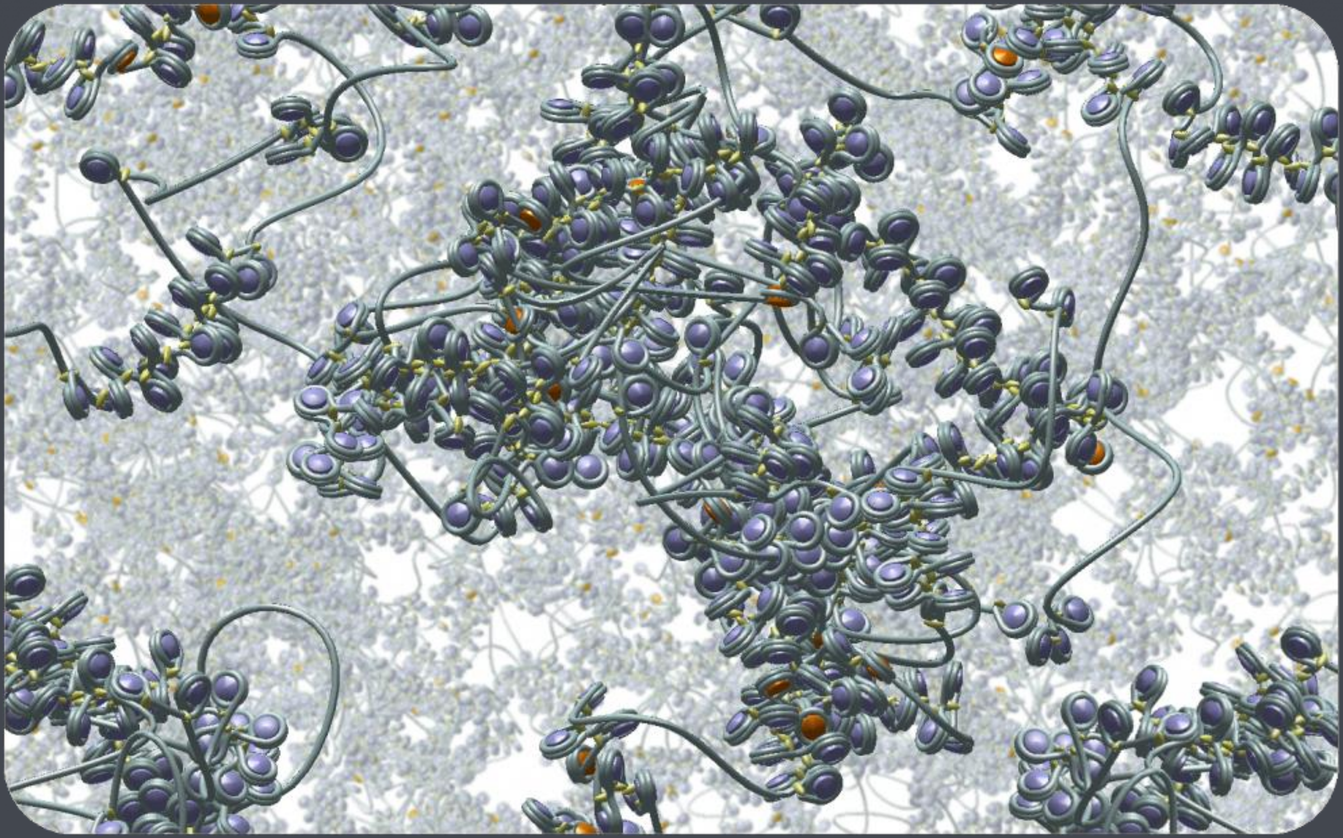
ANALYSIS OF EPIGENETICS AND CHROMATIN STATES IN NORMAL AND CANCER GENOMES

Valentina BOEVA

Institut Curie, Inserm U900, Mines ParisTech

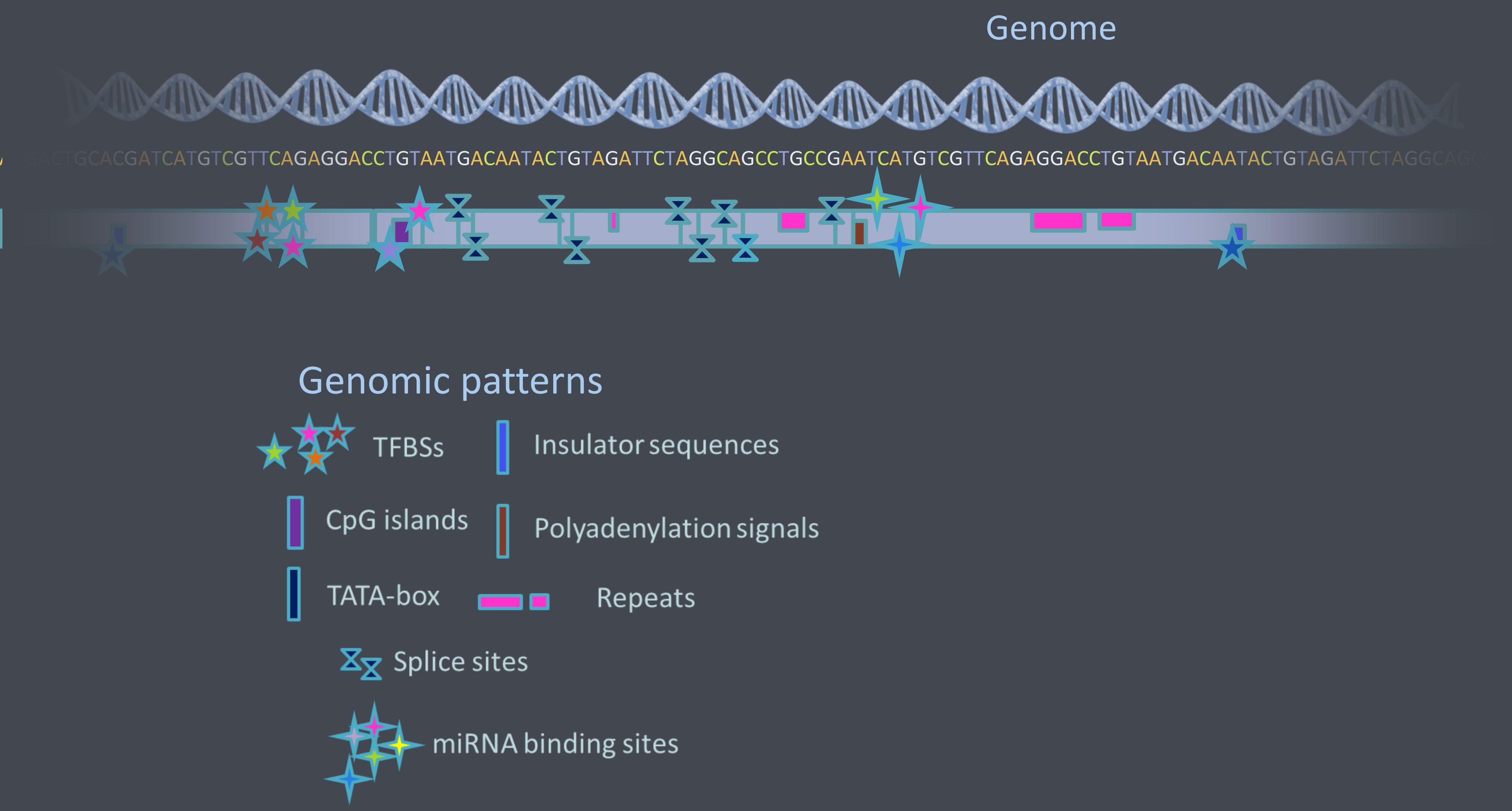
DNA is bound around histone proteins to form chromatin

