RNA Highlight referenced papers ▶	3. Epigenetic changes during differentiation and development Highlight referenced papers ▶	4. Regulatory models: networks, motifs, modules, sequence drivers and predictive models Highlight referenced papers ▶	5. Interpreting variation: GWAS, cancer, genotype, evolution and allelic Highlight referenced papers ▶	6. Epigenomic changes in human disease and during cancer progression Highlight referenced papers ▶

Produced with exclusive support from: **illumina®**



8 Research threads

- **Annotation of the non-coding genome**
- **Relationship between different epigenomic marks: DNA accessibility and methylation, histone marks, and RNA**

8 Research threads

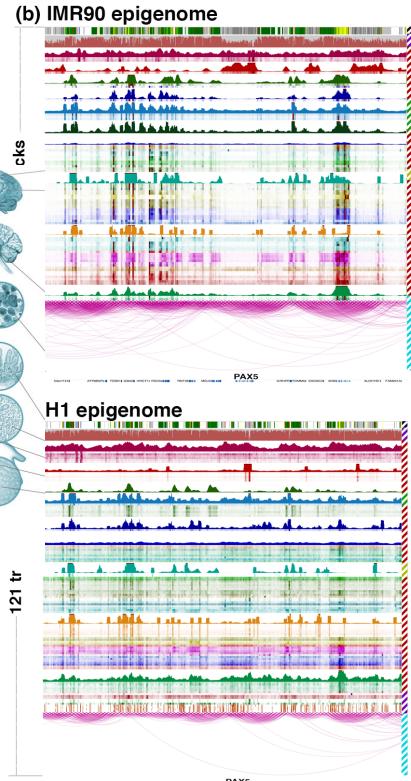
21 papers in one bundle

~60 previously published roadmap papers

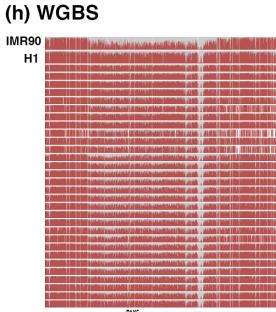
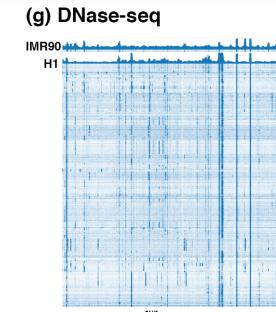
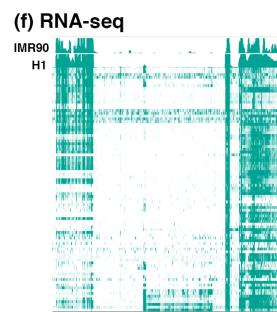
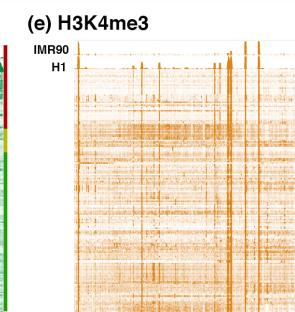
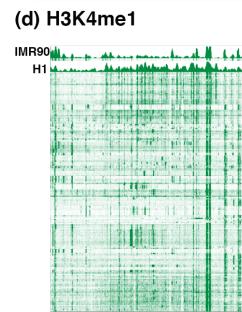
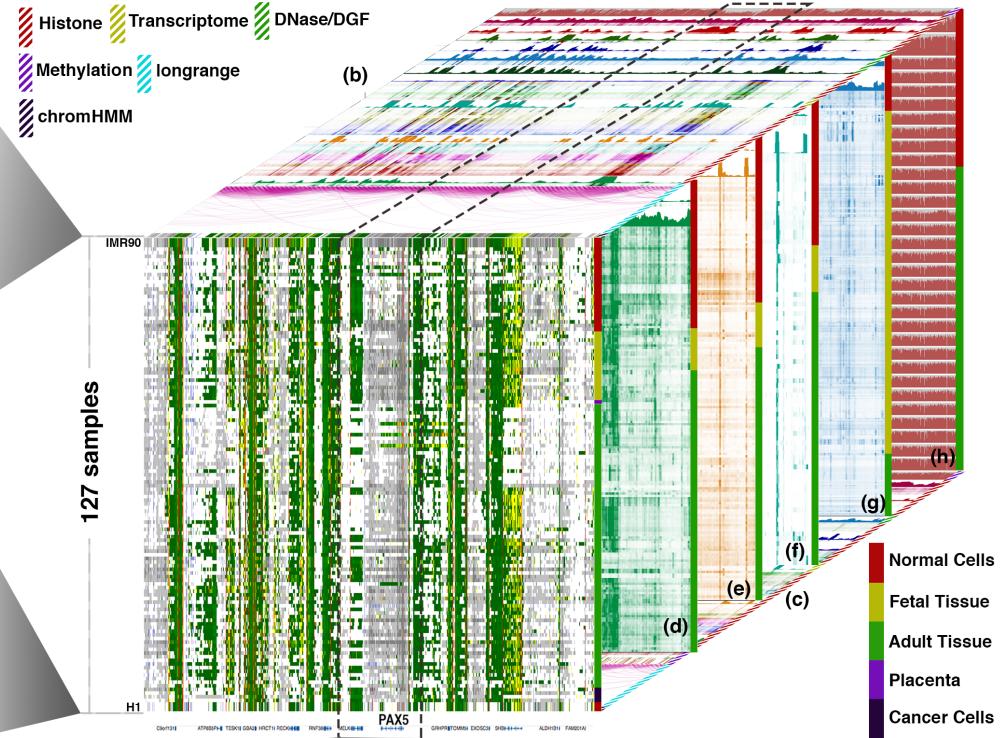
genotype, evolution and allele

- **Epigenomic changes in human disease and during cancer progression**
- **Brain epigenome**
- **Computational tools and methods**

The Map!



(a) complete epigenomes overview





Genomes

Genome Browser

Tools

Mirrors

Downloads

My Data

About Us

View

Help

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

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

chr17:7,569,035-7,593,494 24,460 bp. TPS3

go

chr17 (p13.1) p13.3 p13.2 p13.1 17p12 17p11.2 17q11.2 17q12 q21.31 17q22 23.2 24.2 q24.3 q25.1 17q25.3

The UCSC Browser

The linear genome

Scale chr17: 7,575,000 7,580,000 7,585,000 7,590,000 hg19

CpG Islands

The gene

Vertebrate Cons

Conservation

Mammal Cons

Chimp

Rhesus

Mouse

Dog

Opossum

Chicken

X_tropicalis

Zebrafish

Function

Layered H3K27Ac

2 17 42 3 11 9 3 2 11 4 7 8 4
17 3 18 125 125 125 4 125 4 125

A549 DEX100nM 1

A549 DEX100pM 1

A549 DEX1nM 1

A549 DEX500pM 1

A549 DEX5nM 1

A549 ETOH 1

Variation

RepeatMasker

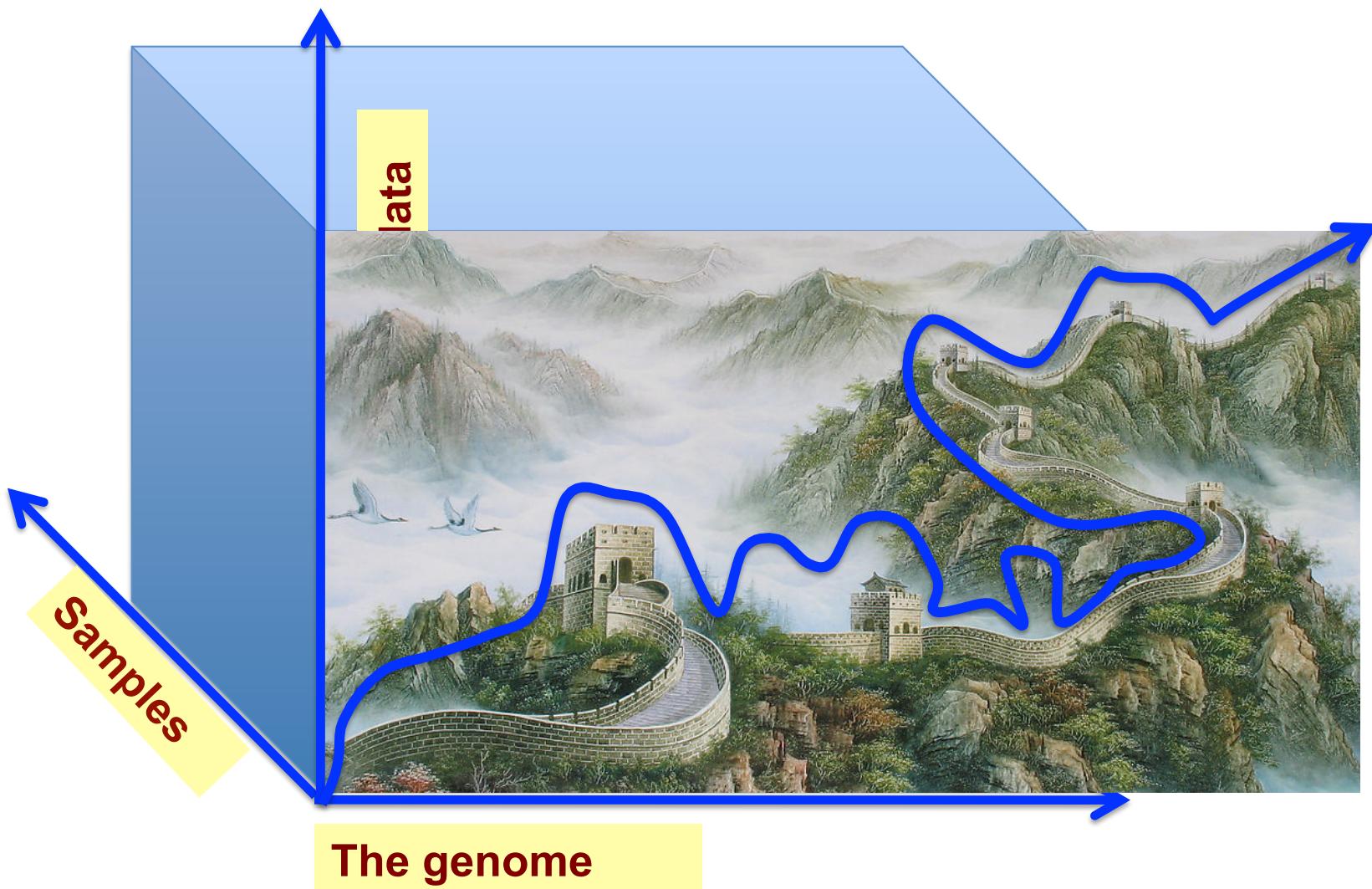
Common SNPs(137)

rs1625895

Sequences

SNPs

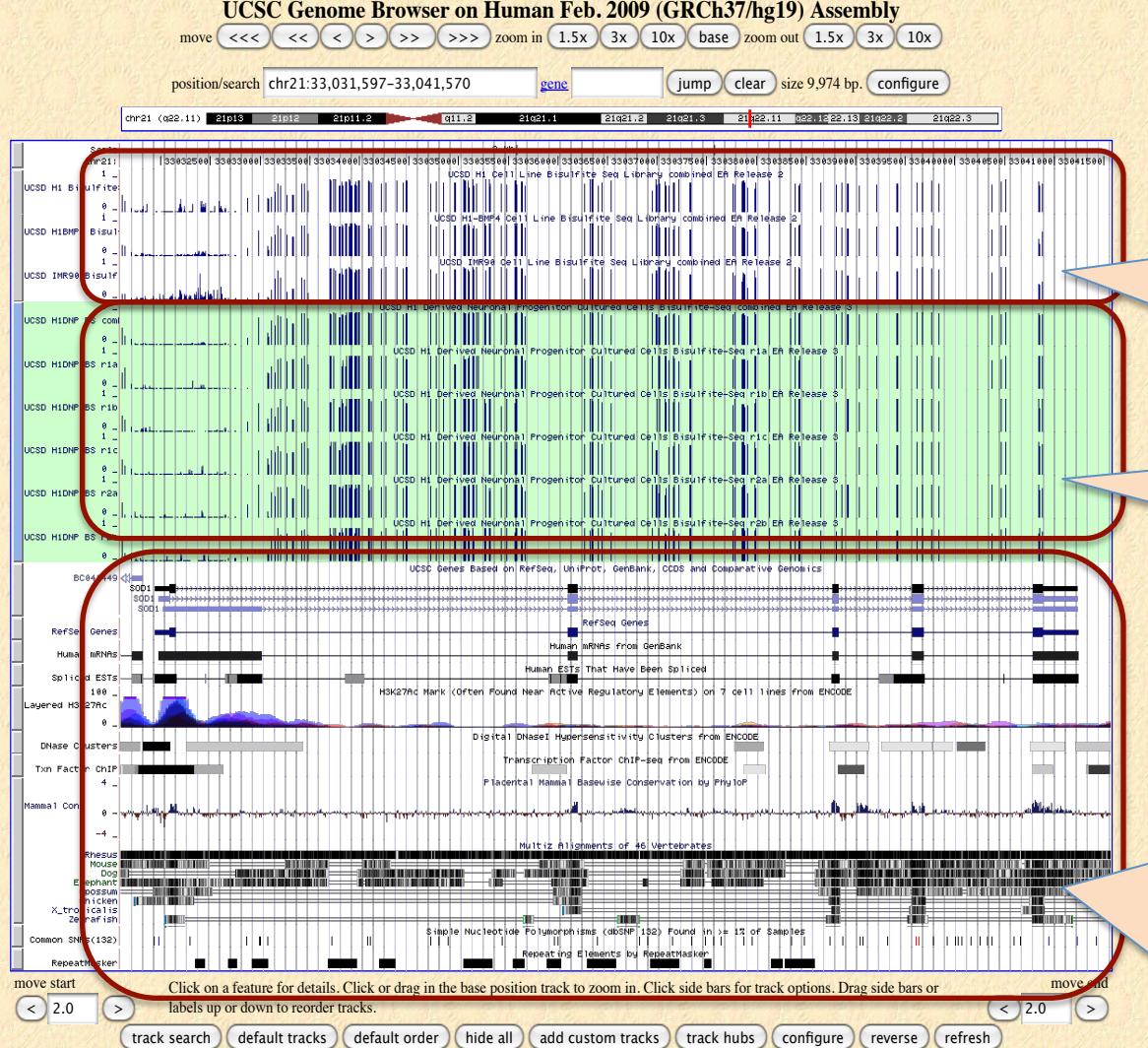
Navigating the genomic data landscape



Mechanisms for accessing Roadmap data

- **Current datasets available through Wash U Hub**
 - Roadmap experimental data: **4737** tracks
 - Roadmap integrative analysis: **18181** tracks
 - ENCODE data: **4253** tracks
 - *IHEC data* (<http://epigenomesportal.ca/ihec/>): **5667** tracks
- **UCSC Genome Browser Mirror**
 - <http://epigenome.wustl.edu>
 - Fully functional UCSC Mirror for accessing complete set of Roadmap data
- **Remote Data Hub for UCSC Browser**
 - <http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg19&hubUrl=http://vizhub.wustl.edu/VizHub/RoadmapReleaseAll.txt>
 - Data are physically located at remote site (St. Louis), and directly displayed at UCSC Genome Browser
- **Wash U Epigenome Browser**
 - <http://epigenomegateway.wustl.edu>
 - Next generation genome browser for exploring Roadmap, ENCODE, and other data
 - Community outreach and workshops
 - <http://epigenomegateway.wustl.edu/support/workshop.html>

