DukeNUS Medical School

GMS6850 – Core Concepts in Bioinformatics

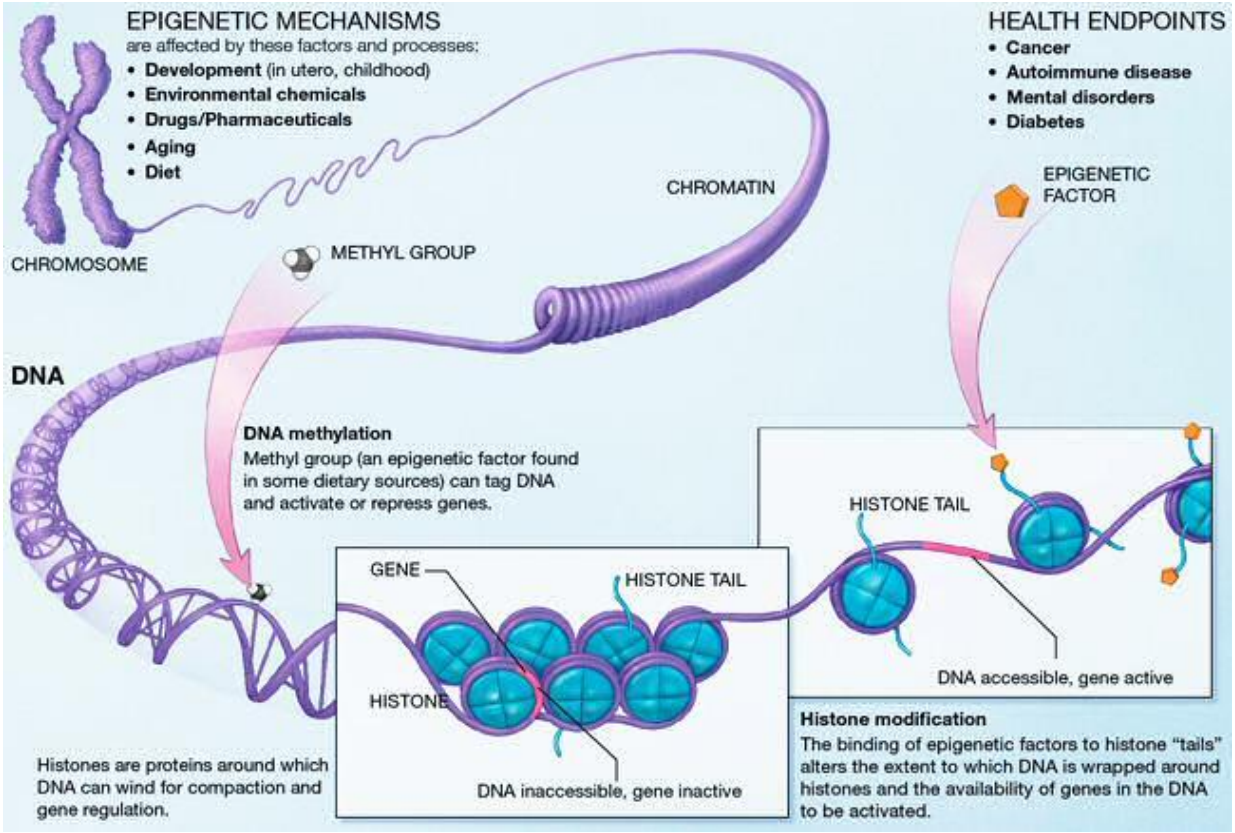
Epigenetics / Epigenomics

2022 Jan 19

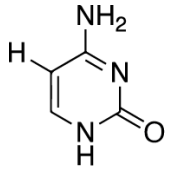
Steve Rozen

Epigenetics/epigenomcis overview will be on the final

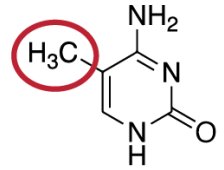
DukeNUS Medical School



By National Institutes of Health - http://commonfund.nih.gov/epigenomics/figure.aspx, Public Domain, https://commons.wikimedia.org/w/index.php?curid=9789221