1 Mb of chromatin

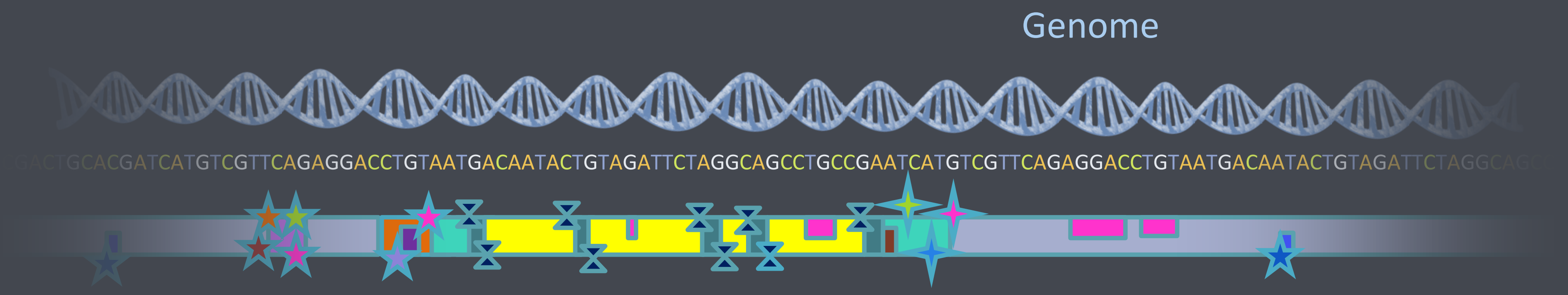


Diesinger and Heermann, *Biophys J.*, 2009

There are many various sequence patterns and motifs in the genome



Pattern and motifs (directly or indirectly) affect functional properties of genomic regions



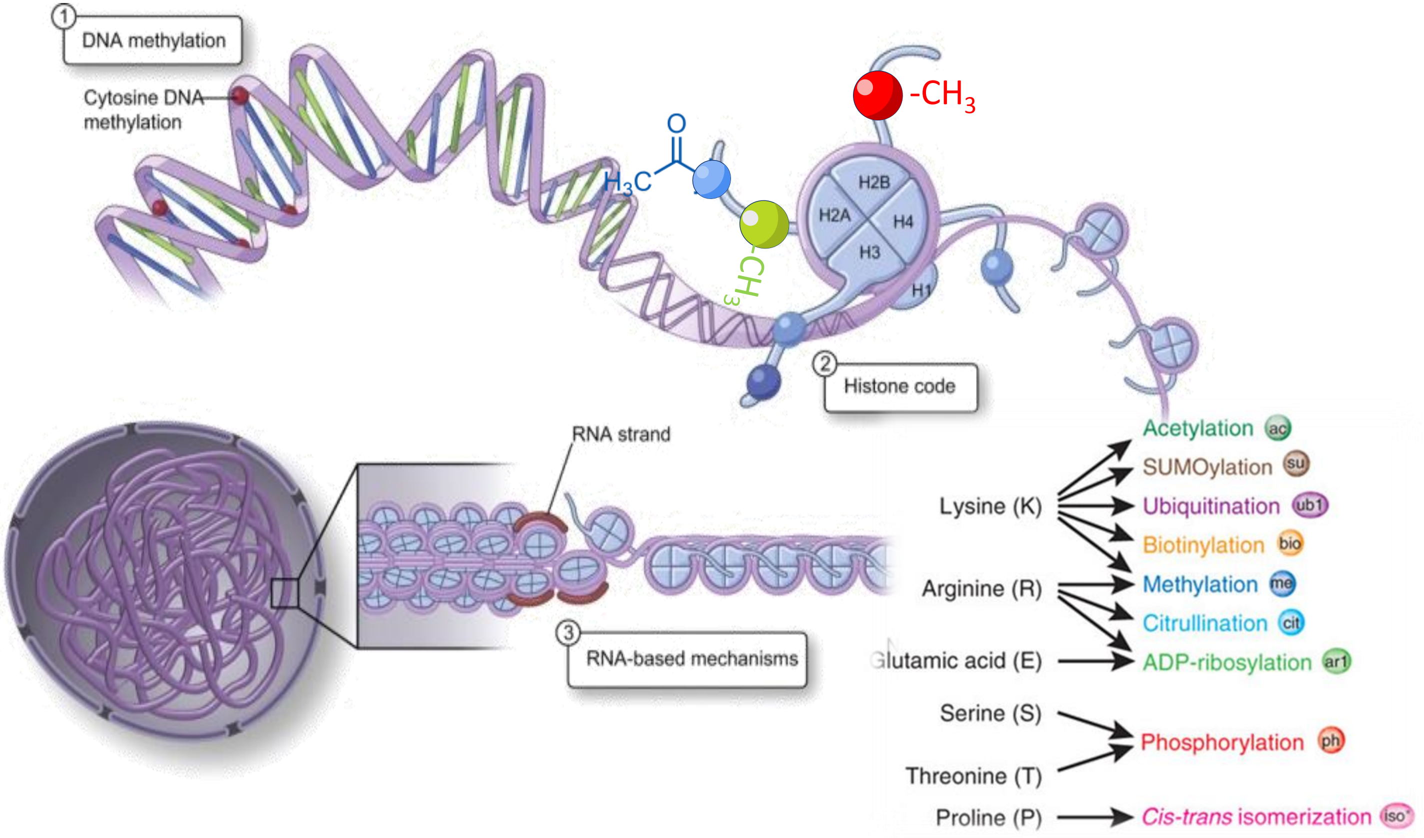
Genomic patterns

- ★ TFBSs
- ▮ Insulator sequences
- ▮ CpG islands
- ▮ Polyadenylation signals
- ▮ TATA-box
- ▬ ▪ Repeats
- ⧗ Splice sites
- ✦ miRNA binding sites