UCSC remote access



Roadmap default tracks

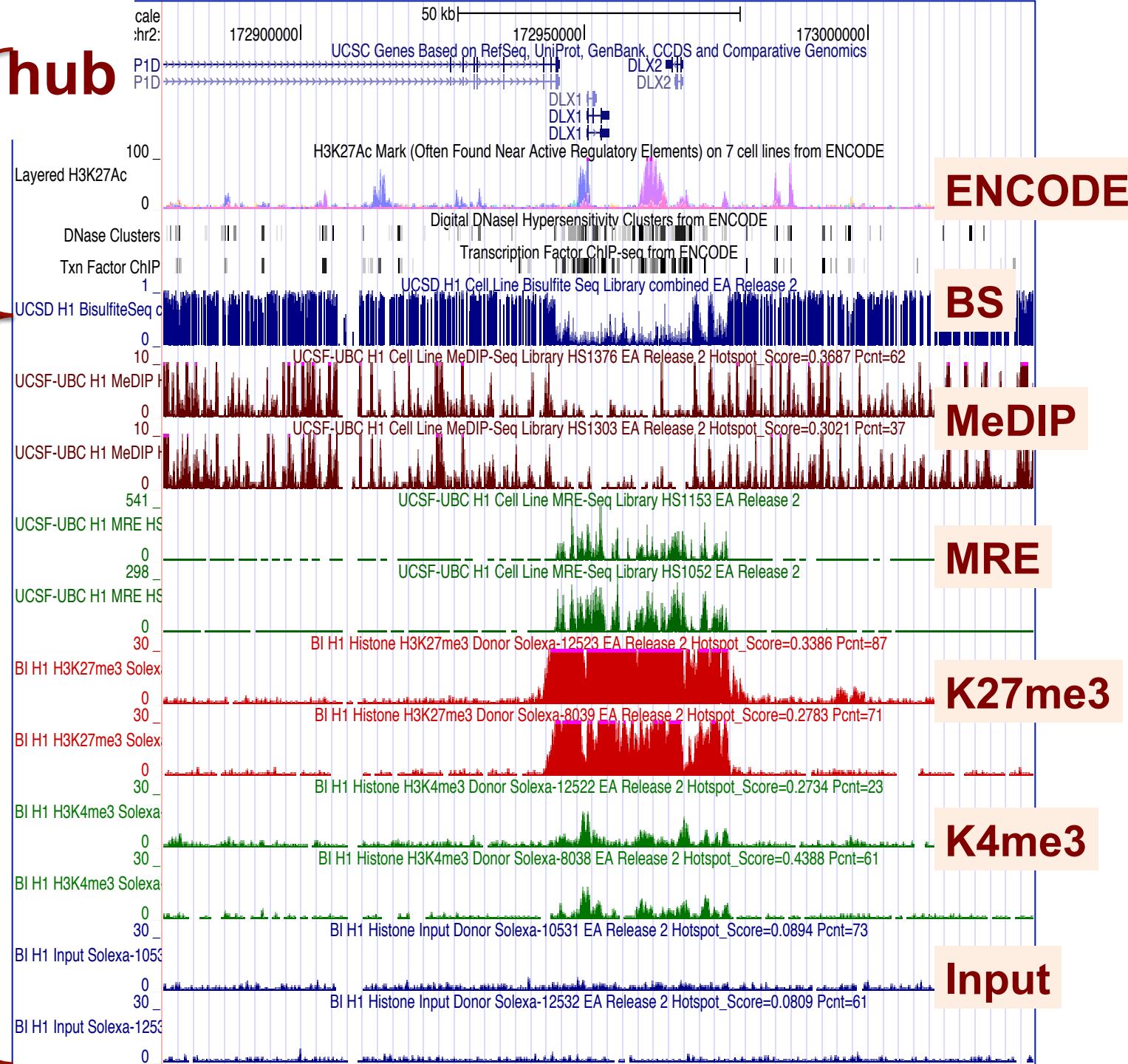
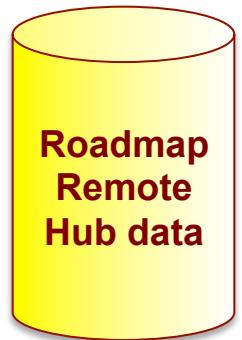
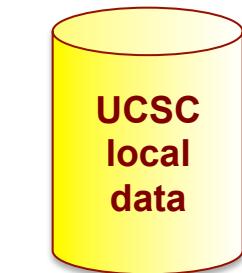
Roadmap select tracks

Native UCSC tracks **(ENCODE, etc)**

Roadmap track control

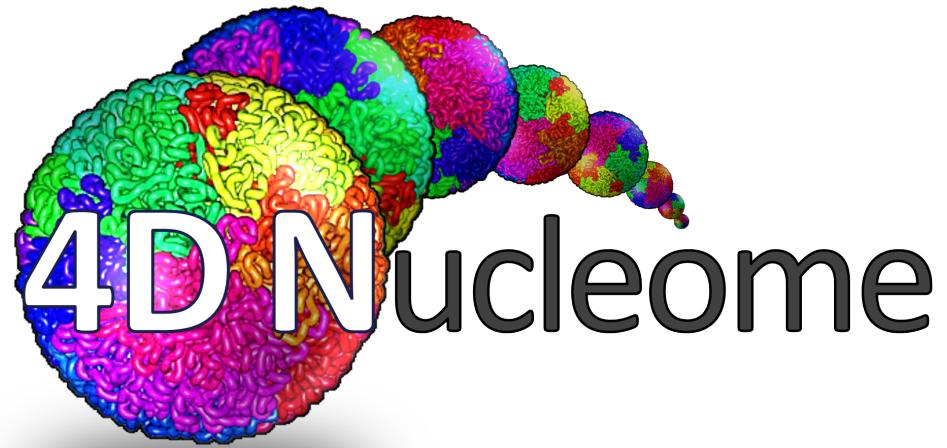
Roadmap track control

Remote hub



WashU Epigenome Browser: Innovations in exploring epigenomic data

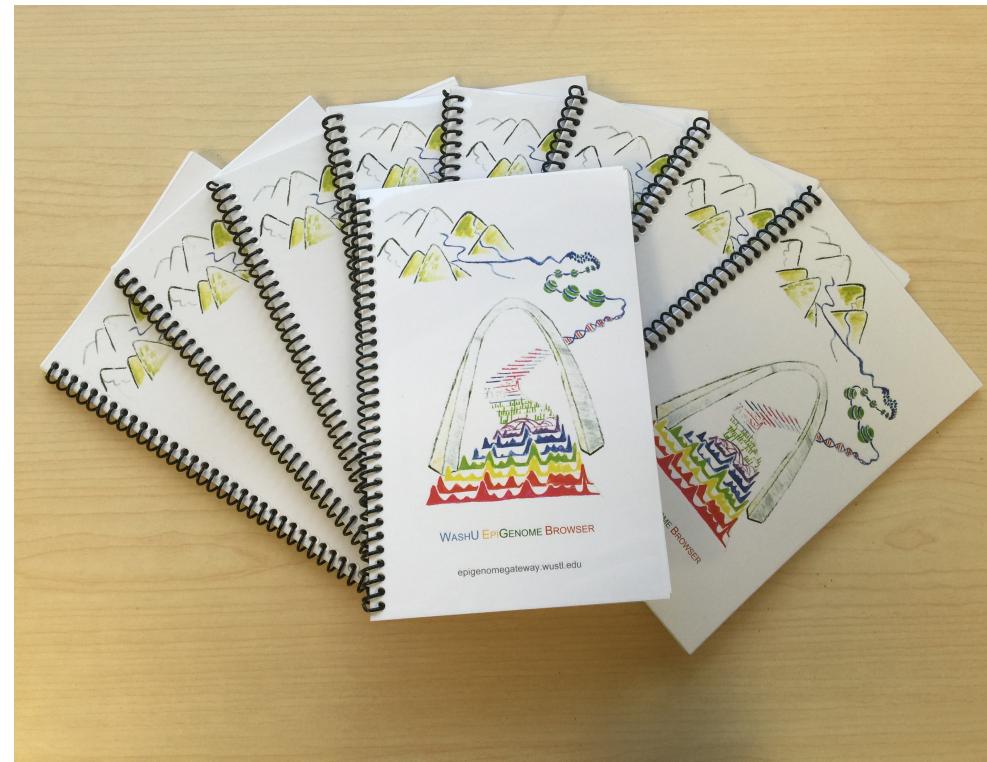
- **Community resource for multi-dimensional data**
 - Browse hundreds of data tracks in one view
 - Display epigenomic data alongside with their metadata
 - Viewing data on specific genomic features, genesets, or pathways
 - Outreach activities
 - **Google-map style browsing**
 - Heatmap, wiggle map, going from whole genome to single base
 - Zoom and pan, drag and drop, extensive user configuration
 - Sessions, custom track, custom hub, hub cluster
 - **Visualizing non-linear genome**
 - Display chromatin interaction data (HiC, ChIA-PET, etc)
 - **Visual bioinformatics engine**
 - Real time statistical analysis
 - Common genomic data summary graph
 - **Continued development**
 - Novel, more expressive visual format
 - Comparative epigenomics
 - Cancer genome browser
 - Browser of transposable elements
- Nat Methods 2011, 8:989-990
Curr Protoc Bioinformatics. 2012
Nat Methods 2013, 10:375–376
Bioinformatics, 2014, 30(15):2206-7
Nat Biotechnology, 2015, 33, 345–346



TaTARGET

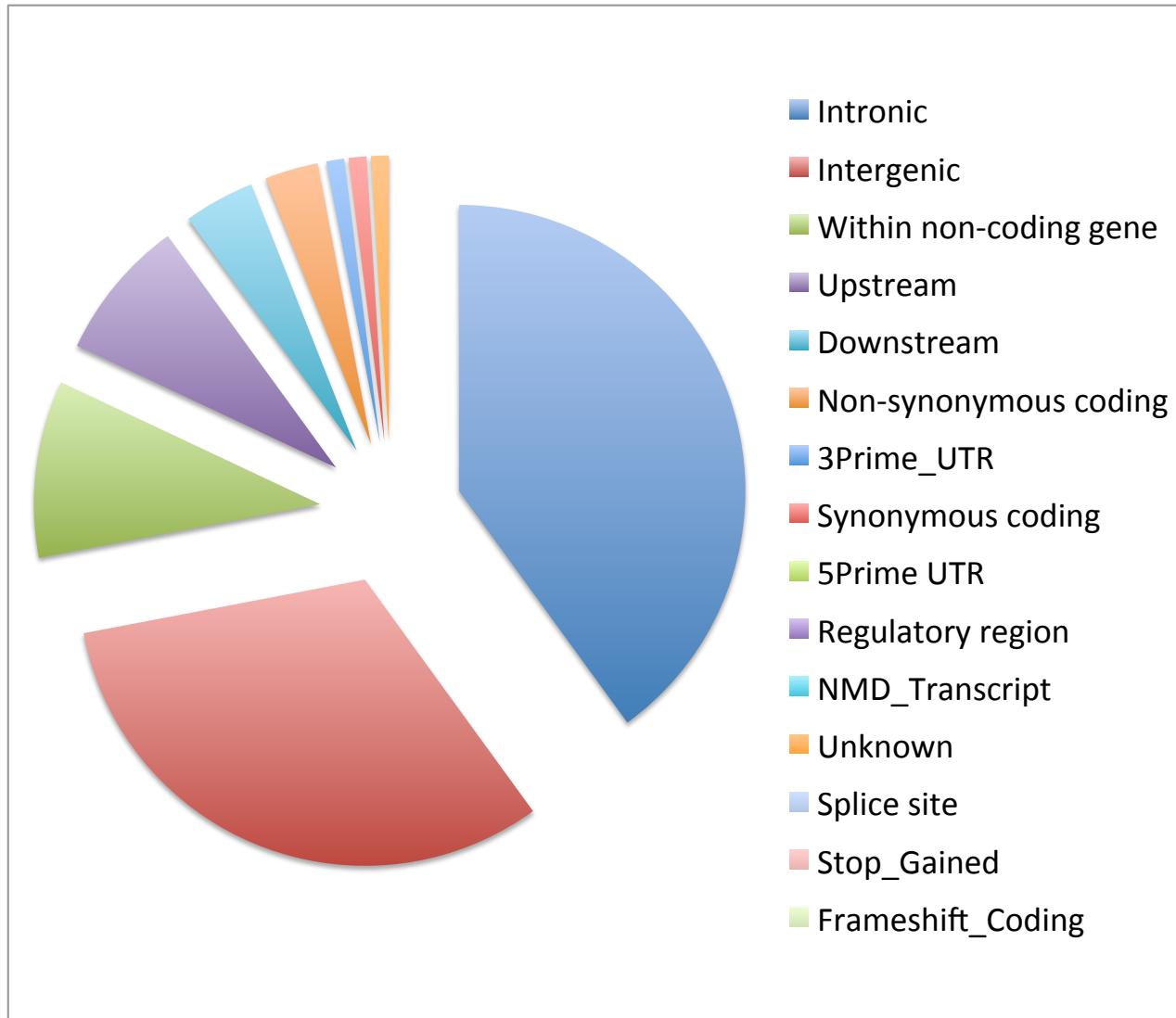
Outreach and education

- Support websites
 - Wiki, helpdesk, mailing list
 - Documentations, tutorial book, video demo
- User community and social media
 - Facebook
 - Google+
- Workshops
 - CSHL
 - ASHG 2014
 - Keystone 2015
 - Shanghai 2015
 - Puerto Rico 2016
 - SOT 2016
 - Keystone 2016
 - ASHG 2016



1. Epigenomic annotation of genetic variants

I got my GWAS hits! Now what?



I got my GWAS hits! Now what?