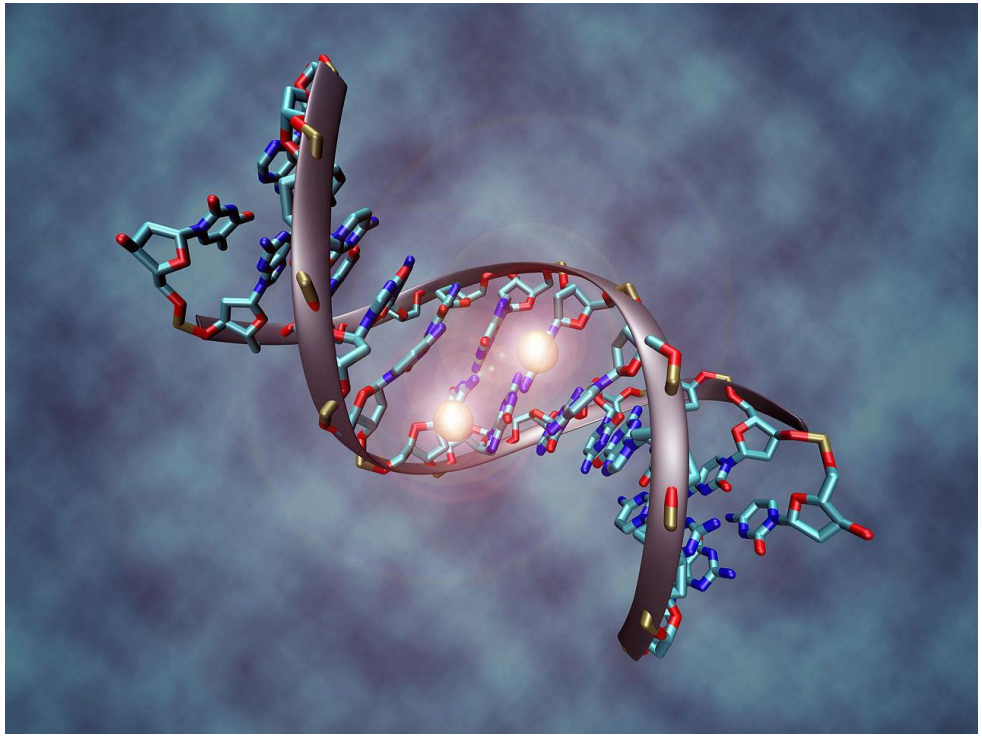
Cytosine



methylated Cytosine

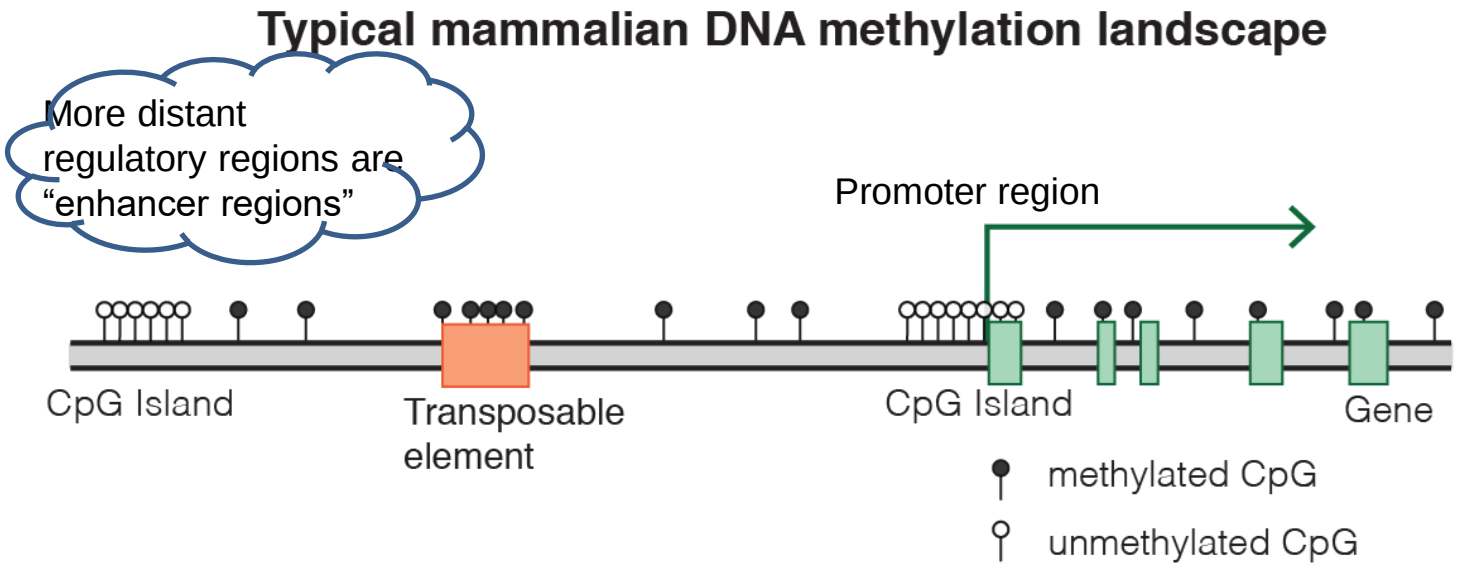
By Mariuswalter - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=54318073

DNA methylation

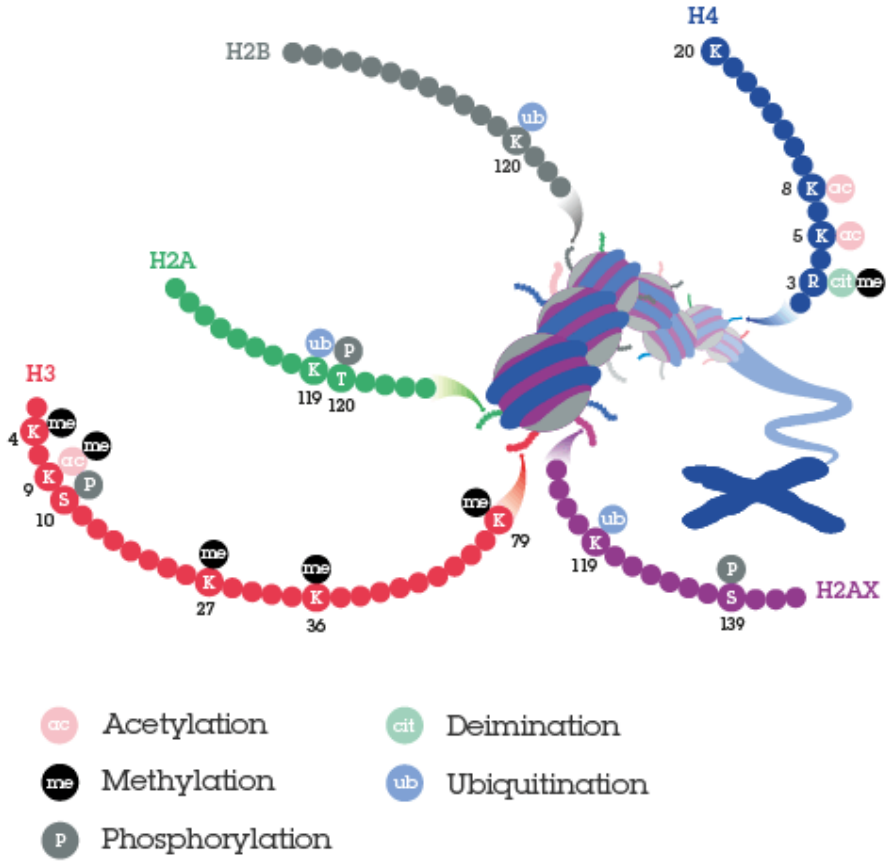


By Christoph Bock, Max Planck Institute for Informatics - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=17066877

DNA methylation



Histone modifications / histone code



Go to:

https://en.wikipedia.org/wiki/Histone_code

“Histone code”