Genomic regions

- ⬜ Intergenic regions
- 🟪 Enhancer
- 🟧 Promoter
- 🟦 Exons
- 🟨 Introns
- 🟩 UTRs

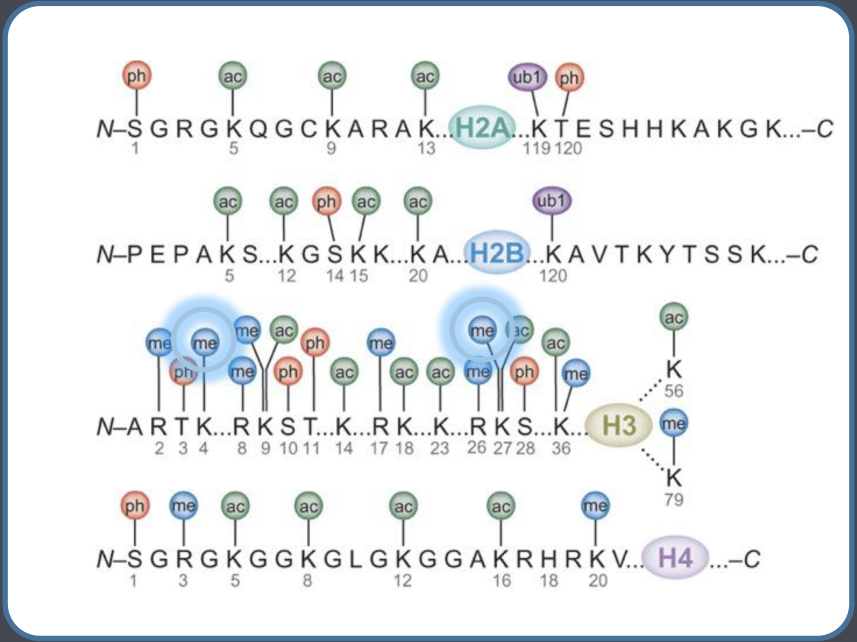
Epigenetic profiles = combination of CpG methylation of DNA and histone modifications