LETTER

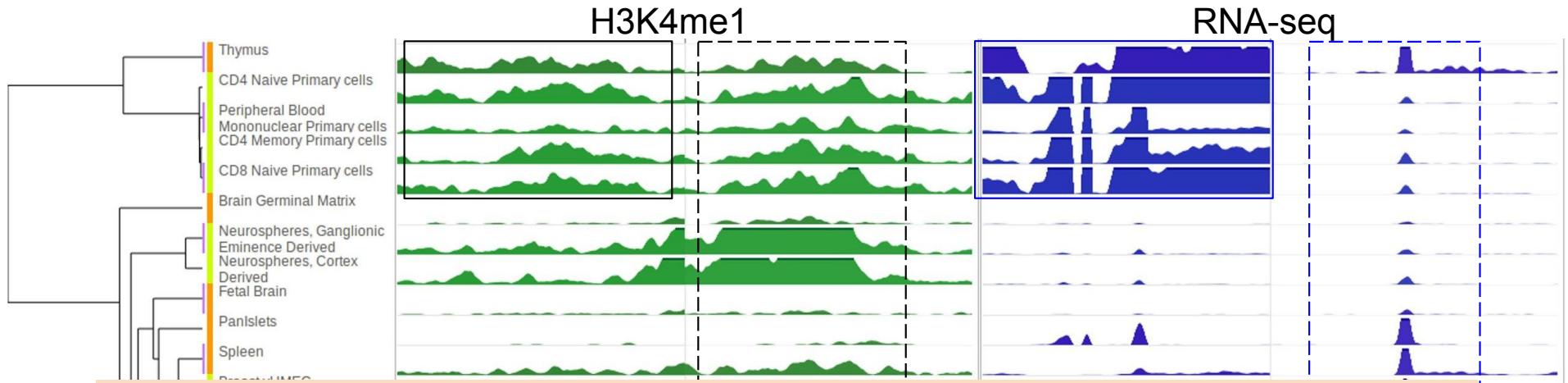
doi:10.1038/nature10251

Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis

The International Multiple Sclerosis Genetics Consortium* & the Wellcome Trust Case Control Consortium 2*

Multiple Sclerosis:

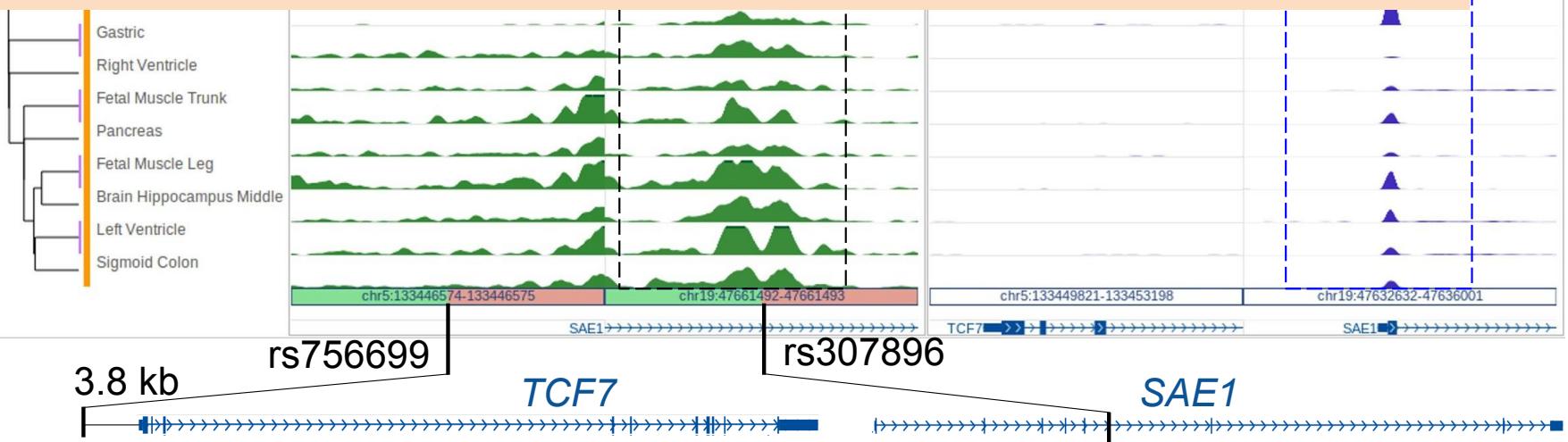
- an autoimmune disease that affects the brain and spinal cord (central nervous system).
- a collaborative GWAS involving 9,772 cases of European descent collected by 23 research groups working in 15 different countries



Roadmap Epigenome Browser

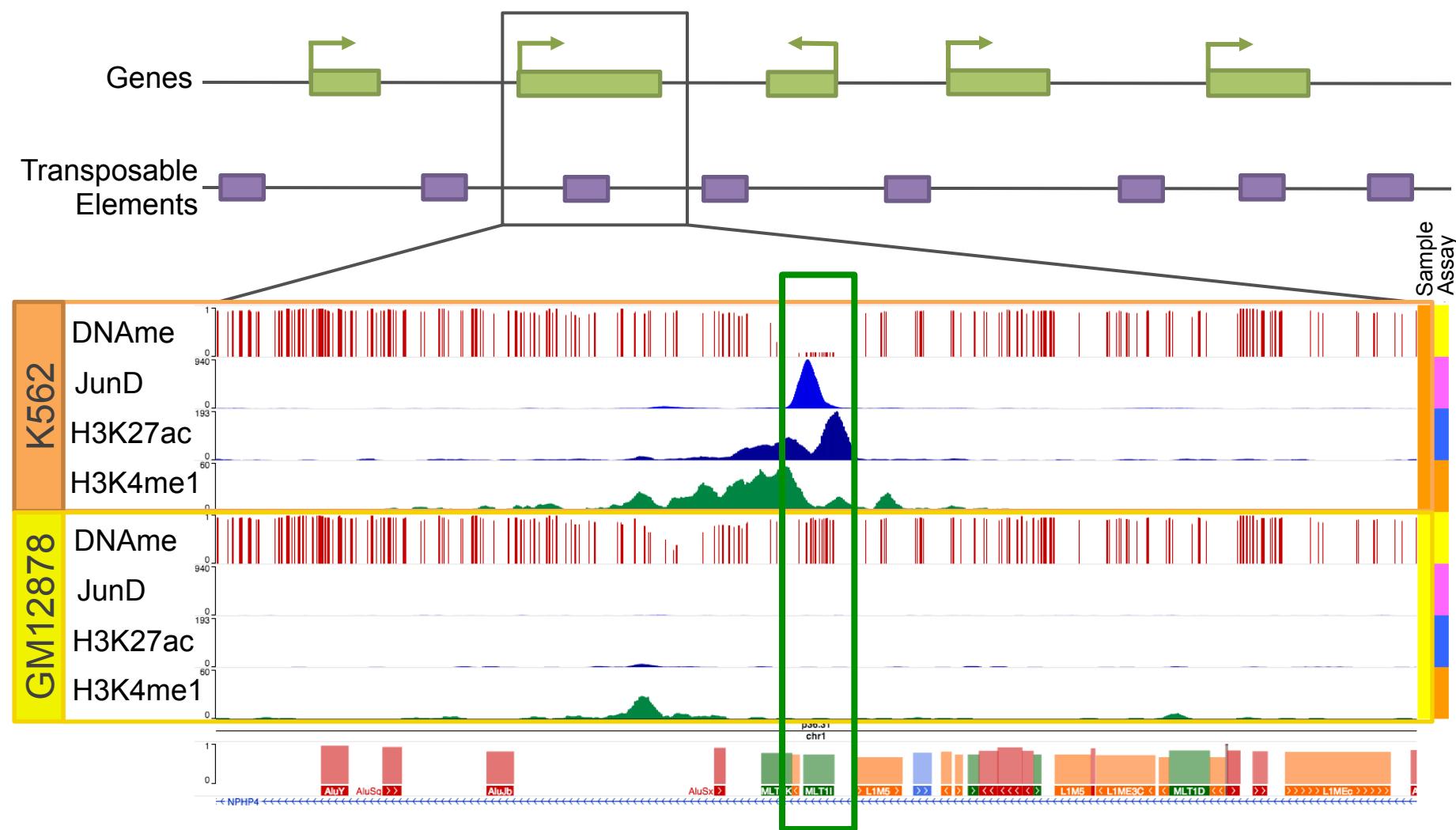
<http://epigenomegateway.wustl.edu/browser/roadmap/>

Nat Biotechnology, 2015

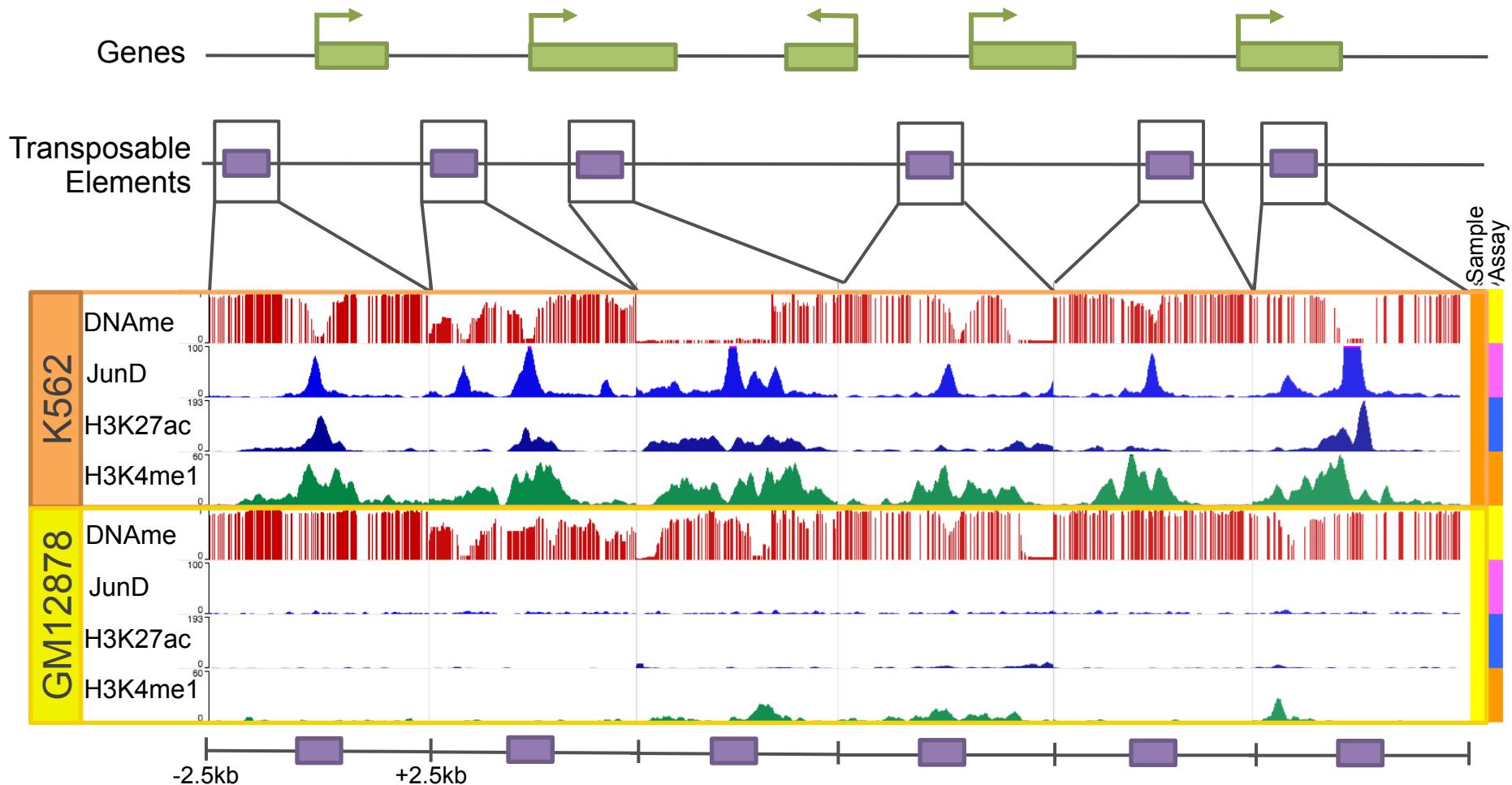


2. Transposable elements and gene regulation

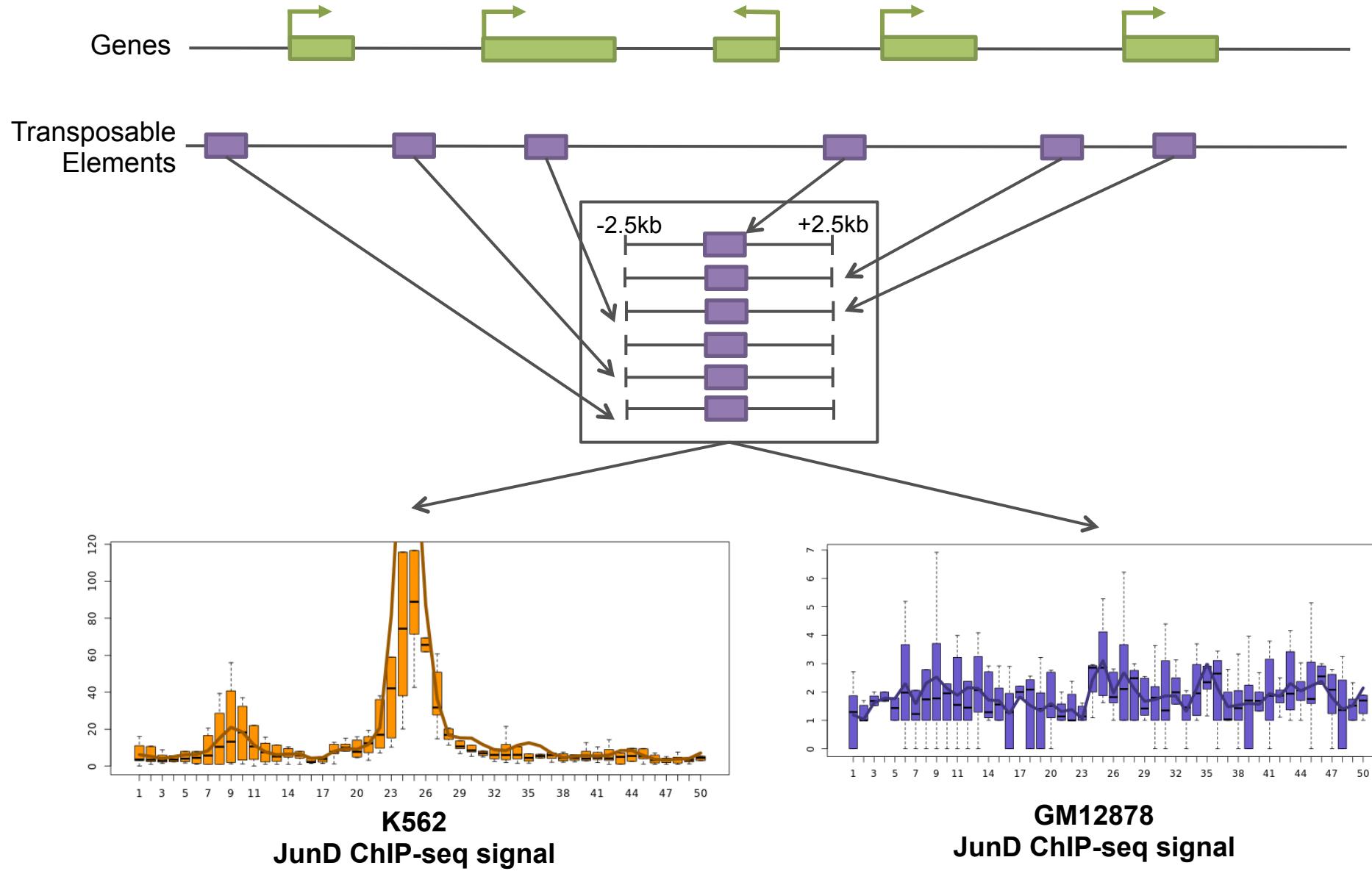
Linear genome view (using ENCODE data)