Type of modification	Histone							
	H3K4	H3K9	H3K14	H3K27	H3K79	H3K122	H4K20	H2BK5
mono-methylation	activation[7]	activation[8]		activation[8]	activation[8][9]		activation[8]	activation[8]
di-methylation	repression[10]	repression[3]		repression[3]	activation[9]			
tri-methylation	activation[11]	repression[8]		repression[8]	activation,[9] repression[8]			repression[3]
acetylation		activation[11]	activation[11]	activation[12]		activation[13]		

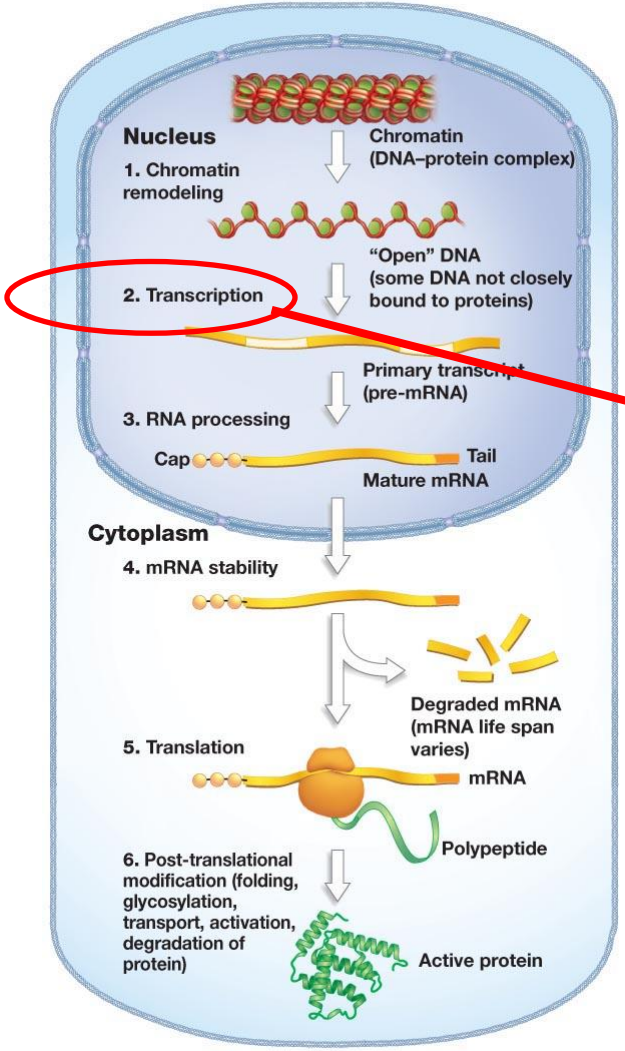
https://en.wikipedia.org/wiki/Histone_code

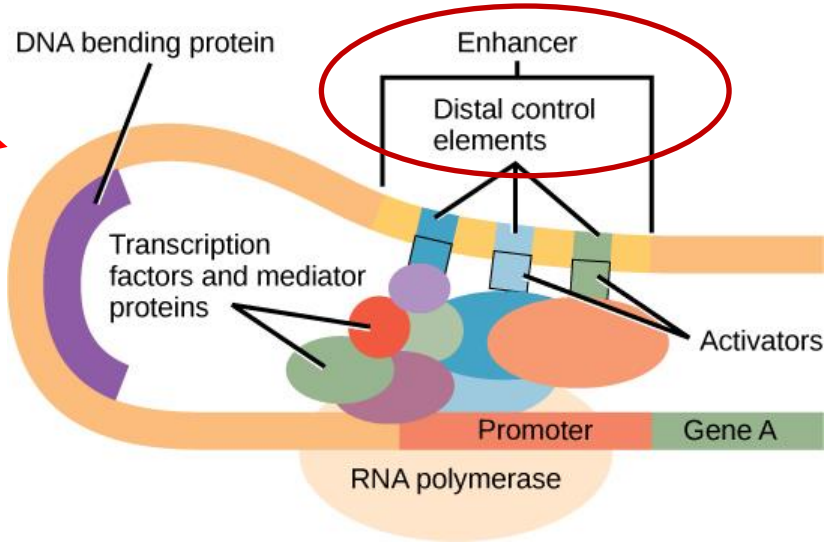
“Histone code”

- H3K4me1 - primed enhancers
- H3K4me3 is enriched in transcriptionally active promoters.[14]
- H3K9me2 -repression
- H3K9me3 is found in constitutively repressed genes.
- H3K27me3 is found in facultatively repressed genes.[8]
- H3K36me
- H3K36me2
- H3K36me3 is found in actively transcribed gene bodies.
- H3K79me2
- H3K9ac is found in actively transcribed promoters.
- H3K14ac is found in actively transcribed promoters.
- H3K23ac
- H3K27ac distinguishes active enhancers from poised enhancers.
- H3K36ac
- H3K56ac is a proxy for de novo histone assembly.[15]
- H3K122ac is enriched in poised promoters and also found in a different type of putative enhancer that lacks H3K27ac.

https://en.wikipedia.org/wiki/Histone_code

Interpreting the genetic program





https://courses.lumenlearning.com/wmopen-biology1/chapter/eukaryotic-gene-regulation/

nature Vol 459|7 May 2009|doi:10.1038/nature07829

Figure 1 | Chromatin modifications at promoters are generally cell-type-invariant whereas those at enhancers are cell-type-specific. We used ChIP-chip to map histone modifications (H3K4me1, H3K4me3 and H3K27ac) in the ENCODE regions in five cell types (HeLa, GM, K562, ES, dES). **a**, We performed *k*-means clustering on the chromatin modifications found ±5 kb from 414 promoters, and observe them to be generally invariant across cell types. **b**, As in **a**, but clustering on 1,423 non-redundant enhancers predicted on the basis of chromatin signatures, revealing the cell-type-specificity of enhancers. LogR is the log ratio of enrichment of each marker as determined by ChIP-chip. Promoters and predicted enhancers are located at the centre of 10-kb windows as indicated by black triangles.