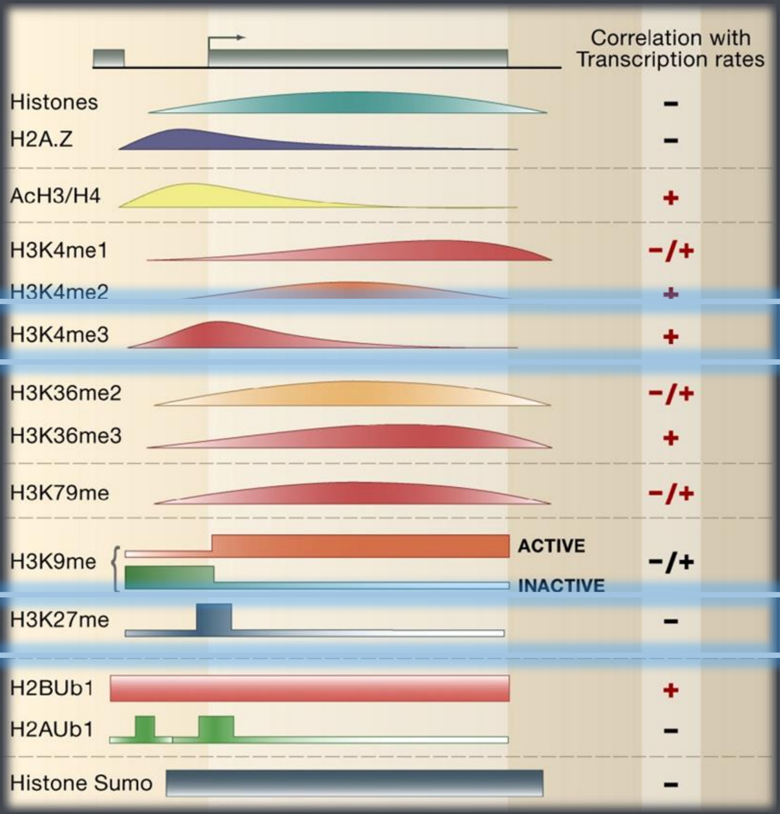


M. S. Yan et al, *J. Appl. Physiol.*, 2010

Histone modifications have direct effect on gene transcription

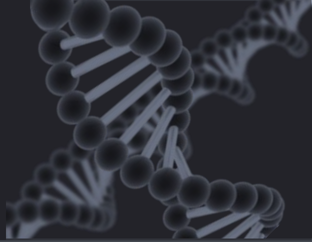
- Histone modifications

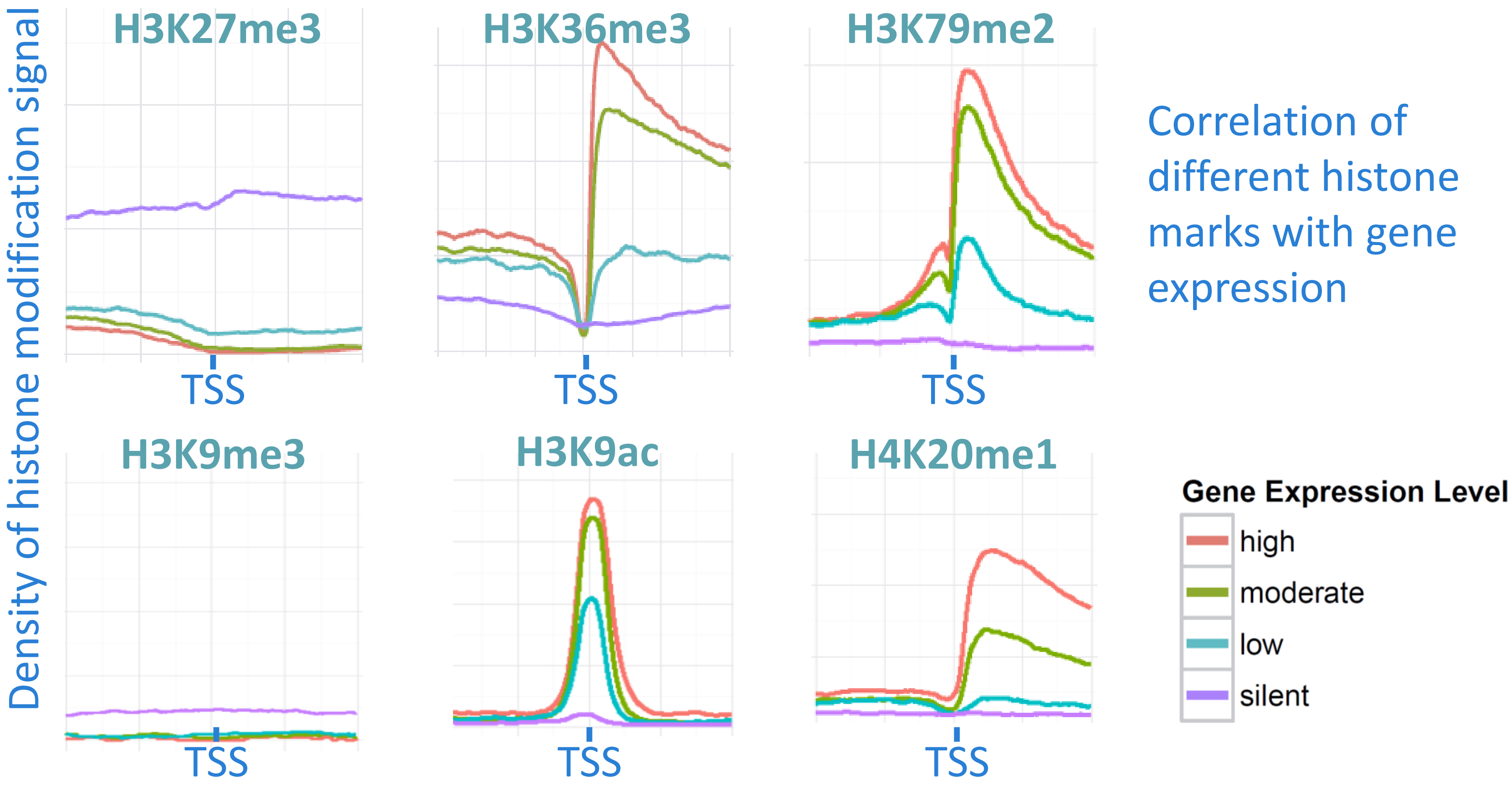




Bhaumik et al, *Nat Str & Mol Biol*, 2007
Li et al, *Cell,* 2007

Histone modifications have direct effect on gene transcription



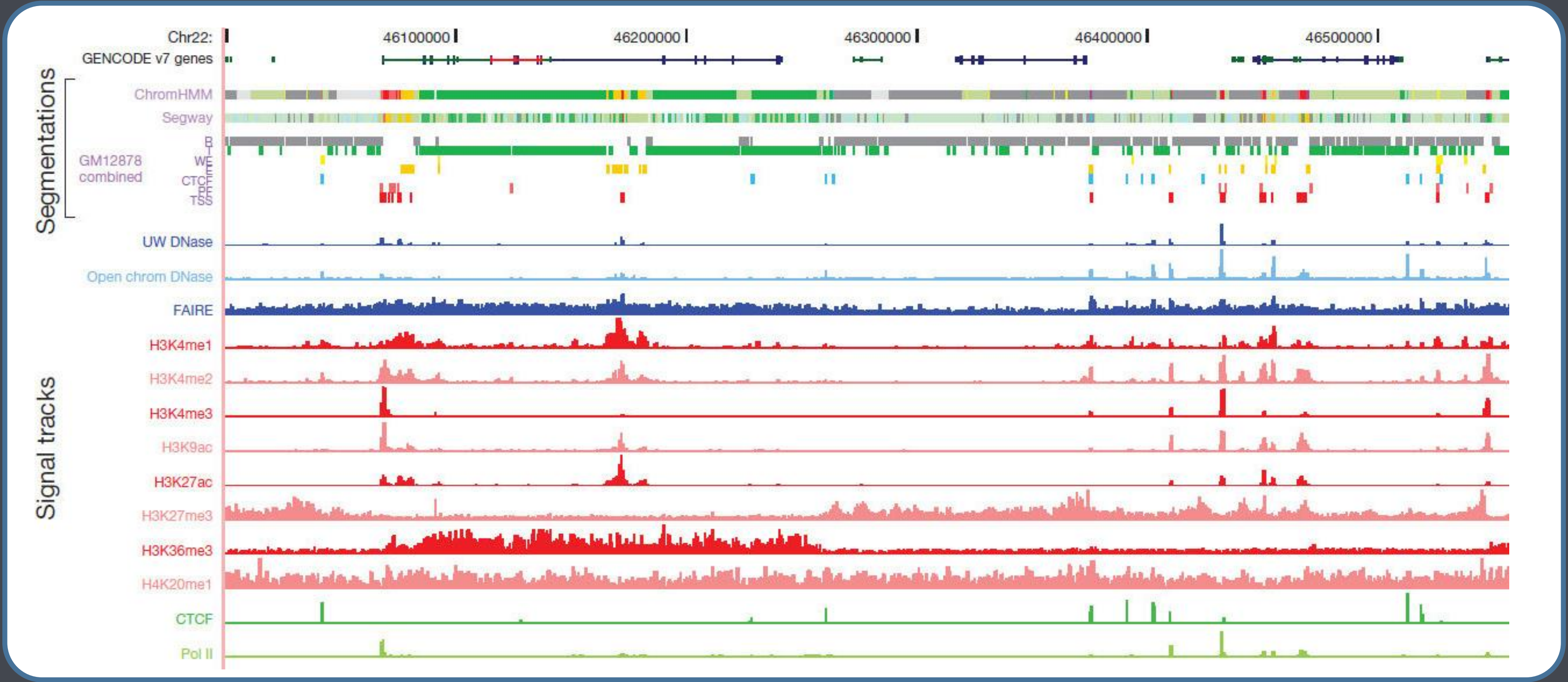


Correlation of different histone marks with gene expression

HeLa-S3 cell line

Histone modifications form groups and indicate distinct chromatin states

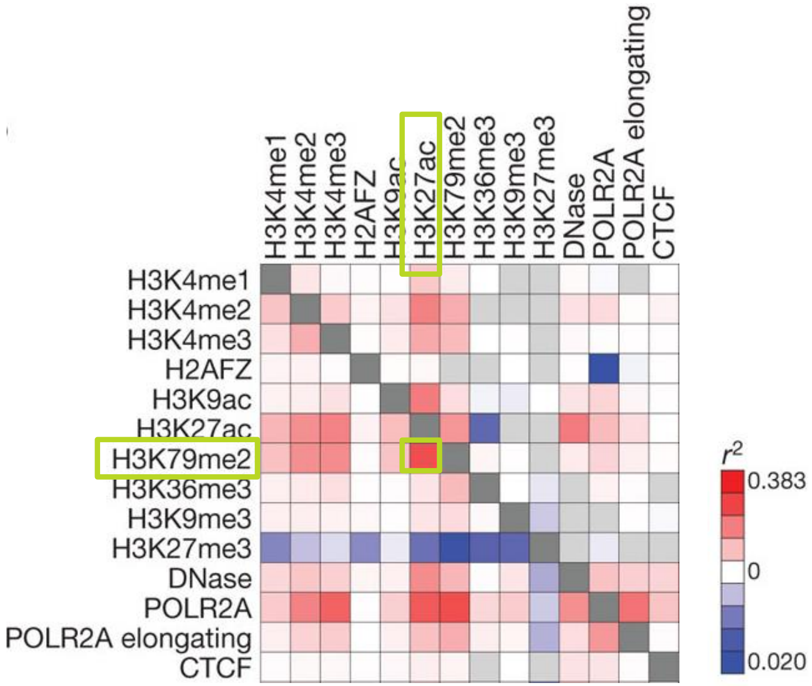
- Histone modifications, histone variants, binding sites (Pol II, CTCF, p300,…) ➡ chromatin states



ENCODE Project Consortium, *Nature,* 2012

Histone modifications form groups and indicate distinct chromatin states

- Histone modifications, histone variants, binding sites (Pol II, CTCF, p300,…) ➡ chromatin states