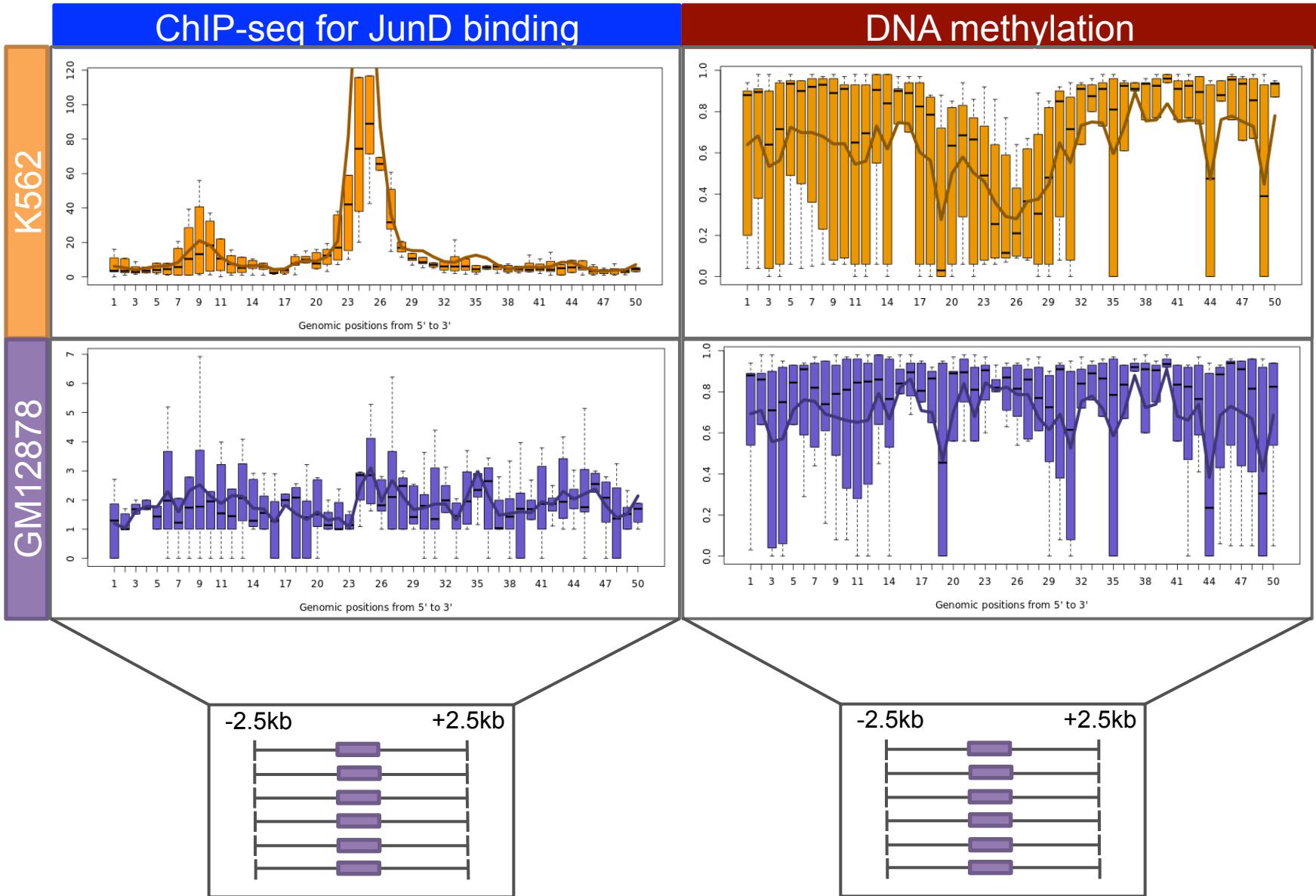
Gene/Region-set view



Summarized view



Summarized view



3. Cancer cell identity crisis hypothesis

Cancer cell identity crisis hypothesis

Brain
Cancer

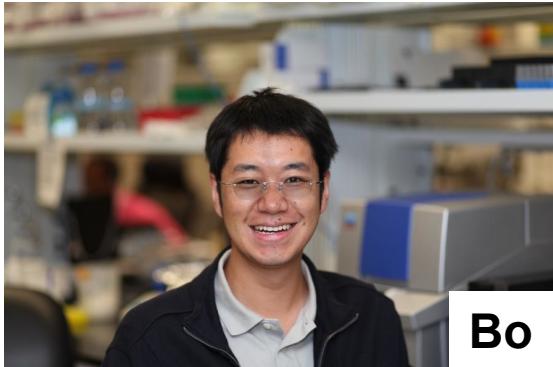


Joseph Costello
UCSF

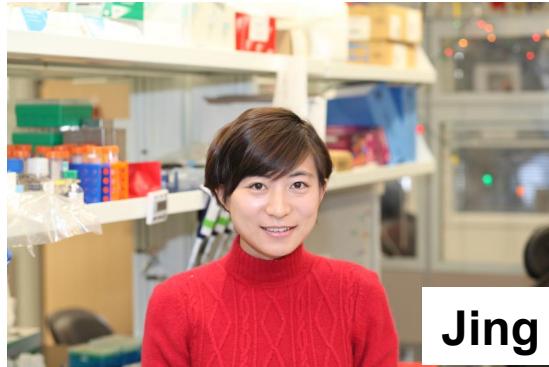
Endometrial
Cancer



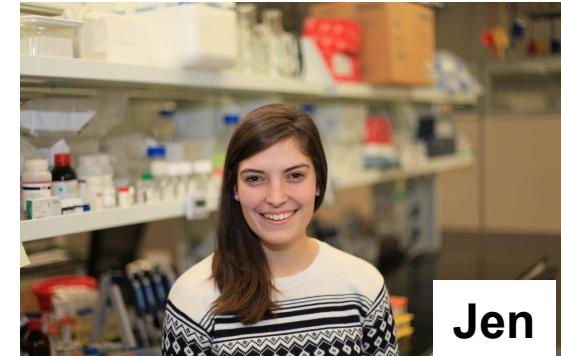
Paul Goodfellow
Ohio State



Bo



Jing



Jen

Nagarajan, Genome Research 2014
Zhang, BMC Genomics, 2014

DNA Methylocomics

Methylated CpGs

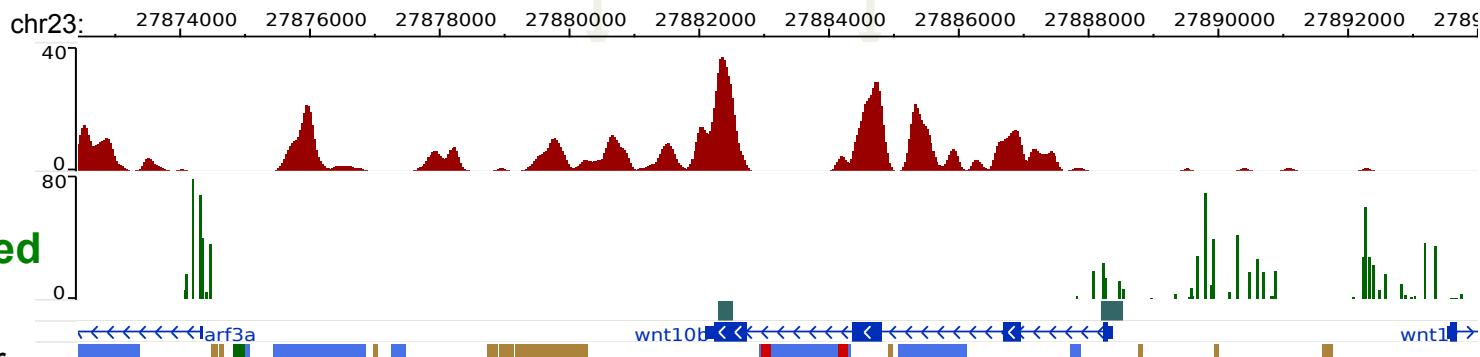
methyl DNA
Immunoprecipitation

MeDIP

Unmethylated CpGs

methylation-sensitive
restriction digestion

MRE



3' CpG island is
methylated

5' CpG island is
unmethylated



Joe Costello



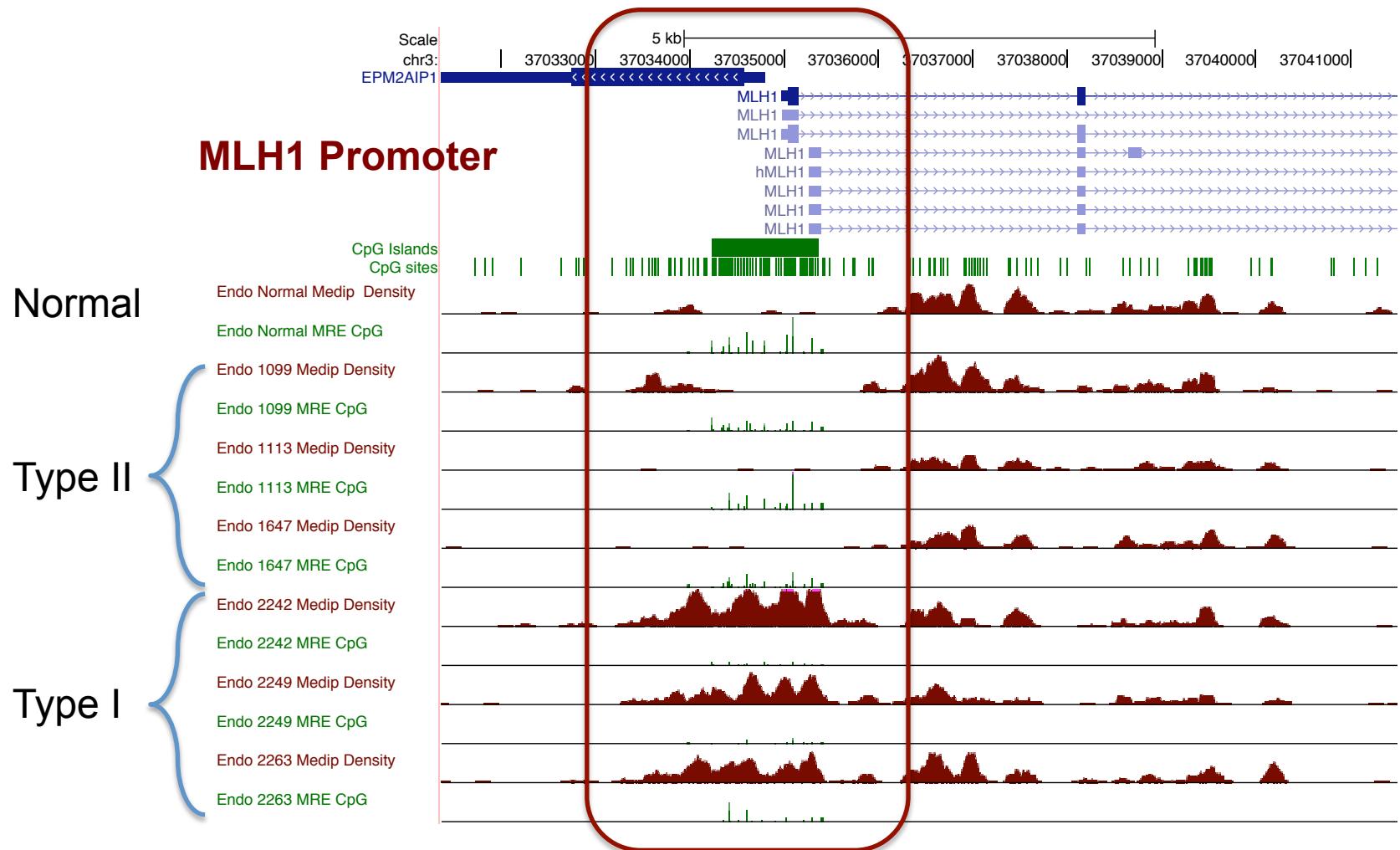
Xiaoyun Xing



Daofeng Li

In collaboration with Joe
Costello, UCSF
Nature 2010
Li et al., Methods, 2015

Endometrial cancer type-specific methylation: MLH promoter

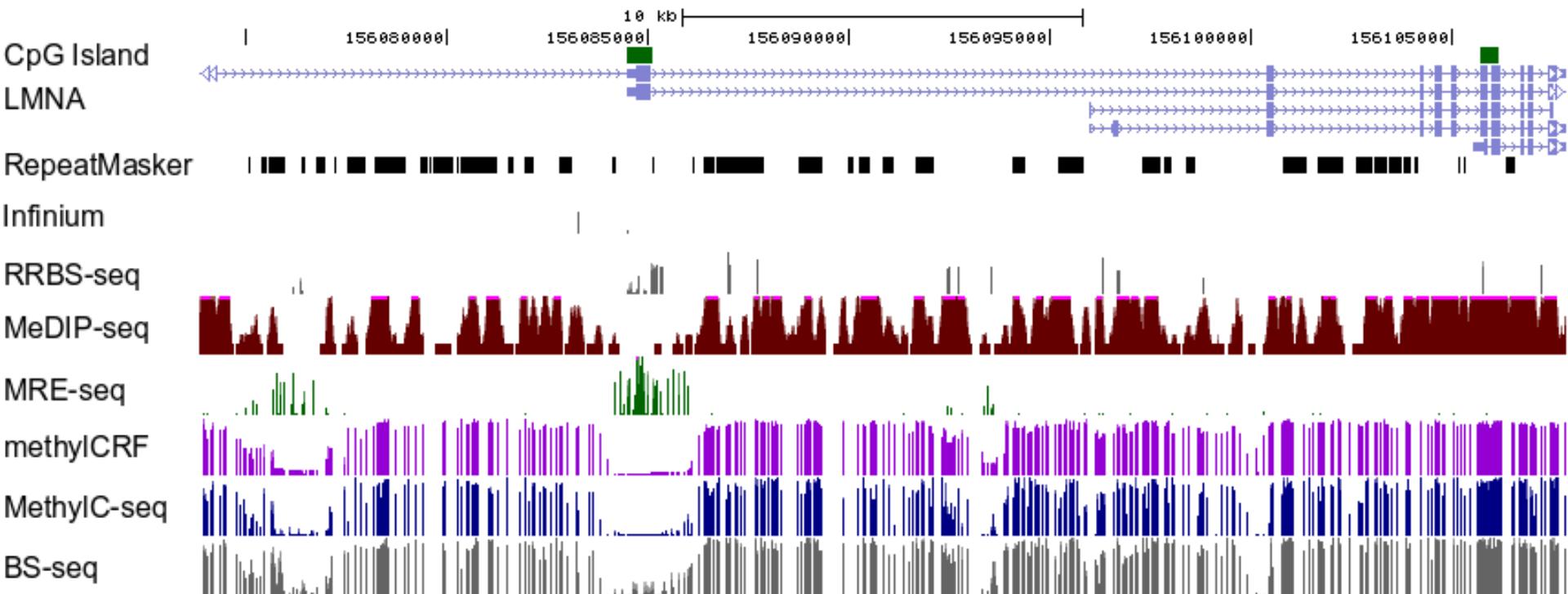