DukeNUS Medical School

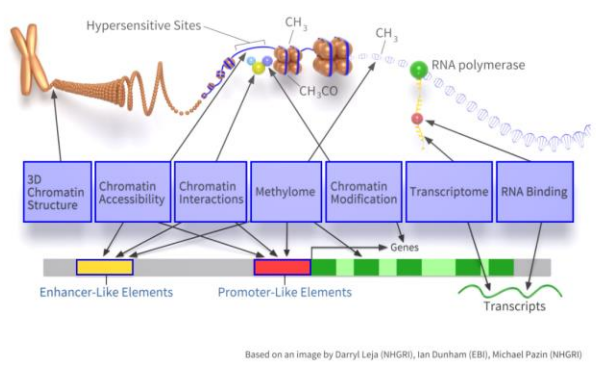
Histone modifications a
global cell-type-specific

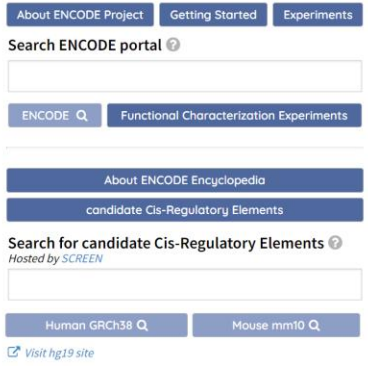
Nathaniel D. Heintzman[1,2]*, Gary C. Hon[1,3]*, R. Da
Lindsey F. Harp[1], Zhen Ye[1], Leonard K. Lee[1], Rhona
Jessica E. Antosiewicz-Bourget[7], Hui Liu[8], Xinmin Z
James A. Thomson[7,10], Gregory E. Crawford[11], Man