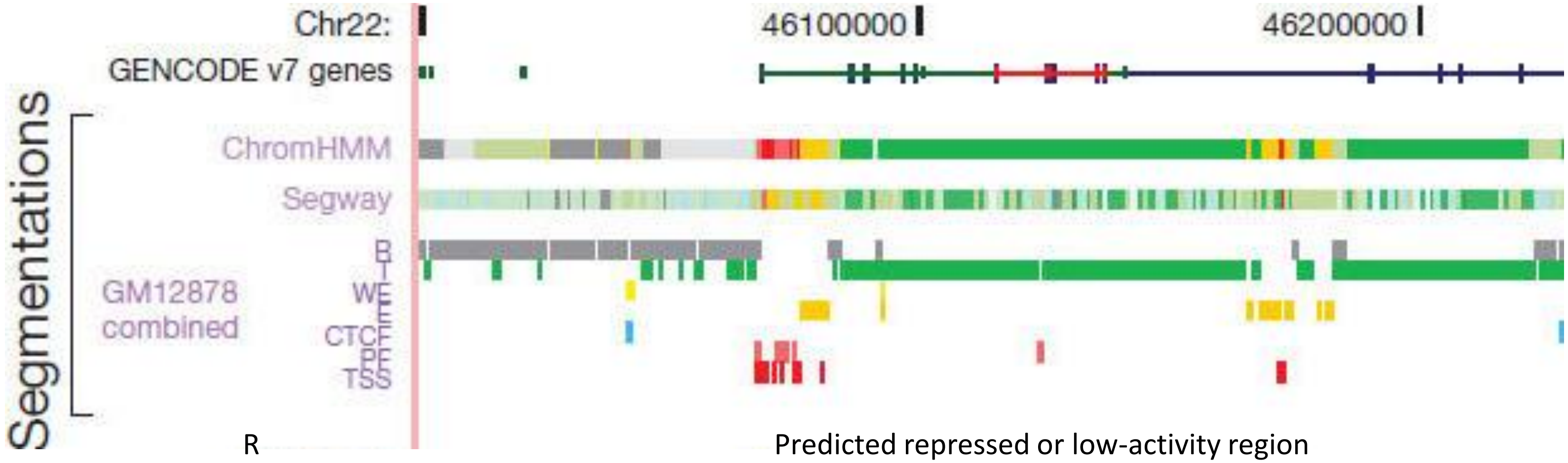


Pair-wise correlations of allele-specific signal within single genes (below the diagonal) or within individual ChromHMM segments across the whole genome for selected DNase-seq and histone modification ChIP-seq assays

ENCODE Project Consortium, *Nature*, 2012

Histone modifications form groups and indicate distinct chromatin states

- Histone modifications, histone variants, binding sites (Pol II, CTCF, p300,…) ➡ chromatin states



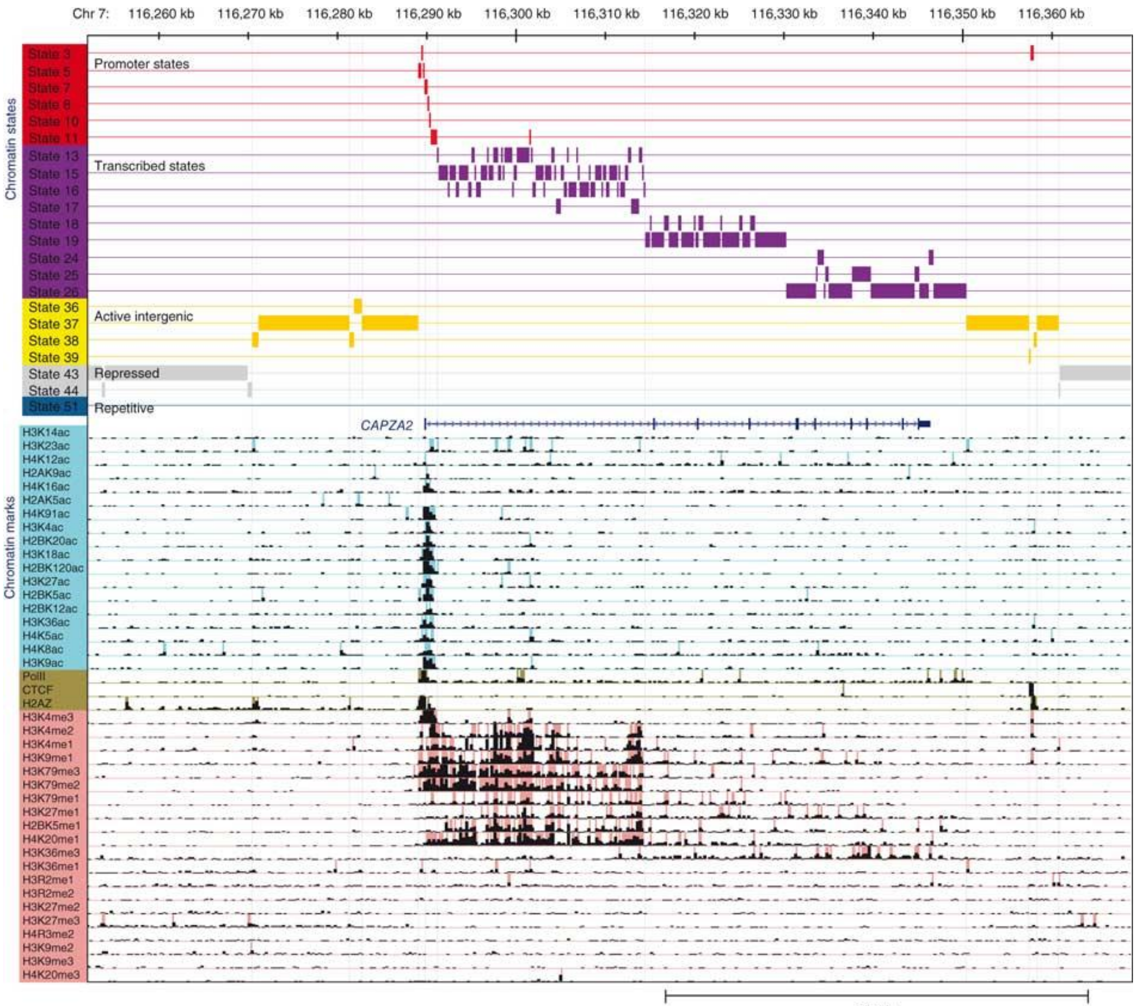
7 states

R	Predicted repressed or low-activity region
T	Predicted transcribed region
WE	Predicted weak enhancer or open chromatin *cis*-regulatory element
E	Predicted enhancer
CTCF	CTCF-enriched element
PF	Predicted promoter flanking region
TSS	Predicted promoter region including TSS