In collaboration with Paul Goodfellow

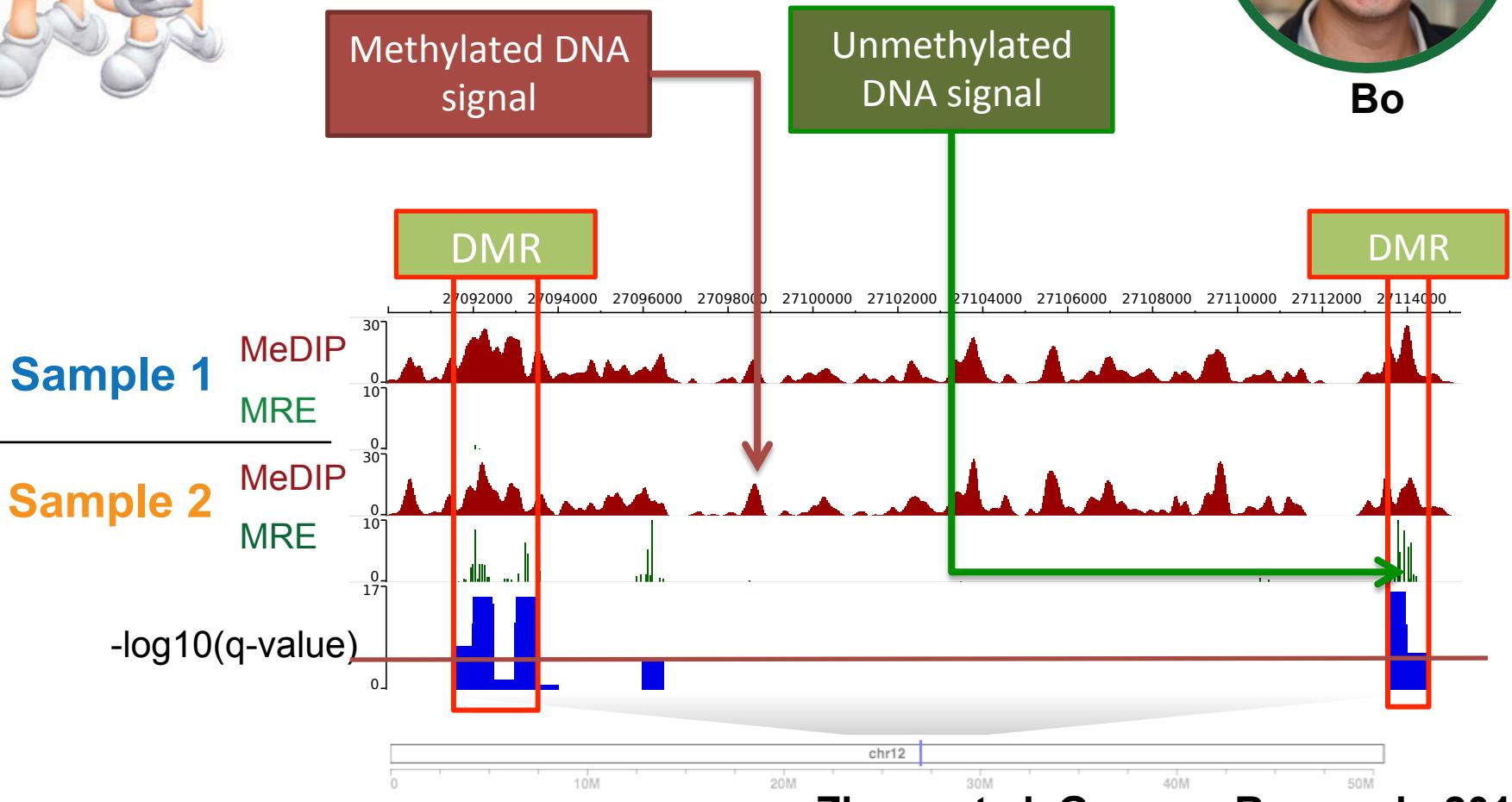
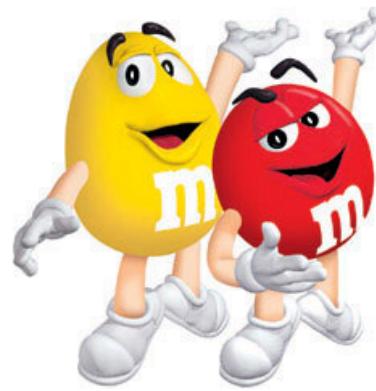


Michael

MethylCRF: a Conditional Random Field-based algorithm to predict DNA methylation level at single base resolution



M&M finds differentially methylated regions (DMRs) genome-wide





Vast majority of tissue-, cell type-, developmental, and/or disease-specific DMRs are marking enhancer elements



Rebecca

Skin, surface ectoderm

Nature Comm., 2014



Hyung Joo

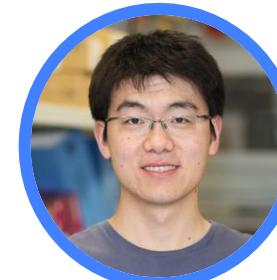
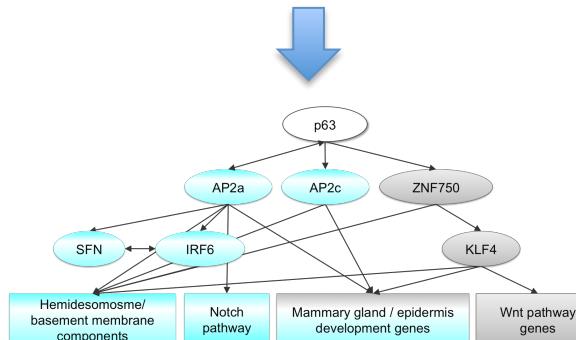
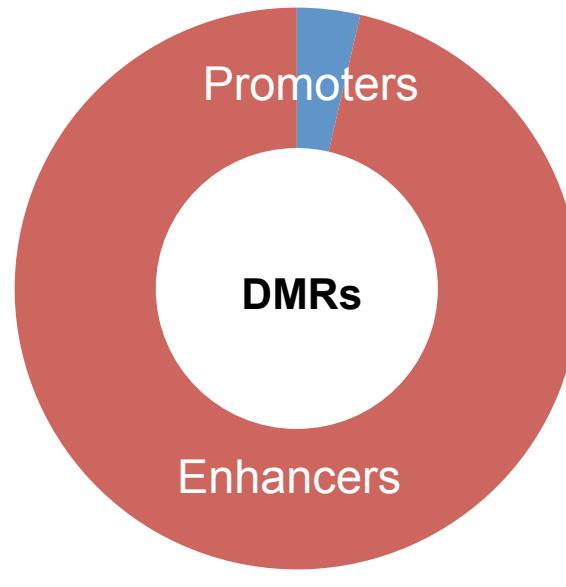
Zebrafish embryogenesis

Nature Comm., 2015



GiNell

Intermediate methylation,
Nature Comm., 2015



Junchen

Variable methylation,

G3, 2016



Bo



Jing

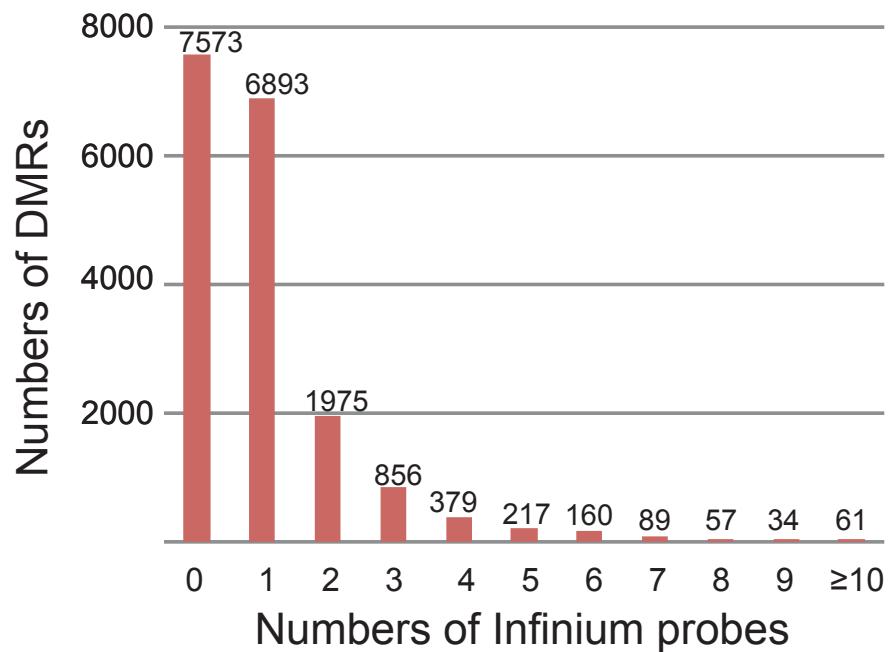
Endometrial cancer, GBM

Genome Research 2014

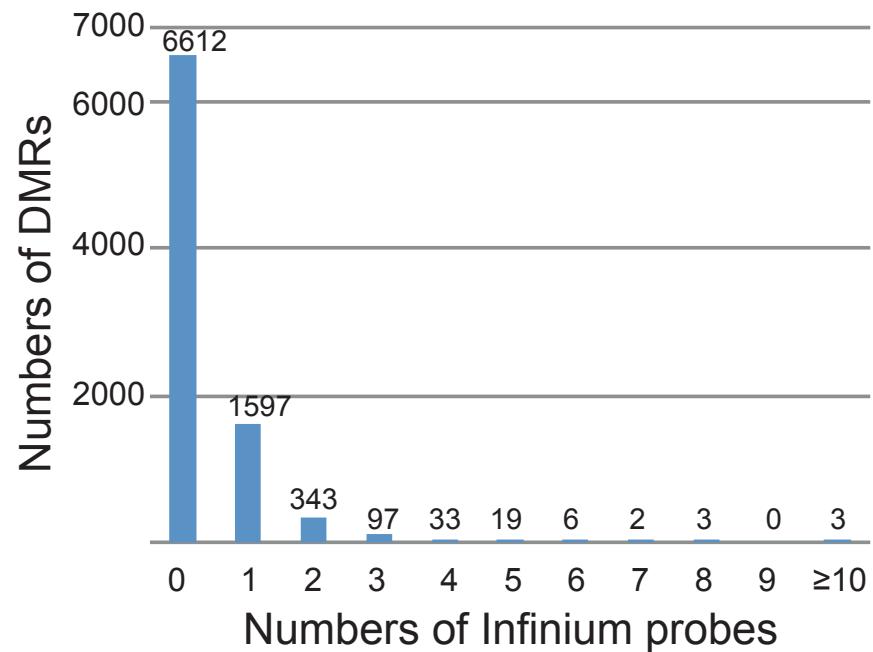
BMC Genomics, 2014

DMRs do not fall into classic “regions of interest”