http://compbio.mit.edu/publications/33_Heintzman_Nature_09.pdf

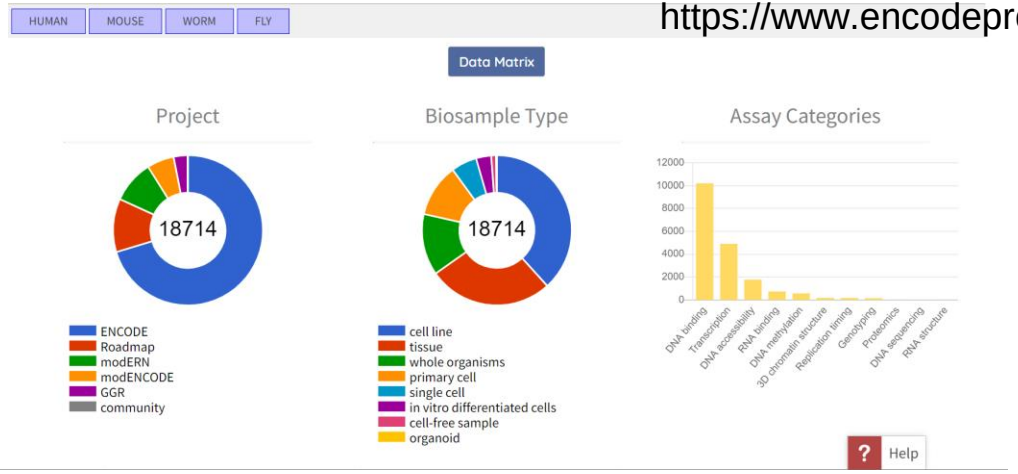
ENCODE: Encyclopedia of DNA Elements

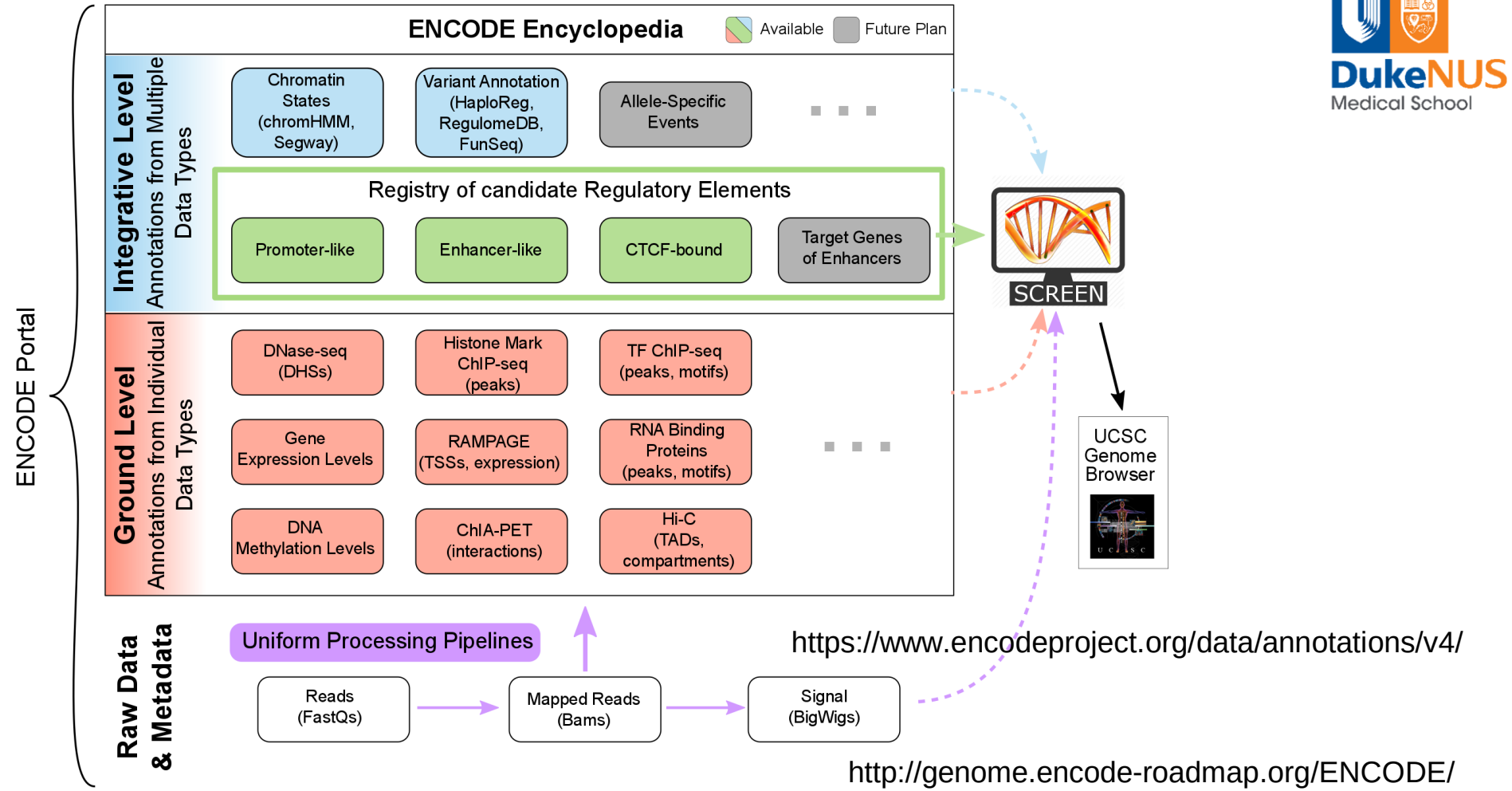
https://www.encodeproject.org/





https://www.encodeproject.org/data/annotations/v4/




ENCODE Portal
ENCODE Encyclopedia
Available
Future Plan
Integrative Level
Annotations from Multiple Data Types
Chromatin States (chromHMM, Segway)
Variant Annotation (HaploReg, RegulomeDB, FunSeq)
Allele-Specific Events
Registry of candidate Regulatory Elements
Promoter-like
Enhancer-like
CTCF-bound
Target Genes of Enhancers
Ground Level
Annotations from Individual Data Types
DNase-seq (DHSs)
Histone Mark ChIP-seq (peaks)
TF ChIP-seq (peaks, motifs)
Gene Expression Levels
RAMPAGE (TSSs, expression)
RNA Binding Proteins (peaks, motifs)
DNA Methylation Levels
ChIA-PET (interactions)
Hi-C (TADs, compartments)
SCREEN
UCSC Genome Browser
Raw Data & Metadata
Uniform Processing Pipelines
Reads (FastQs)
Mapped Reads (Bams)
Signal (BigWigs)
https://www.encodeproject.org/data/annotations/v4/
http://genome.encode-roadmap.org/ENCODE/

ARTICLE

OPEN

doi:10.1038/nature14248

Integrative analysis of 111 reference human epigenomes

Roadmap Epigenomics Consortium†, Anshul Kundaje[1,2,3]*, Wouter Meuleman[1,2]*, Jason Ernst[1,2,4]*, Misha Bilenky[5]*, Angela Yen[1,2], Alireza Heravi-Moussavi[5], Pouya Kheradpour[1,2], Zhizhuo Zhang[1,2], Jianrong Wang[1,2], Michael J. Ziller[2,6], Viren Amin[7], John W. Whitaker[8], Matthew D. Schultz[9], Lucas D. Ward[1,2], Abhishek Sarkar[1,2], Gerald Quon[1,2], Richard S. Sandstrom[10], Matthew L. Eaton[1,2], Yi-Chieh Wu[1,2], Andreas R. Pfenning[1,2], Xinchen Wang[1,2,11], Melina Claussnitzer[1,2], Yaping Liu[1,2], Cristian Coarfa[7], R. Alan Harris[7], Noam Shoresh[2], Charles B. Epstein[2], Elizabeta Gjoneska[2,12], Danny Leung[8,13], Wei Xie[8,13], R. David Hawkins[8,13], Ryan Lister[9], Chibo Hong[14], Philippe Gascard[15], Andrew J. Mungall[5], Richard Moore[5], Eric Chuah[5], Angela Tam[5], Theresa K. Canfield[10], R. Scott Hansen[16], Rajinder Kaul[16], Peter J. Sabo[10], Mukul S. Bansal[1,2,17], Annaick Carles[18], Jesse R. Dixon[8,13], Kai-How Farh[2], Soheil Feizi[1,2], Rosa Karlic[19], Ah-Ram Kim[1,2], Ashwinikumar Kulkarni[20], Daofeng Li[21], Rebecca Lowdon[21], GiNell Elliott[21], Tim R. Mercer[22], Shane J. Neph[10], Vitor Onuchic[7], Paz Polak[2,23], Nisha Rajagopal[8,13], Pradipta Ray[20], Richard C. Sallari[1,2], Kyle T. Siebenthall[10], Nicholas A. Sinnott-Armstrong[1,2], Michael Stevens[21,42], Robert E. Thurman[10], Jie Wu[24,25], Bo Zhang[21], Xin Zhou[21], Arthur E. Beaudet[26], Laurie A. Boyer[11], Philip L. De Jager[2,23,27], Peggy J. Farnham[28], Susan J. Fisher[29], David Haussler[30], Steven J. M. Jones[5,31,32], Wei Li[33], Marco A. Marra[5,32], Michael T. McManus[34], Shamil Sunyaev[2,23,27], James A. Thomson[35,41], Thea D. Tlsty[15], Li-Huei Tsai[2,12], Wei Wang[8], Robert A. Waterland[36], Michael Q. Zhang[20,37], Lisa H. Chadwick[38], Bradley E. Bernstein[2,39,40]§, Joseph F. Costello[14]§, Joseph R. Ecker[9]§, Martin Hirst[5,18]§, Alexander Meissner[2,6]§, Aleksandar Milosavljevic[7]§, Bing Ren[8,13]§, John A. Stamatoyannopoulos[10]§, Ting Wang[21]§ & Manolis Kellis[1,2]§

The reference human genome sequence set the stage for studies of genetic variation and its association with human disease, but epigenomic studies lack a similar reference. To address this need, the NIH Roadmap Epigenomics Consortium generated the largest collection so far of human epigenomes for primary cells and tissues. Here we describe the integrative analysis of 111 reference human epigenomes generated as part of the programme, profiled for histone modification patterns, DNA accessibility, DNA methylation and RNA expression. We establish global maps of regulatory elements, define regulatory modules of coordinated activity, and their likely activators and repressors. We show that disease- and trait-associated genetic variants are enriched in tissue-specific epigenomic marks, revealing biologically relevant cell types for diverse human traits, and providing a resource for interpreting the molecular basis of human disease. Our results demonstrate the central role of epigenomic information for understanding gene regulation, cellular differentiation and human disease.