ENCODE Project Consortium, *Nature*, 2012

Histone modifications form groups and indicate distinct chromatin states

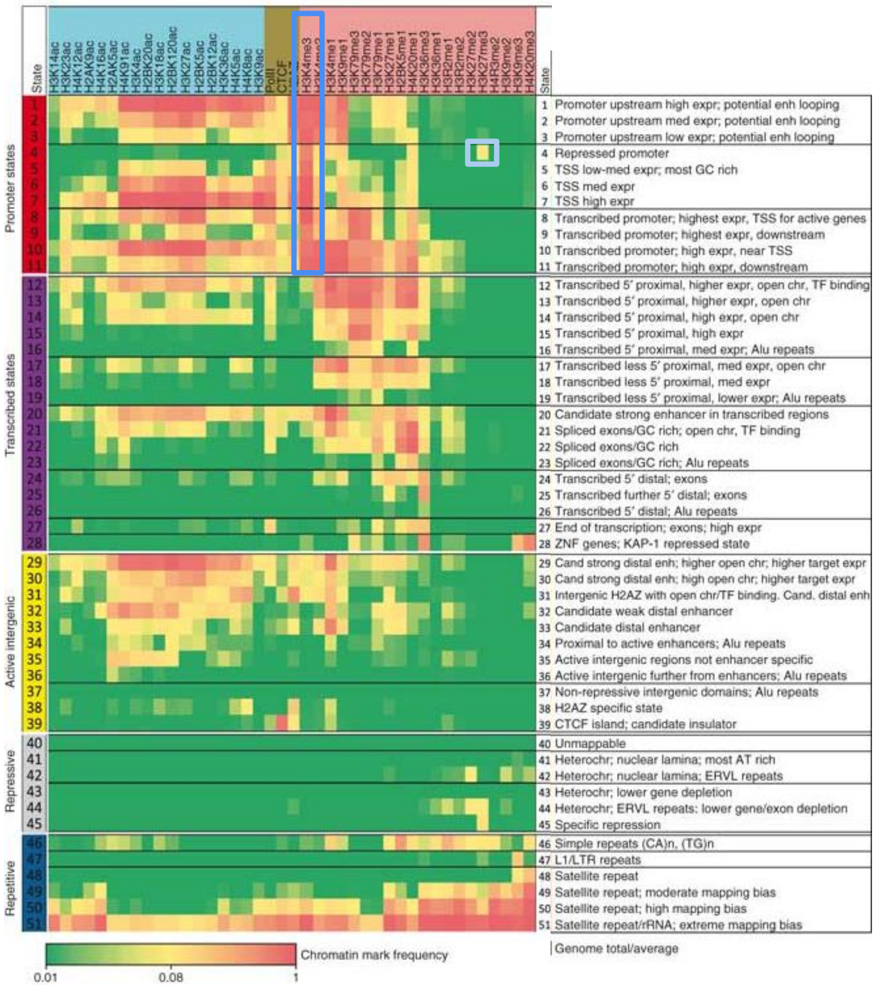
51 states



Input chromatin mark information and resulting chromatin state annotation for a 120-kb region of human chromosome 7 surrounding the *CAPZA2* gene

Ernst & Kellis, *Nature Biotechnology*, 2010

Histone modifications form groups and indicate distinct chromatin states

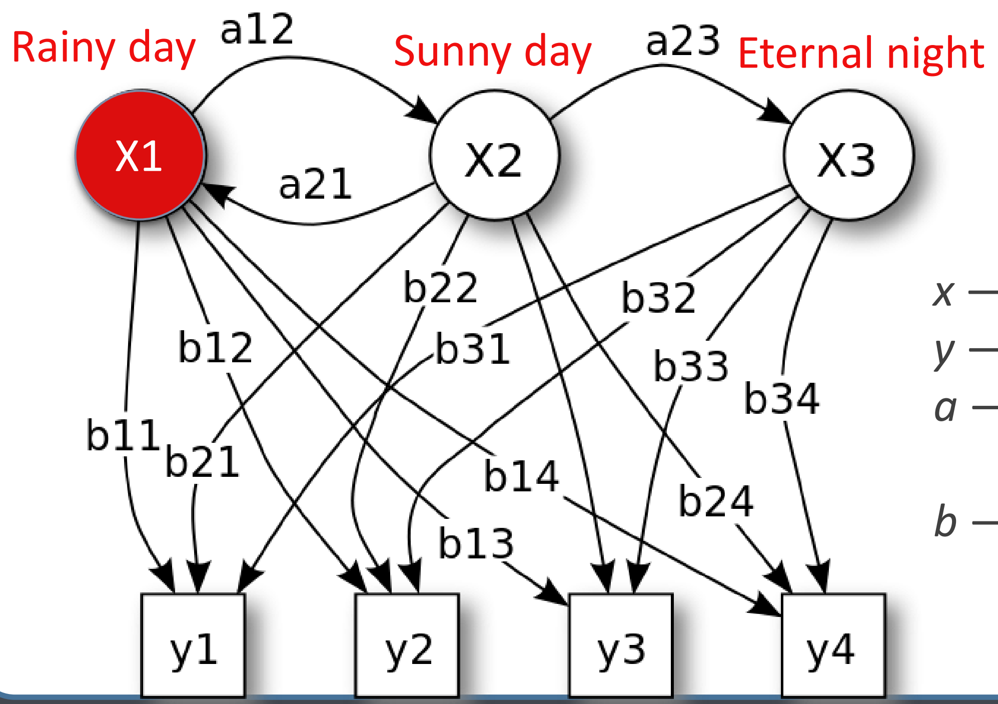


Chromatin mark combinations associated with each state and brief description of biological state function and interpretation

Ernst & Kellis, *Nature Biotechnology*, 2010

Hidden Markov Models allow annotating chromatin states

- Probabilistic parameters of a Hidden Markov Model (HMM)



x — states

y — possible observations

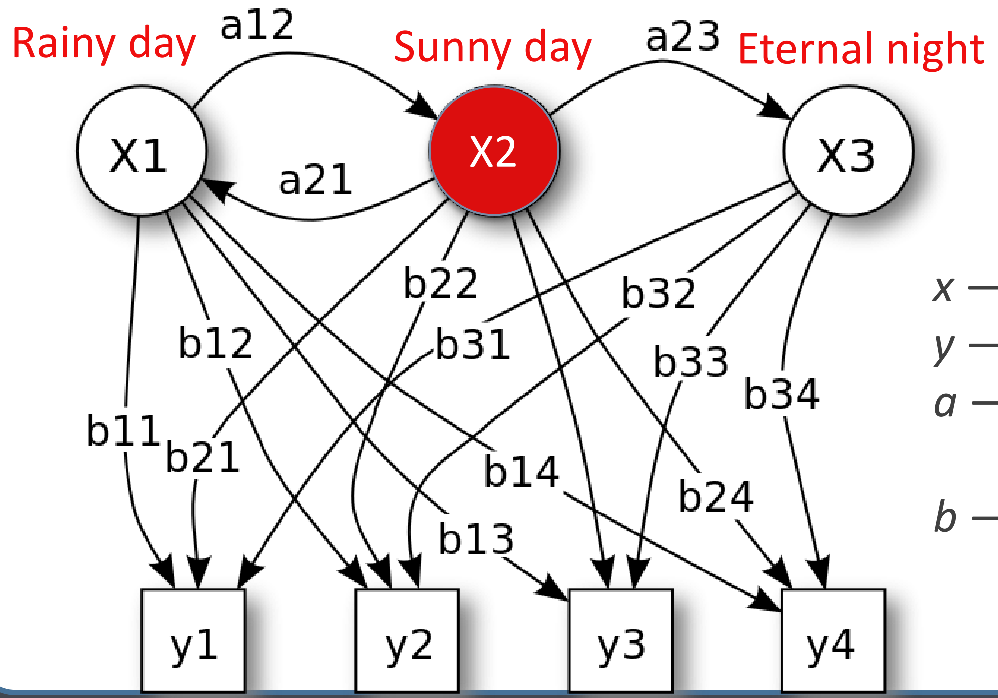
a — state transition probabilities

b — output probabilities

Wikipedia

Hidden Markov Models allow annotating chromatin states

- Probabilistic parameters of a Hidden Markov Model (HMM)