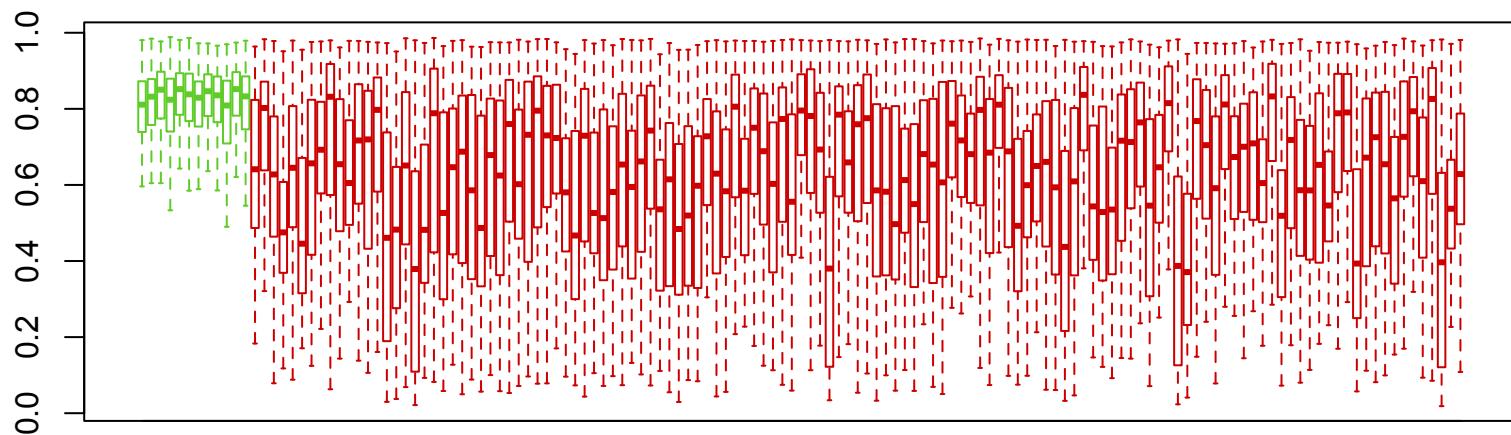
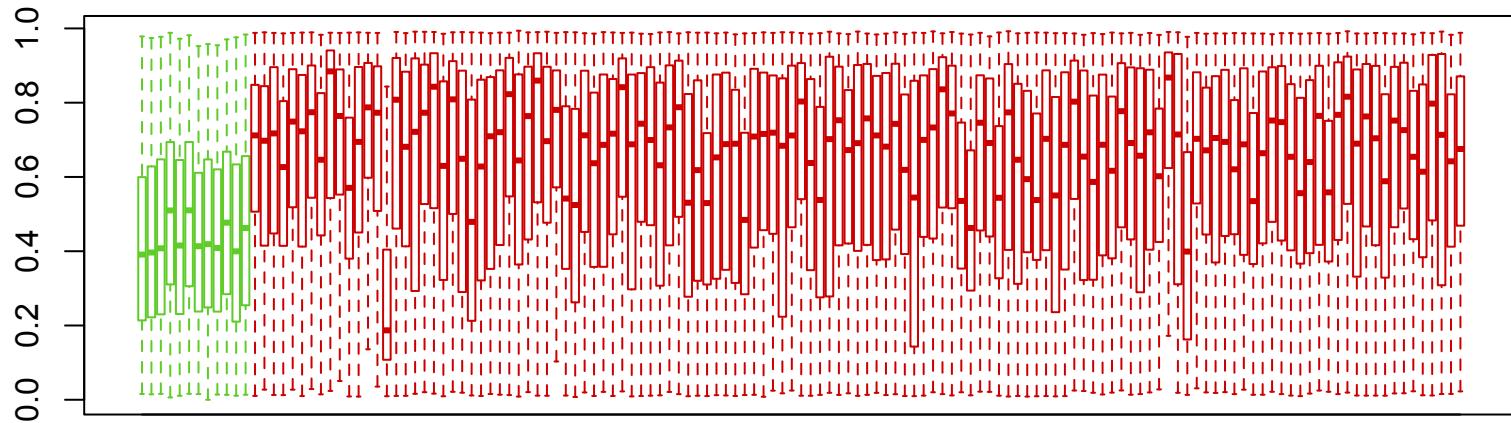
Distribution of Infinium probes in Hypermethylated EAC DMRs



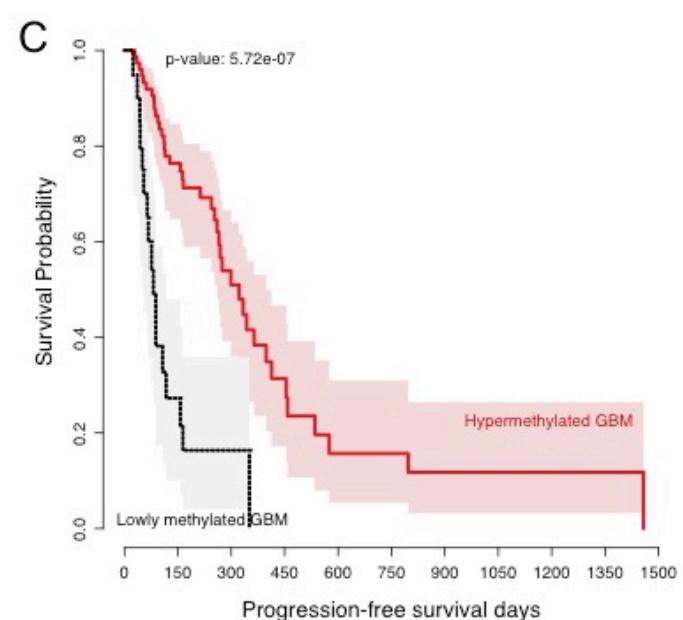
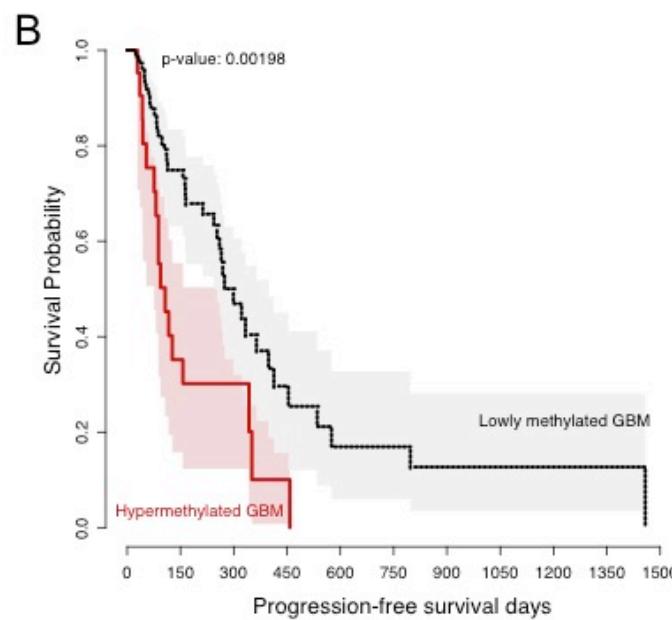
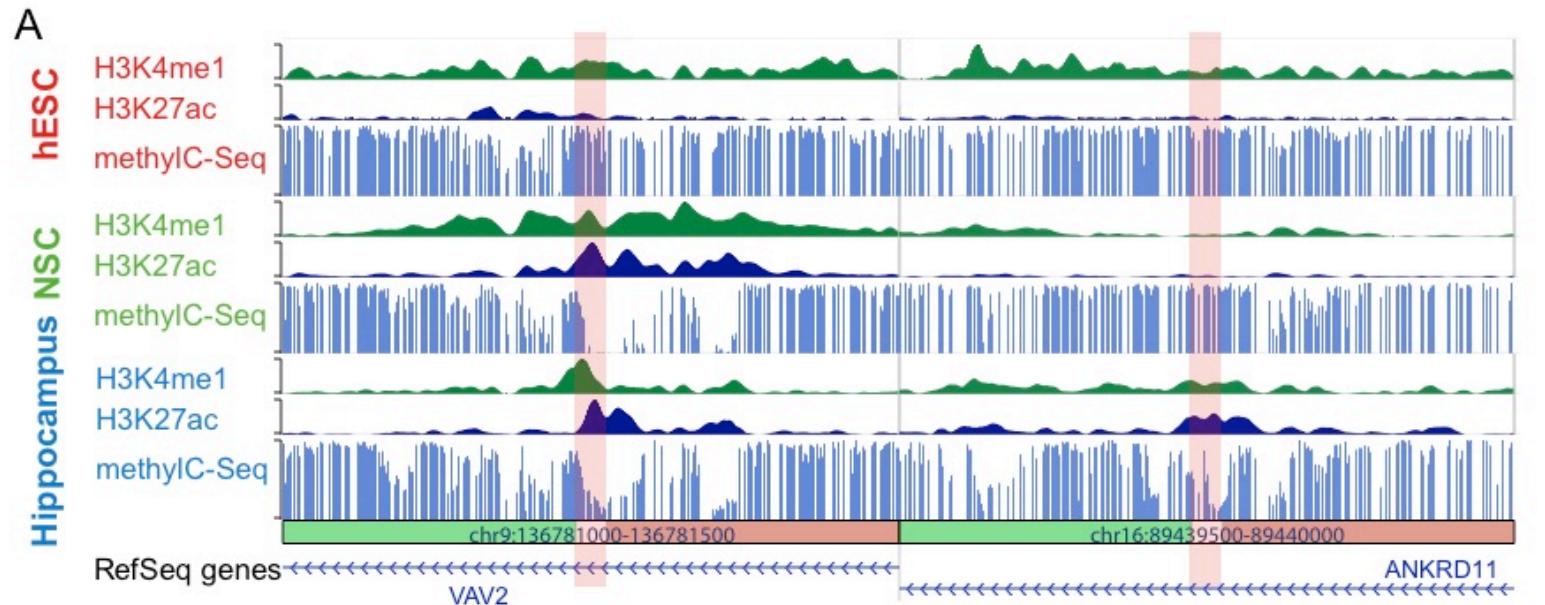
Distribution of Infinium probes in Hypomethylated EAC DMRs



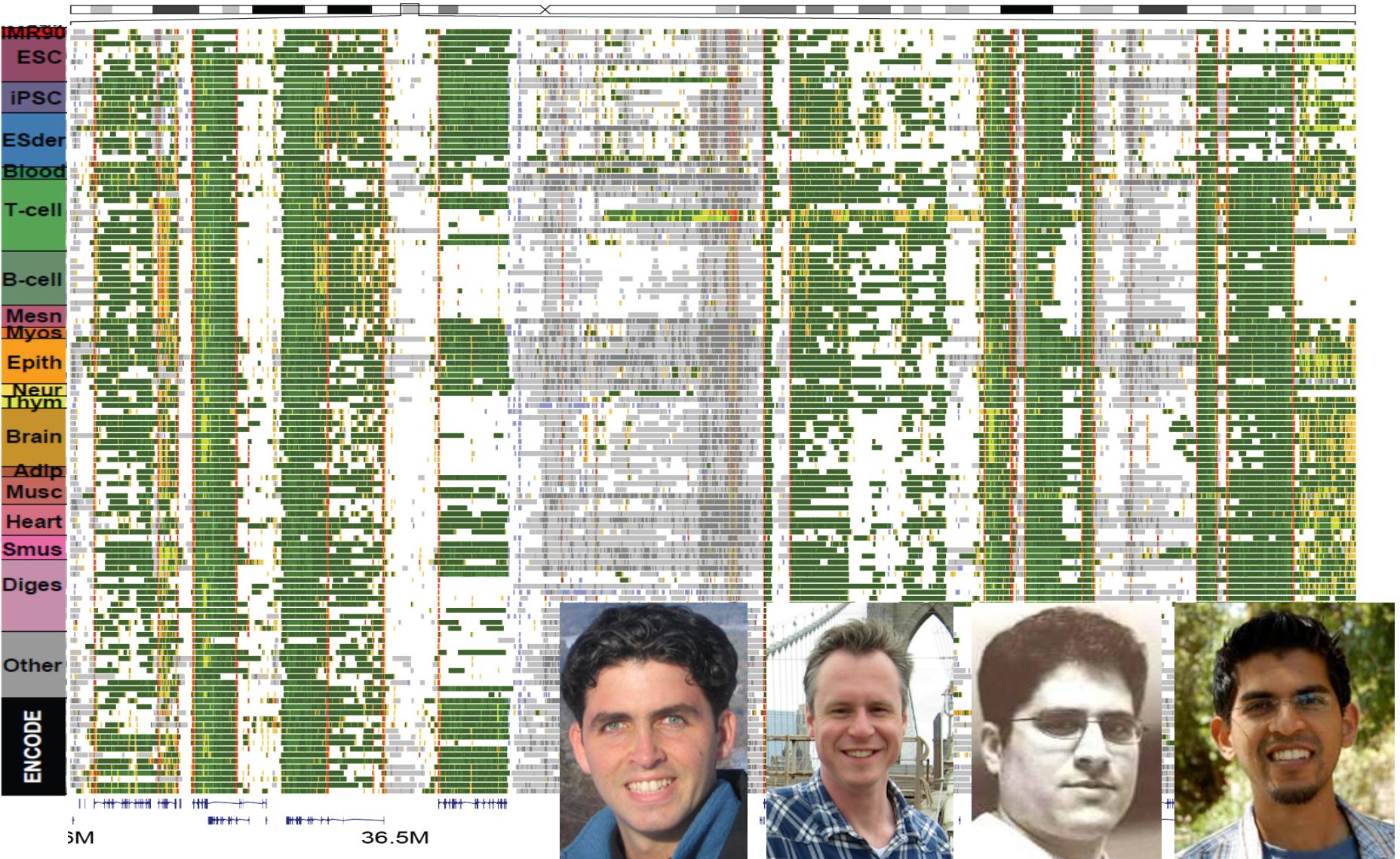
TCGA 450k data as a “validation panel”



DMRs are a novel biomarker



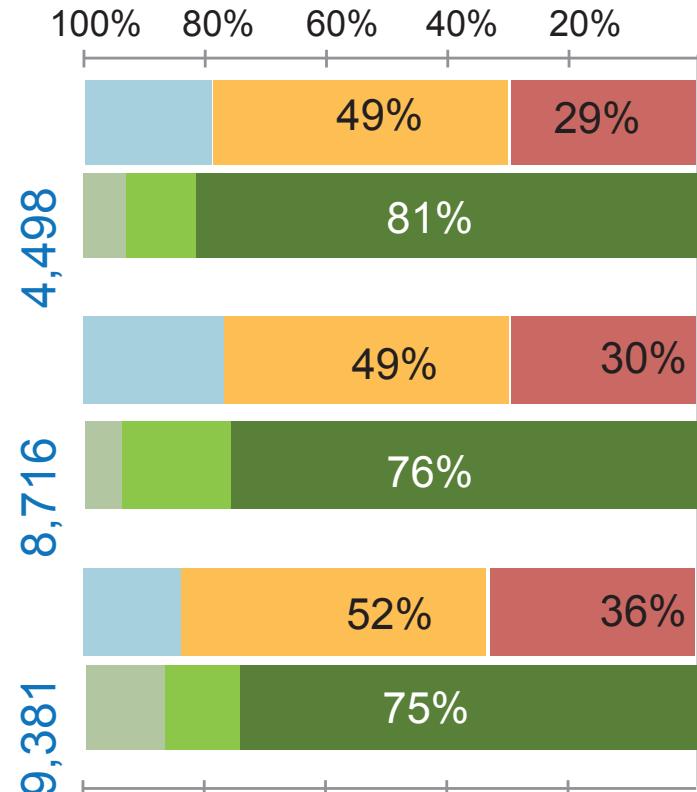
Chromatin state annotations across 127 epigenomes