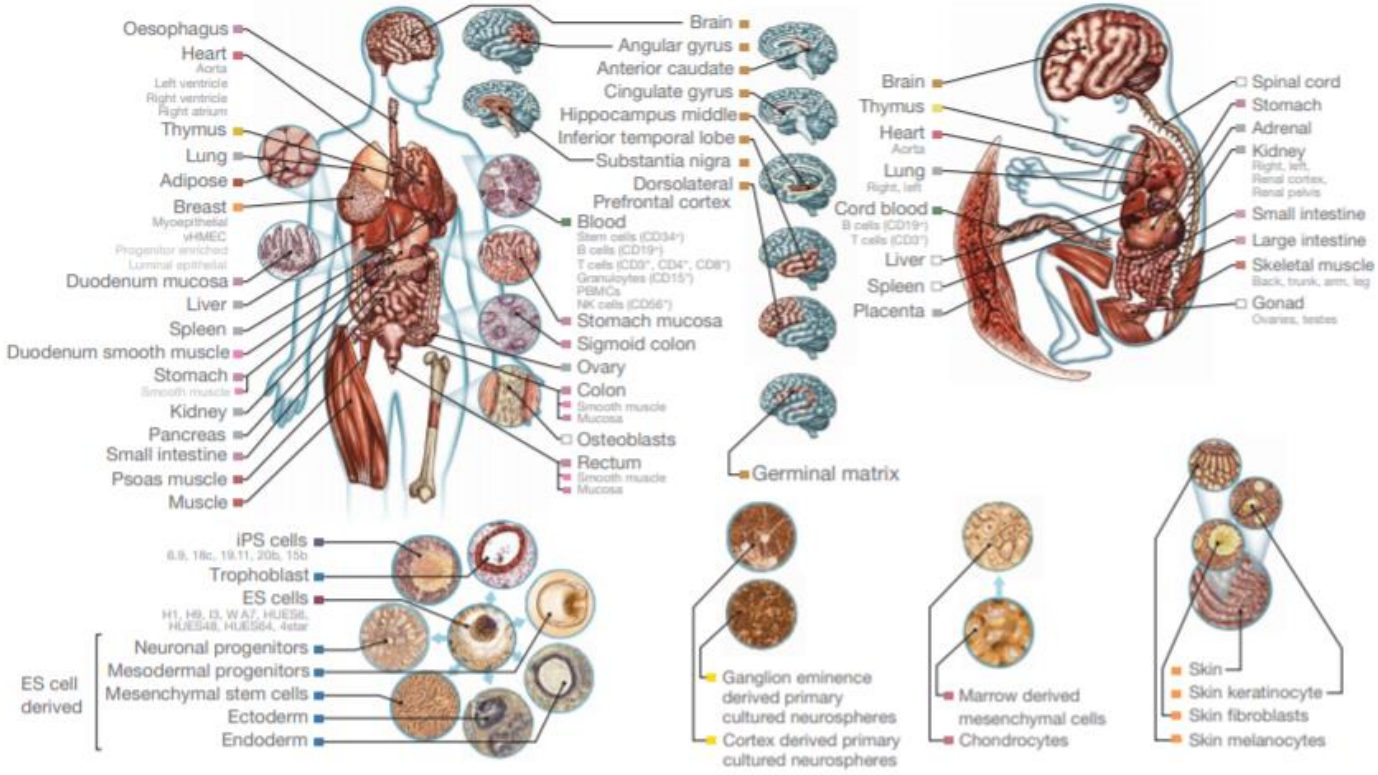


Figure 1 | Tissues and cell types profiled in the Roadmap Epigenomics Consortium. Primary tissues and cell types representative of all major lineages in the human body were profiled, including multiple brain, heart, muscle, gastrointestinal tract, adipose, skin and reproductive samples, as well as immune lineages, ES cells and iPS cells, and differentiated lineages derived from ES cells. Box colours match groups shown in Fig. 2b. Epigenome identifiers (EIDs, Fig. 2c) for each sample are shown in Extended Data Fig. 1.

https://fantom.gsc.riken.jp/

DukeNUS Medical School

FANTOM

FANTOM is an international research consortium established by Dr. Hayashizaki and his colleagues in 2000 to assign functional annotations to the full-length cDNAs that were collected during the Mouse Encyclopedia Project at RIKEN. FANTOM has since developed and expanded over time to encompass the fields of transcriptome analysis. The object of the project is moving steadily up the layers in the system of life, progressing thus from an understanding of the 'elements' - the transcripts - to an understanding of the 'system' - the transcriptional regulatory network, in other words the 'system' of an individual life form.

FANTOM is now in the 6th edition of the project. Project page of each edition is available below:

- FANTOM6 - Functional analysis of non-coding RNAs
- FANTOM5 - Atlases of mammalian promoters, enhancers, lncRNAs and miRNAs
- FANTOM4 - Understanding the transcriptional regulatory network
- FANTOM3 - Transcriptional landscape of mammalian genome
- FANTOM2 - Functional annotation of ~60,000 mouse full-length cDNA collection
- FANTOM1 - Initial functional annotation of ~20,000 mouse cDNA collection

Mouse over the image below for information on FANTOM history and publications.