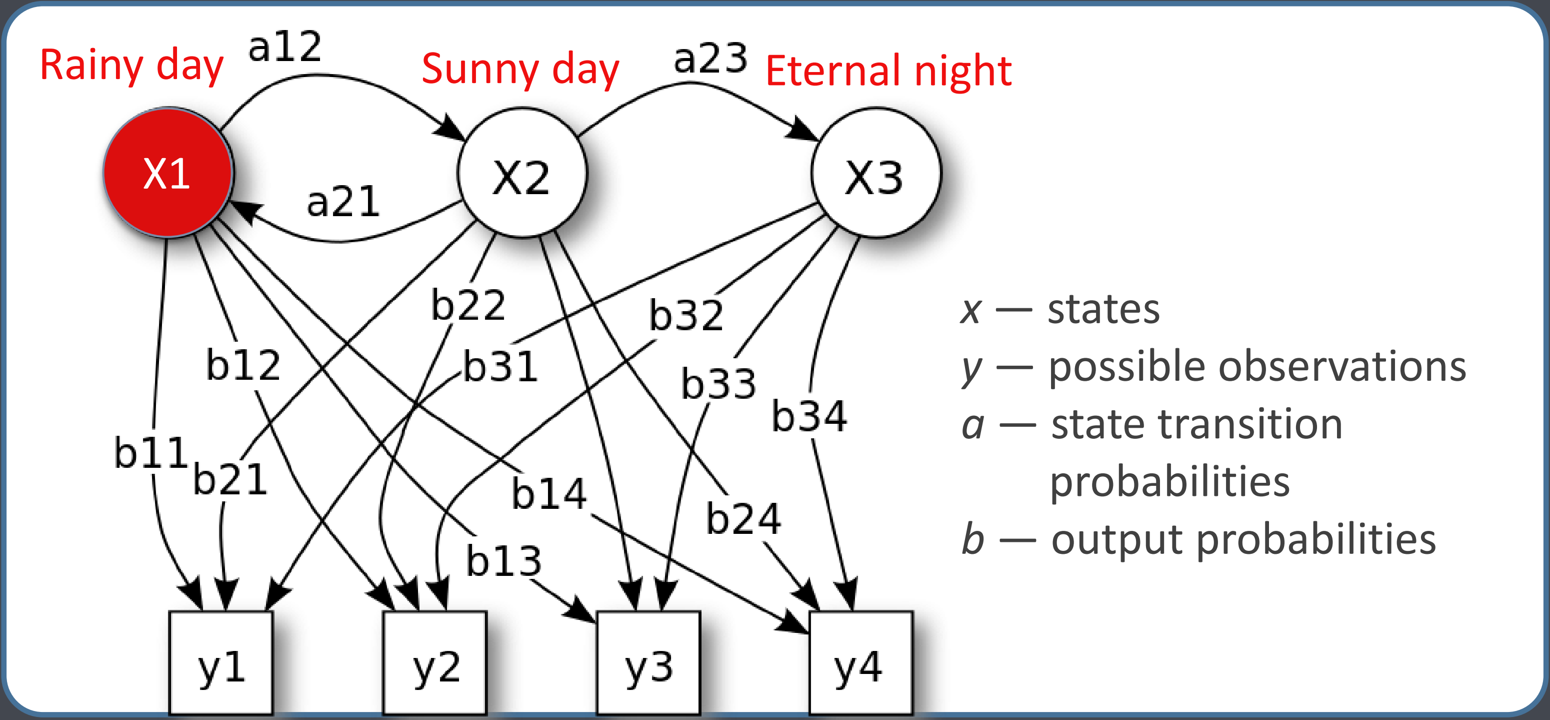


x — states
y — possible observations
a — state transition probabilities
b — output probabilities

Wikipedia

Hidden Markov Models allow annotating chromatin states

- Probabilistic parameters of a Hidden Markov Model (HMM)

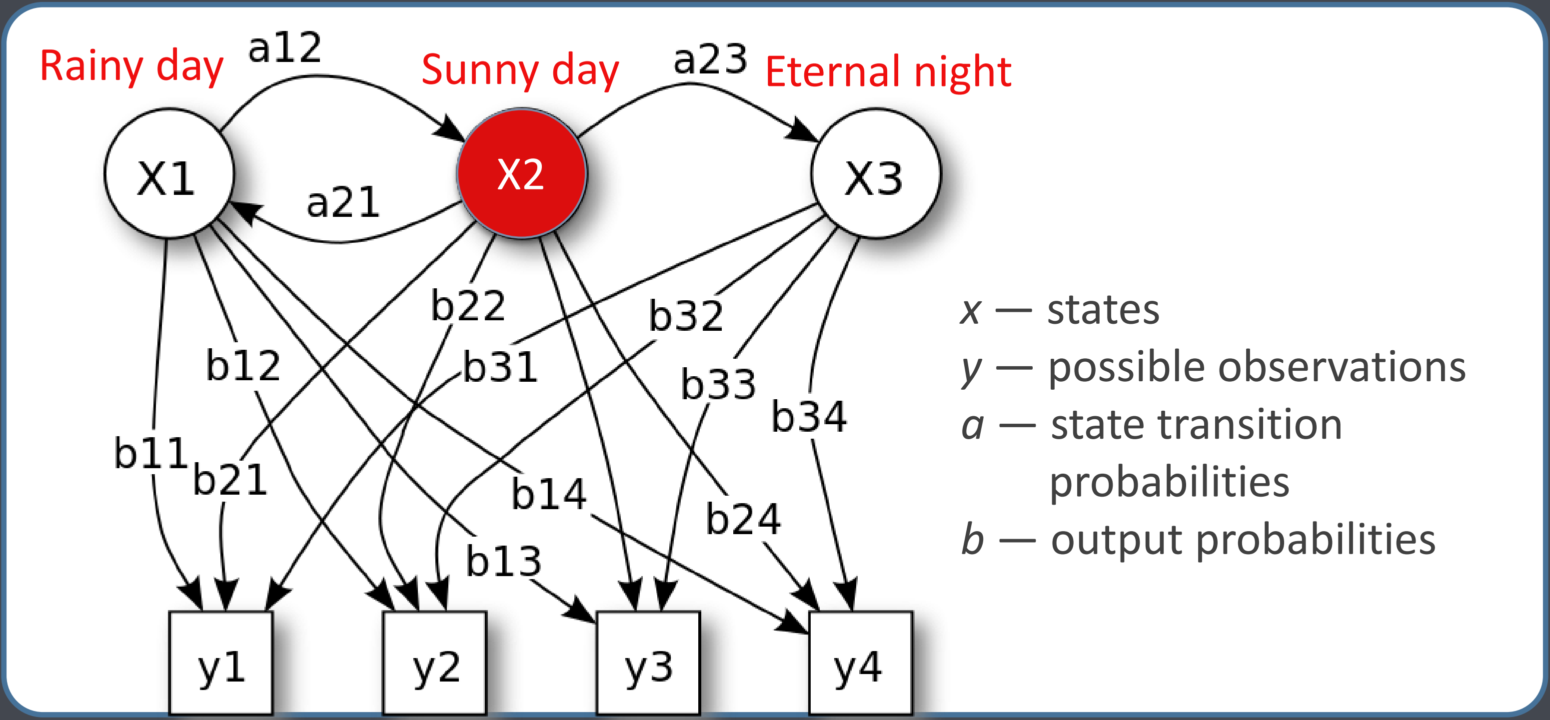


x — states
y — possible observations
a — state transition probabilities
b — output probabilities