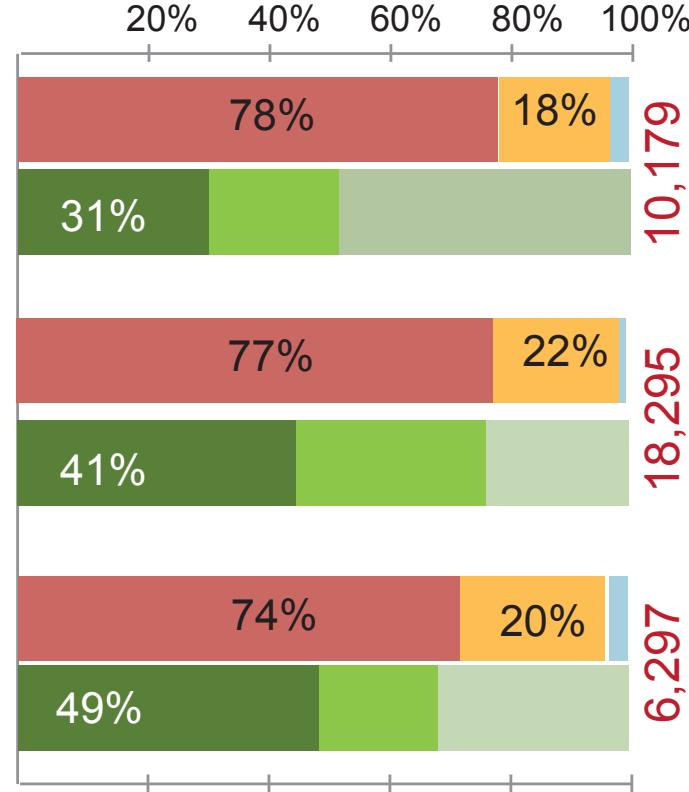
ChromHMM: Manolis Kellis, Wouter Meuleman, Jason Ernst, Anshul Kundaje

DMRs overlap tissue-specific regulatory elements

Cancer hypomethylated DMRs



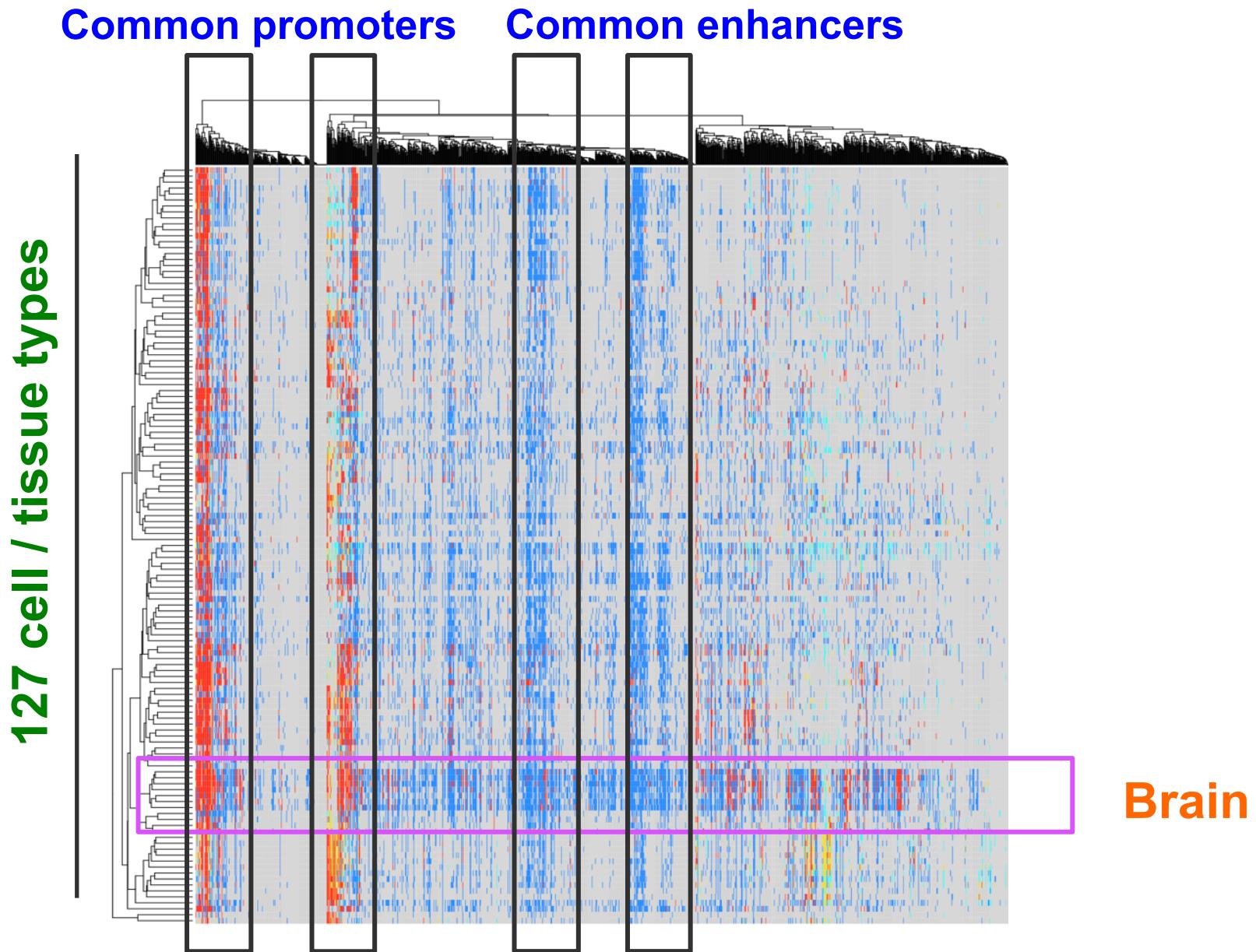
Cancer hypermethylated DMRs



- DMRs with enhancer annotation
- DMRs with other annotation
- DMRs w/o infinium 450K probe
- DMRs with more than 1 infinium 450K probe

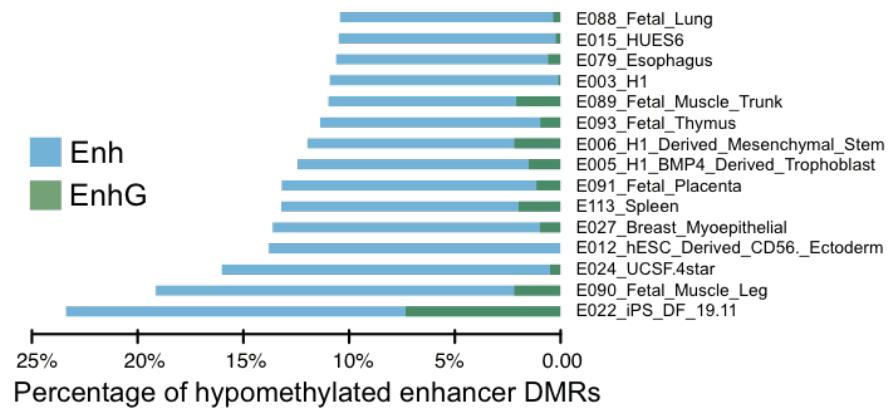
- DMRs with promoter annotation
- DMRs with transposable elements
- DMRs with 1 infinium 450K probe

Hypermethylation of enhancers in GBM

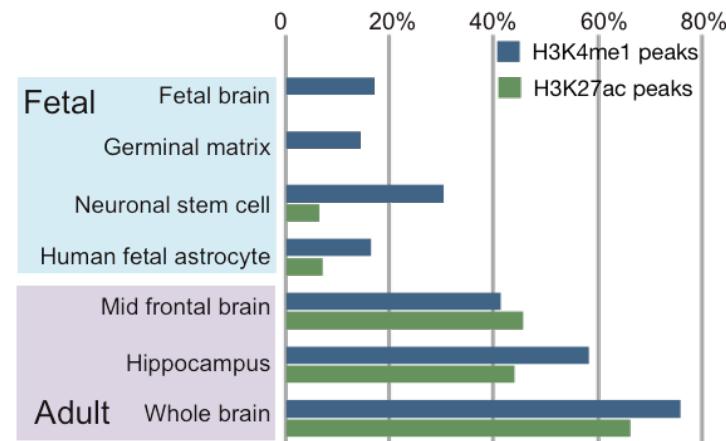


Hyper/hypomethylation of enhancers in GBM

A

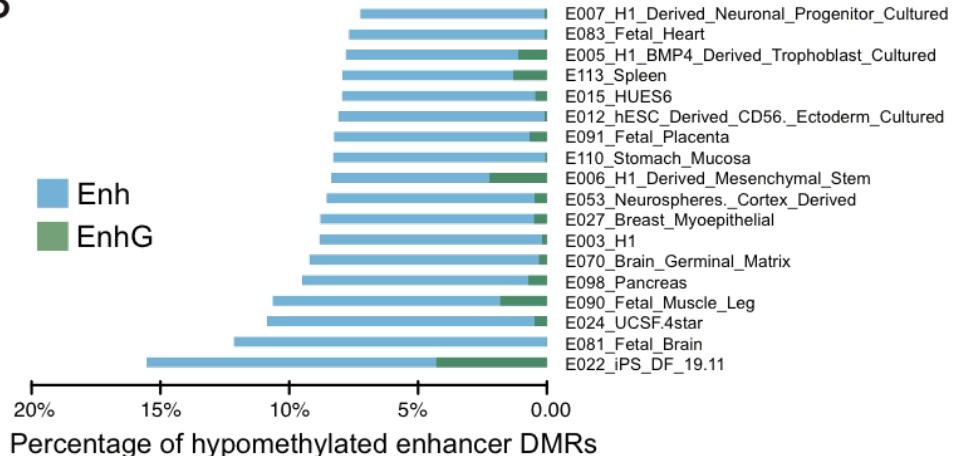


C

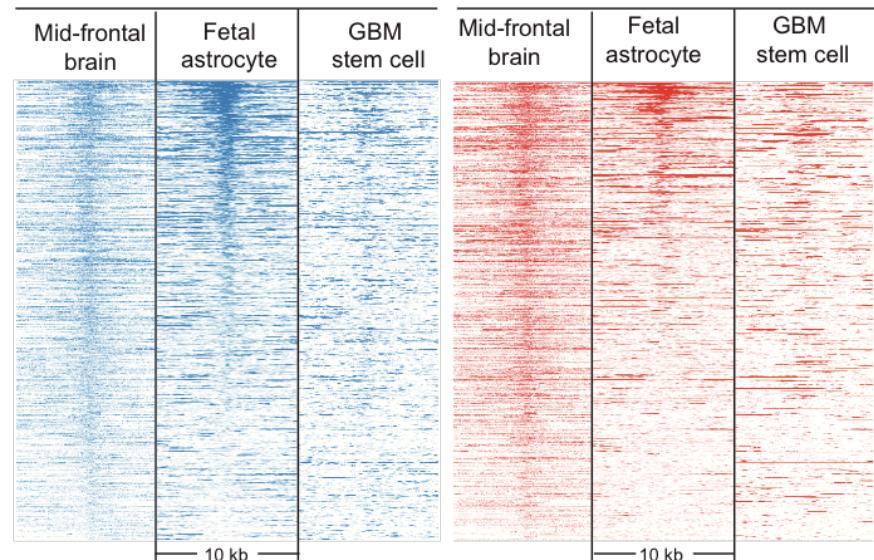


D

B



H3K4me1 signal



Recurrent hypermethylated DMRs Recurrent hypomethylated DMRs

EAC

15,415

2,760

7,332

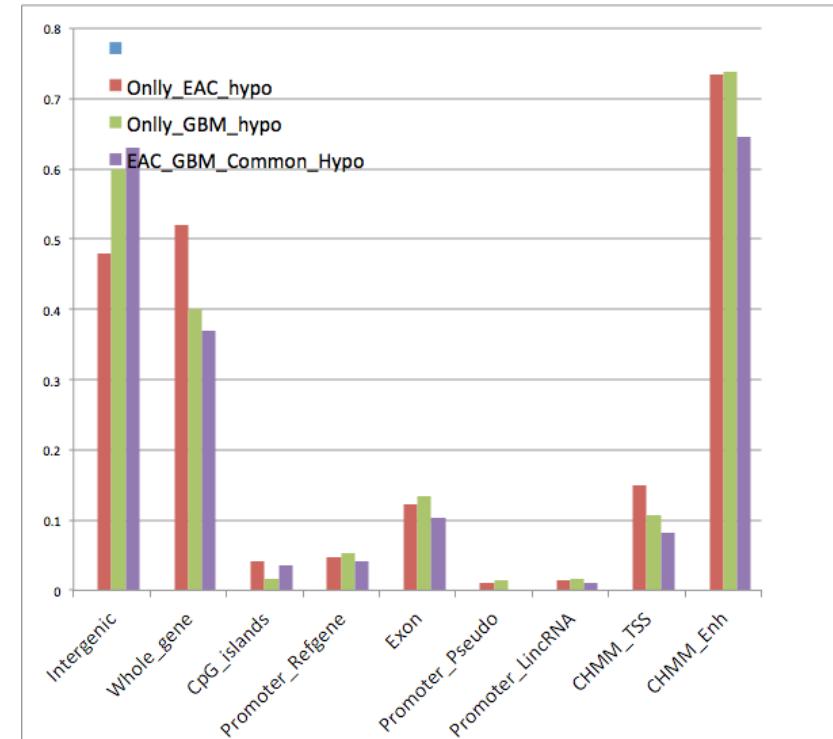
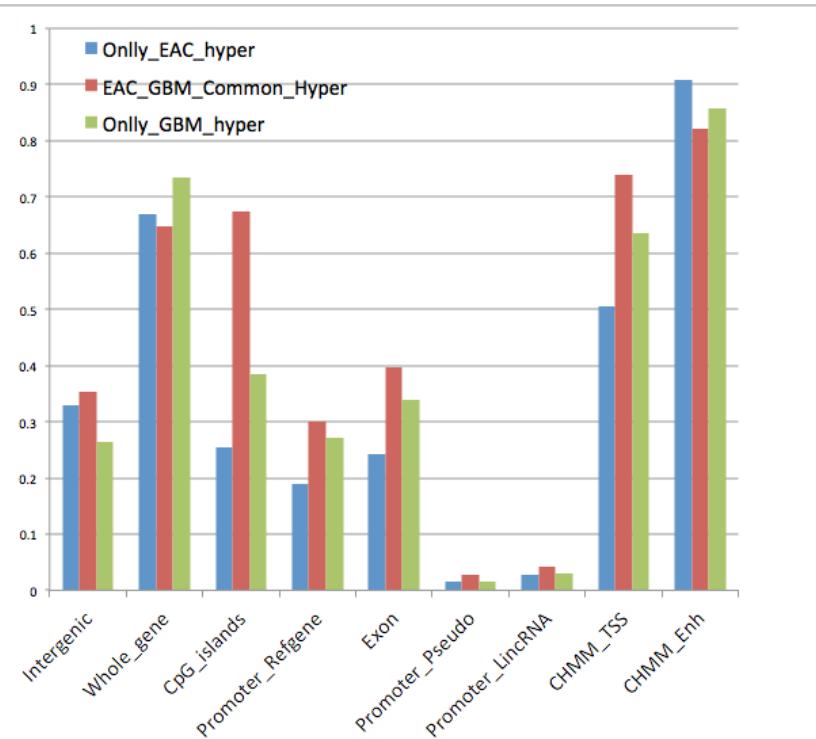
EAC