FANTOM3: Functional Annotation Of Mouse - 3

FANTOM3

- HOME
- Download
- FANTOM3 papers
- History
- Contact
- FAQ
- Photos

RELATED WORKS

- FANTOM portal
- Databases
- References

The Fantom (Functional Annotation of the mouse) aims at providing the ultimate characterization of the mouse transcriptome. O
biological problems and bioinformatics. After the development of original technologies (such as full-length cDNA libraries, CAGE
The combination of original RIKEN full-length cDNAs, CAGE tags and GSC ditags allowed providing the most extensive descript
issue of Science, another milestone one in Nature Genetics, the Genome N
All data and resources used in this project are available to the community fr

Announcements

- [Notice] FANTOM servers, including the cDNA viewer and the CAGE
- FANTOM3 Satellite papers are published in the Genome Network/FAN
- Another milestone paper is published in Nature Genetics. Its supp
- Photo gallery is updated. [2006.01.27]
- Publishing of two milestone papers (Main and extended press release

Databases

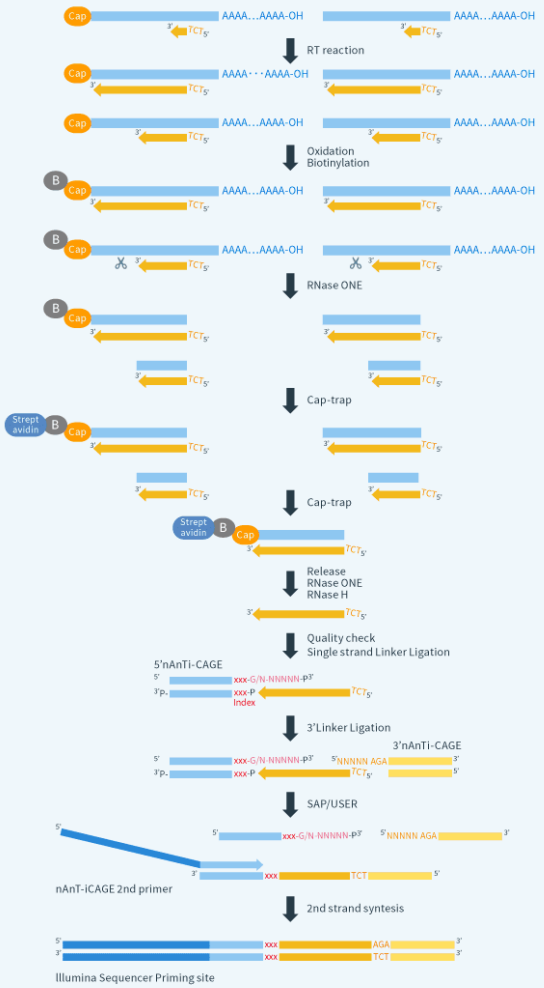
- cDNA Annotation (Annotation strategy)
  - RIKEN cDNA Annotation Viewer
  - Public cDNA Viewer
- Sense/Antisense
  - SADB (Archive files in the Sense/Antise base)
- CAGE (Cap-Analysis Gene Expression)
  - CAGE Basic Viewer (CAGE primary Database)
    [mouse | human (current) | human (Previous ver. of September 2005)]
  - CAGE Analysis Viewer (Promoter Database)
    [mouse | human (current) | human (Previous ver. of September 2005)]
  - CAGE Tree Viewer (Expression clusters)
    [mouse | human]
- Genomic Elements Viewer
  [mouse | human]

In general, hard to find the 5' ends of transcripts, but this is important to find genes with alternative 5' exons, which start transcript variants with different promoter regions and different regulation

"CAGE" deals with this

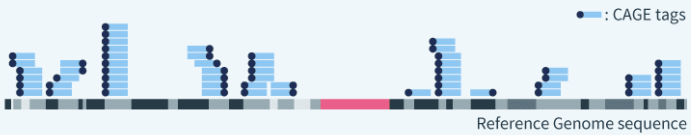
DukeNUS Medical School

CAGE library preparation

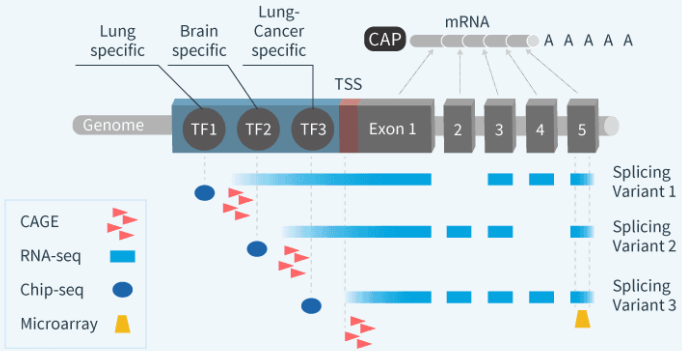


Sequencing, Visualization & Analysis of data

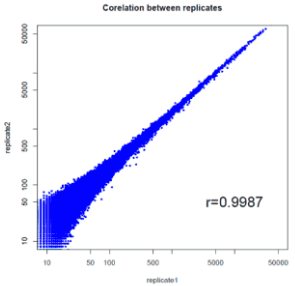
Expression Profiling



Comparison among major gene expression analysis techniques



High reproducibility



https://dnaform.jp/

DukeNUS Medical School

Misuse of the term “epigenetics”

Meaningless advertising gimmicks

https://perma.cc/U2CG-896G

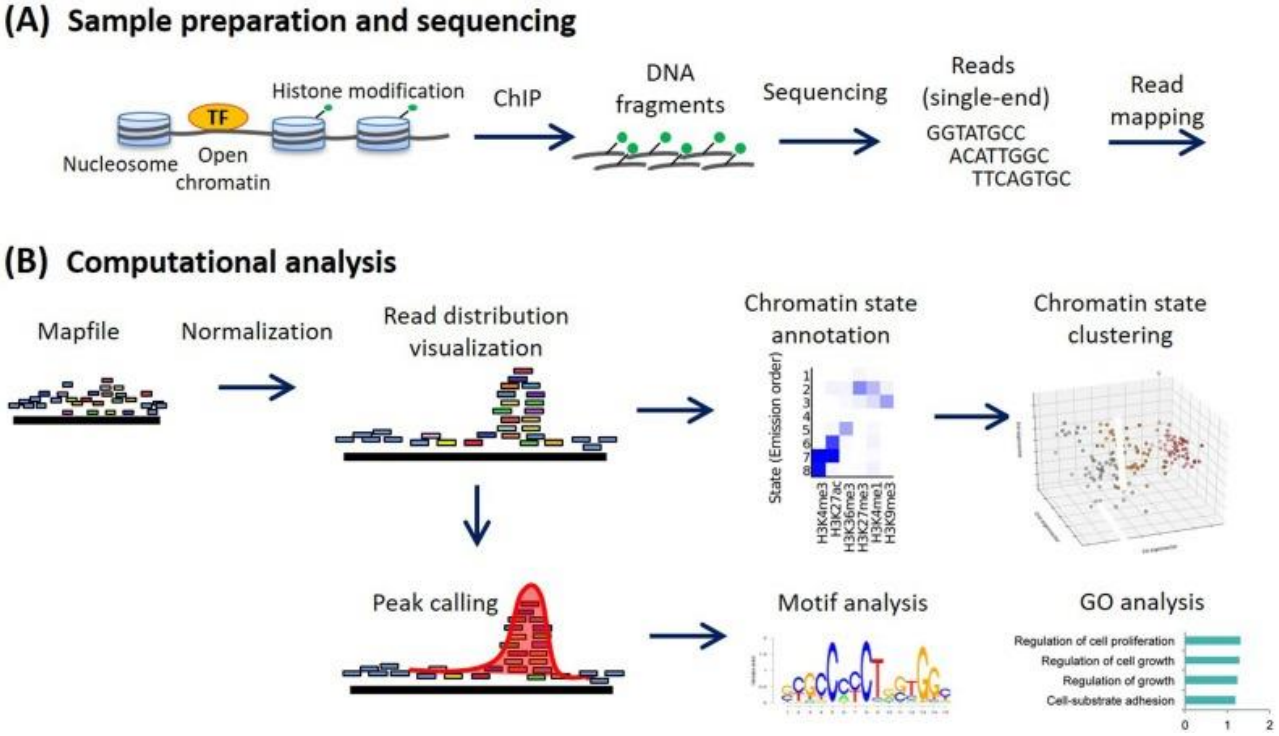
Obsession with transgenerational epigenetic inheritance: https://perma.cc/Z2AZ-RJXZ

More nonsense: https://perma.cc/HPQ8-LBSL

Bottom line: Be more specific, use terms like “chromatin state” (or, more specifically, “DNA methylation”, “histone marks”, “binding of specific transcription factors or regulatory complexes”, “open versus closed chromatin”, etc.)

Methods for ChIP-seq analysis: A practical workflow and advanced applications Ryuichiro Nakato Toyonori Sakata

https://www.sciencedirect.com/science/article/pii/S1046202320300591



Bioinformatics for analyzing ChIP-seq

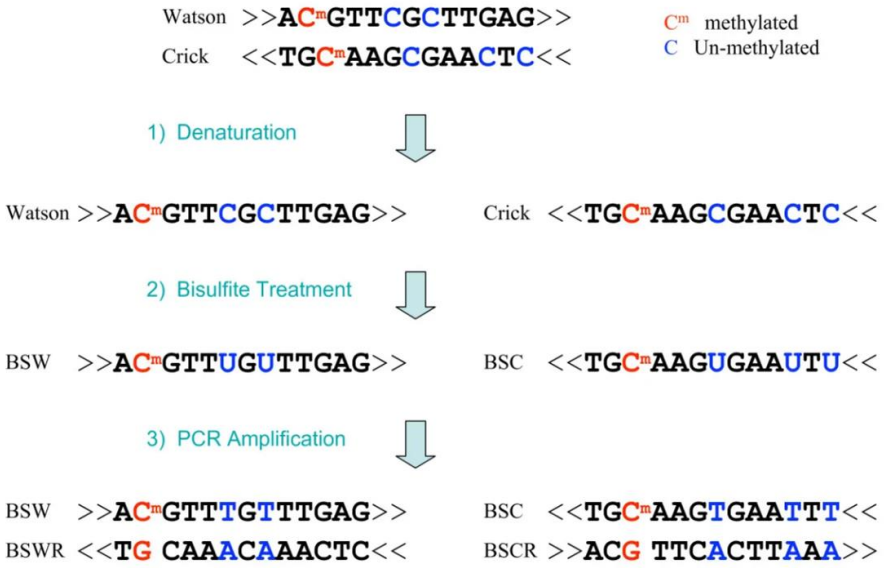
For those with more computational and less genomics background, see https://www.youtube.com/watch?v=nkWGmaYRues for an easy intro to ChiP-seq

There is a ChiP-seq bioinformatics pipeline / recipes here:

https://www.bioconductor.org/packages/devel/data/experiment/vignettes/systemPipeRdata/inst/doc/systemPipeChIPseq.html

(To review in class)

Bisulfite conversion for assaying (5 methy cytosine) DNA methylation



From Xi and Li, 2009, https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-10-232

Pipeline of bisulfite sequencing. 1) Denaturation: separating Watson and Crick strands; 2) Bisulfite treatment: converting un-methylated cytosines (blue) to uracils; methylated cytosines (red) remain unchanged; 3) PCR amplification of bisulfite-treated sequences resulting in four distinct strands: Bisulfite Watson (BSW), bisulfite Crick (BSC), reverse complement of BSW (BSWR), and reverse complement of BSC (BSCR).

Analyzing (5 methyl cytosine) methylation data

DukeNUS Medical School

Technologies

1. Methylation arrays (older and cheaper):

   https://www.bioconductor.org/packages/devel/workflows/vignettes/methylationArrayAnalysis/inst/doc/methylationArrayAnalysis.html

1. Bisulfite sequencing -- not that widely used (?); will not discuss in detail

2.1 Special aligner: http://www.bioinformatics.babraham.ac.uk/projects/bismark/ or https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-10-232

2.2 Downstream analysis: https://www.bioconductor.org/packages/devel/workflows/vignettes/methylationArrayAnalysis/inst/doc/methylationArrayAnalysis.html