8,430

195

4,198

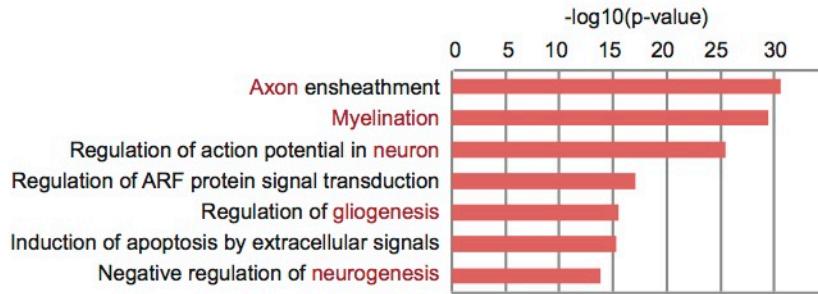
GBM



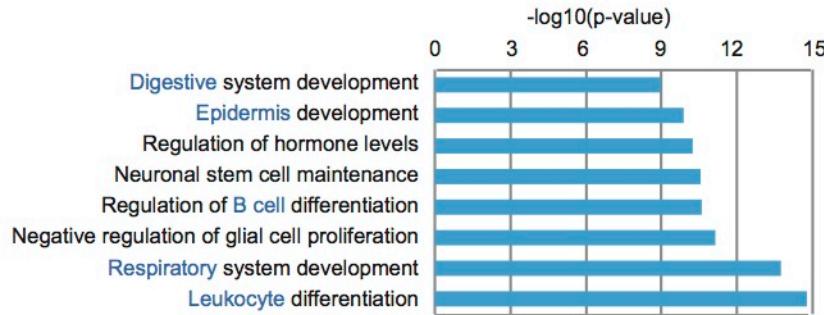
Functional annotation of enhancers with epimutations reveals disruption of distinct cellular processes

Glioblastoma multiforme

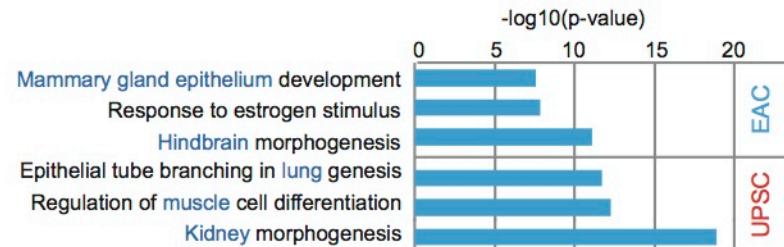
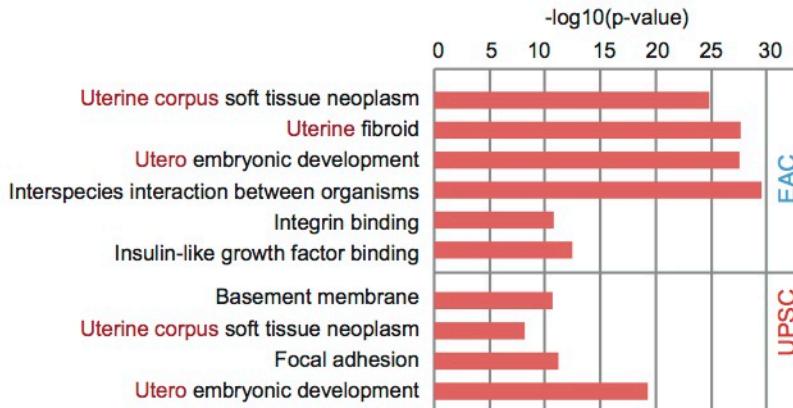
Functional enrichment of hypermethylated DMRs



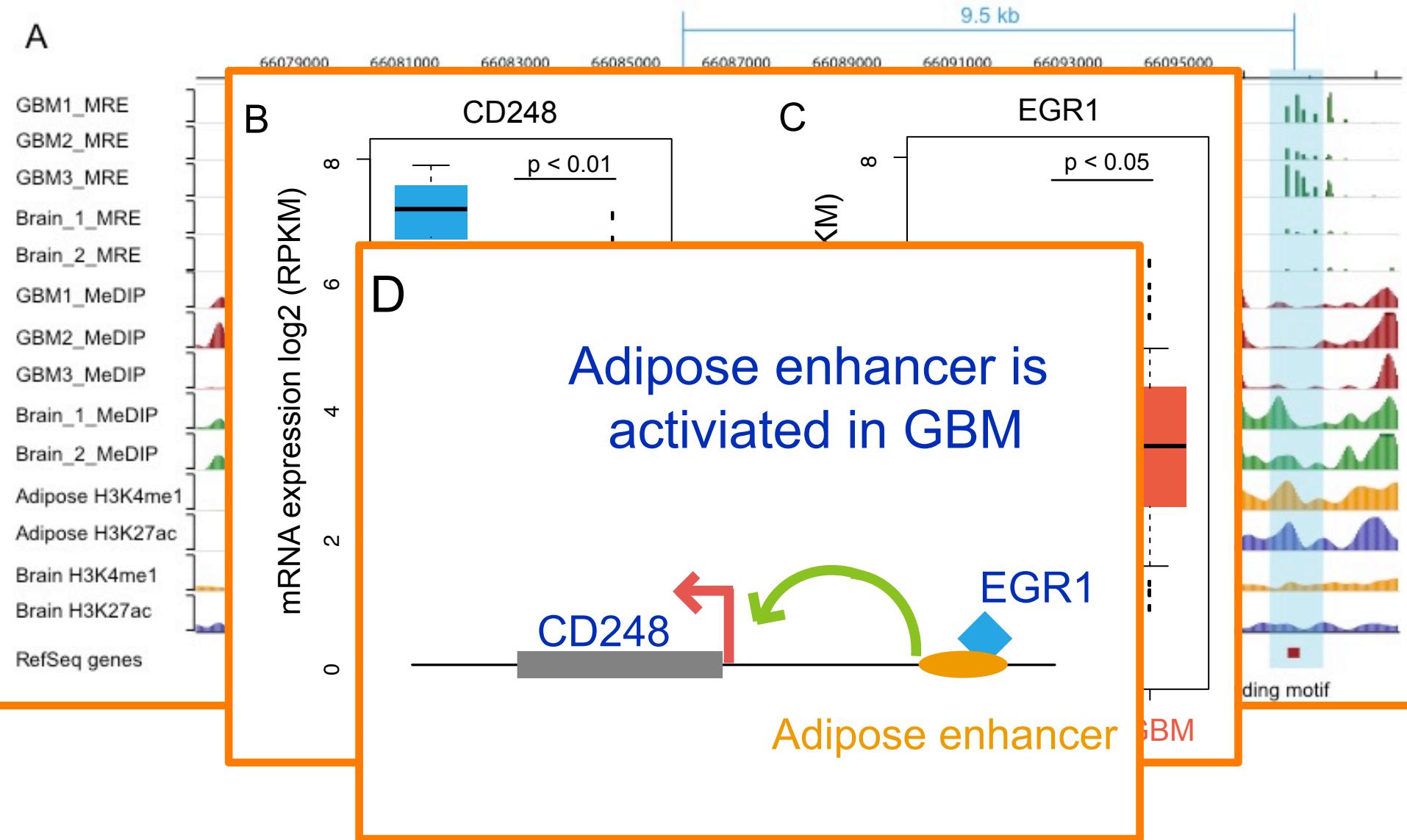
Functional enrichment of hypomethylated DMRs



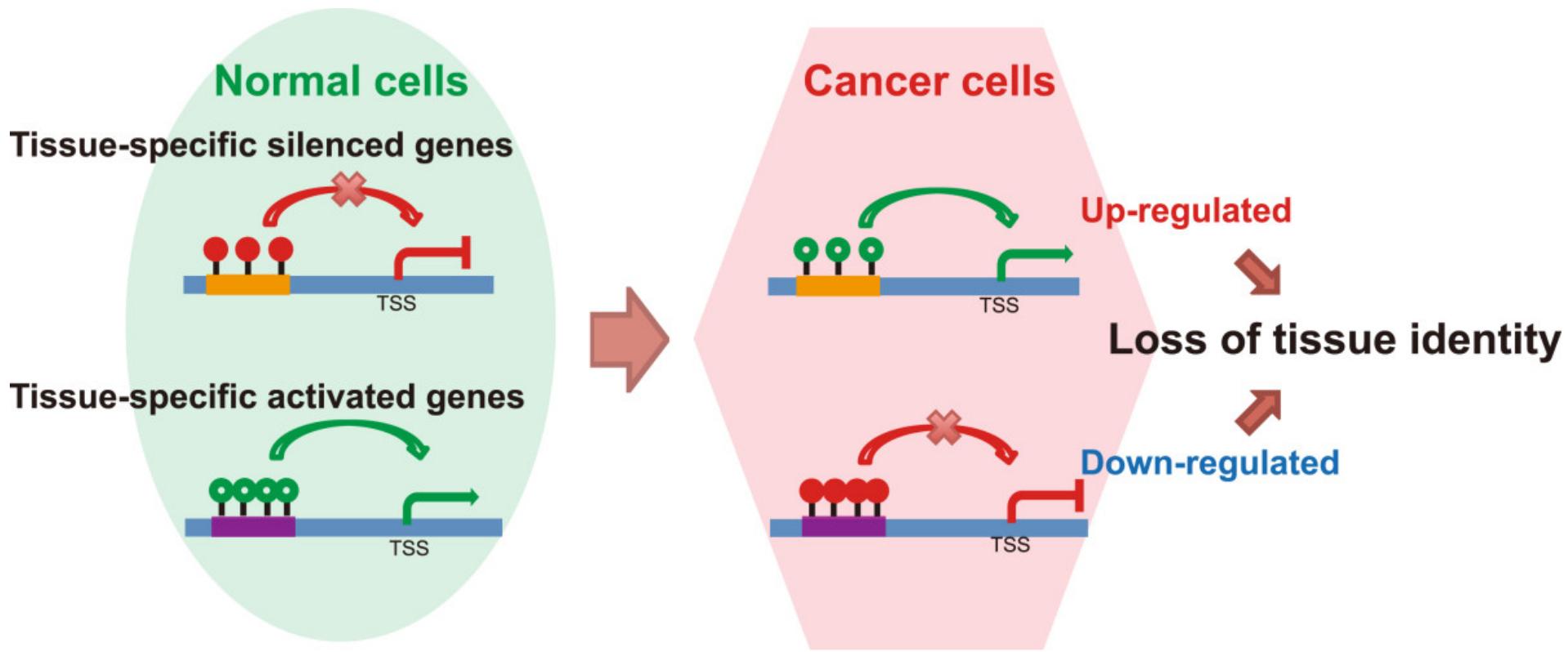
Endometrial cancer



Enhancer epimutation drives unexpected gene expression patterns



Cancer cell identity crisis hypothesis



Tissue-specific enhancers (some are TE-derived)



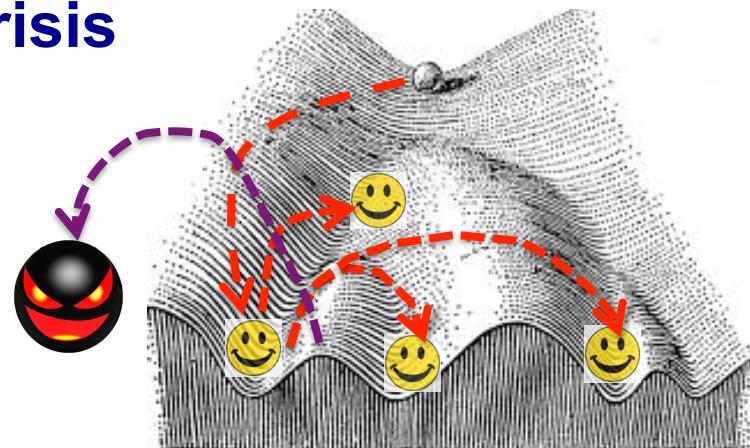
Methylated CpG



Un-methylated CpG

Summary

- Big data in biology
- Using the human reference epigenome map to annotate genetic and epigenetic variants
- DNA methylomics technology
- Enhancers/transposable elements are a major target of dis-regulated DNA methylation in cancer
- Cancer cells undergo identity crisis

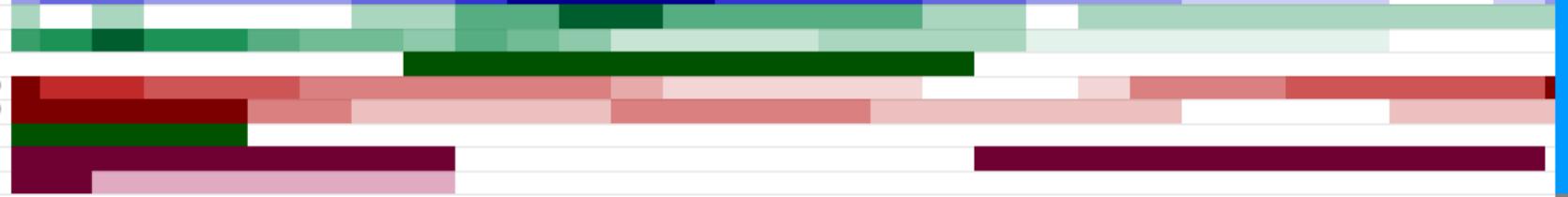
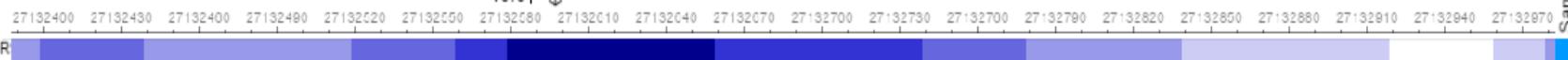


chr7:27132388-27132983

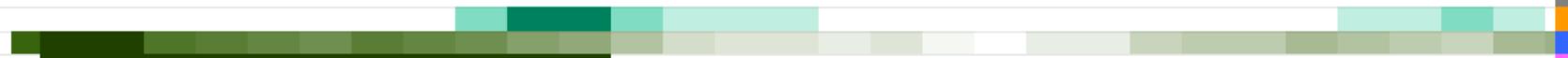
+ -½ -1 -5 ▶

One bp spans 2 pixels

Tracks Apps ⚙

Sample
Assay

QUESTION?



RepeatMasker - NO DATA IN VIEW RANGE

RepeatMasker

1-Divergence %

RefSeq genes

HOXA1 HOXA1