Wikipedia

Hidden Markov Models allow annotating chromatin states

- Probabilistic parameters of a Hidden Markov Model (HMM)

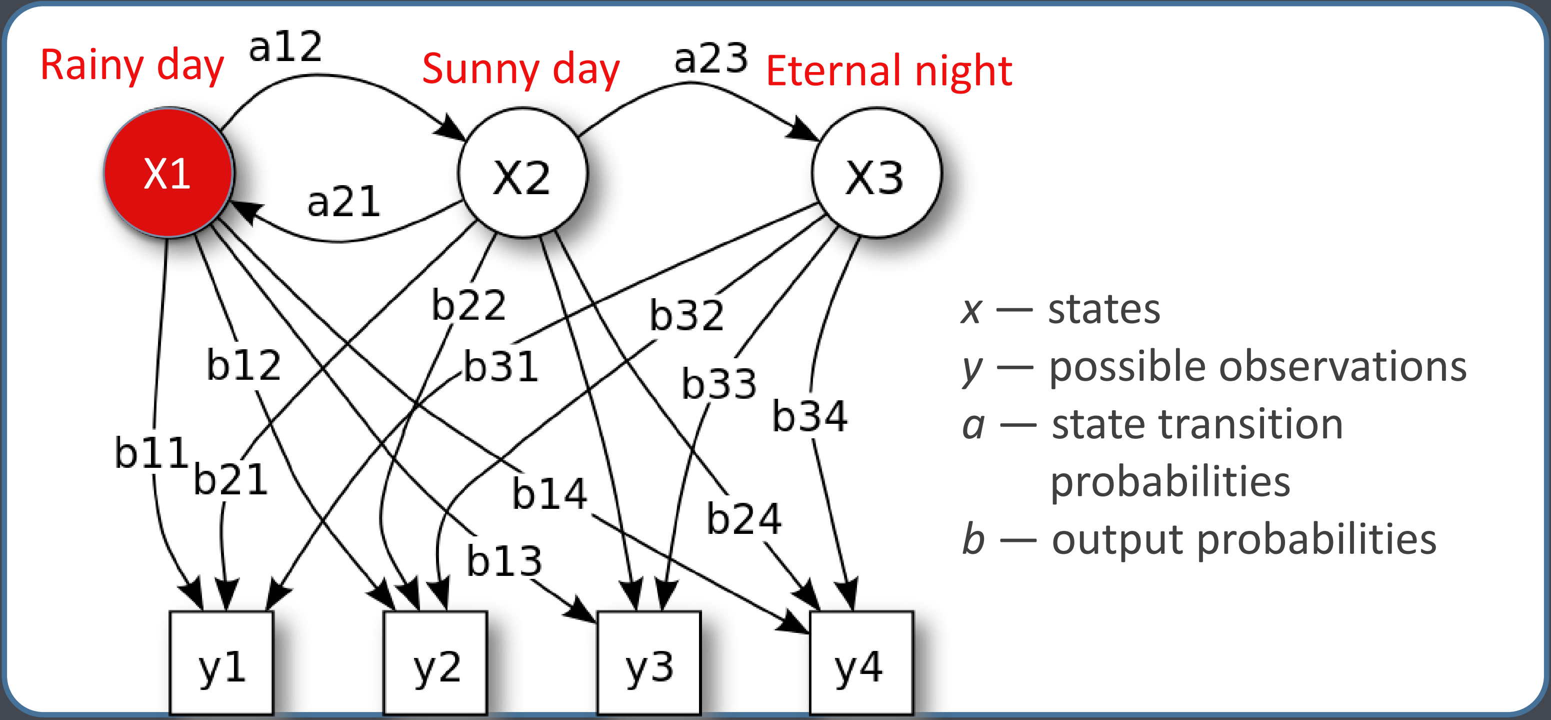


x — states
y — possible observations
a — state transition probabilities
b — output probabilities

Wikipedia

Hidden Markov Models allow annotating chromatin states

- Probabilistic parameters of a Hidden Markov Model (HMM)

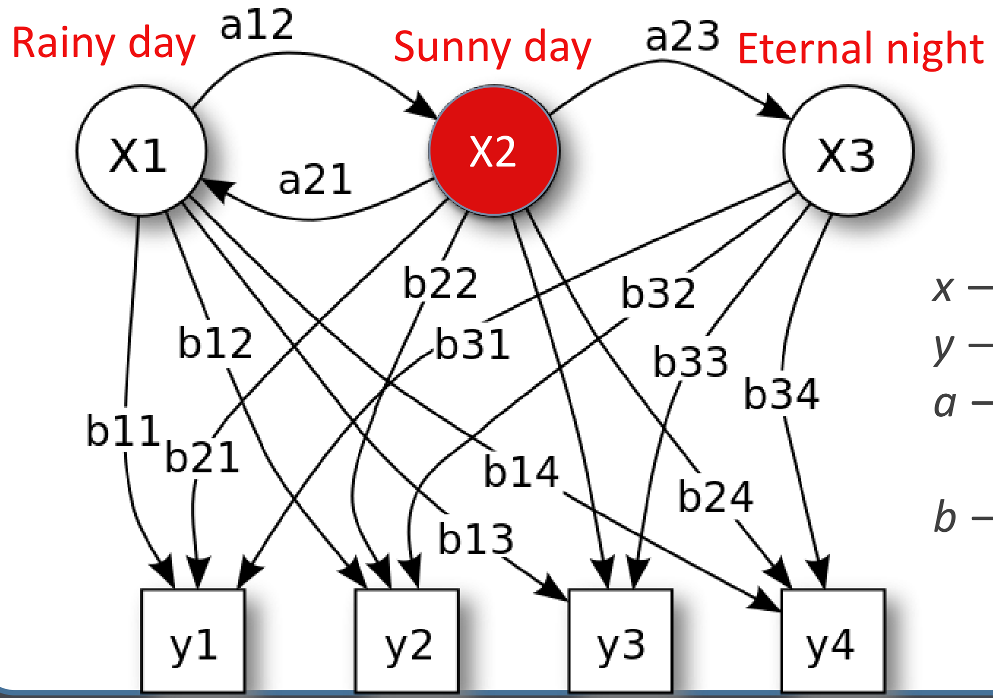


x — states
y — possible observations
a — state transition probabilities
b — output probabilities

Wikipedia

Hidden Markov Models allow annotating chromatin states

- Probabilistic parameters of a Hidden Markov Model (HMM)

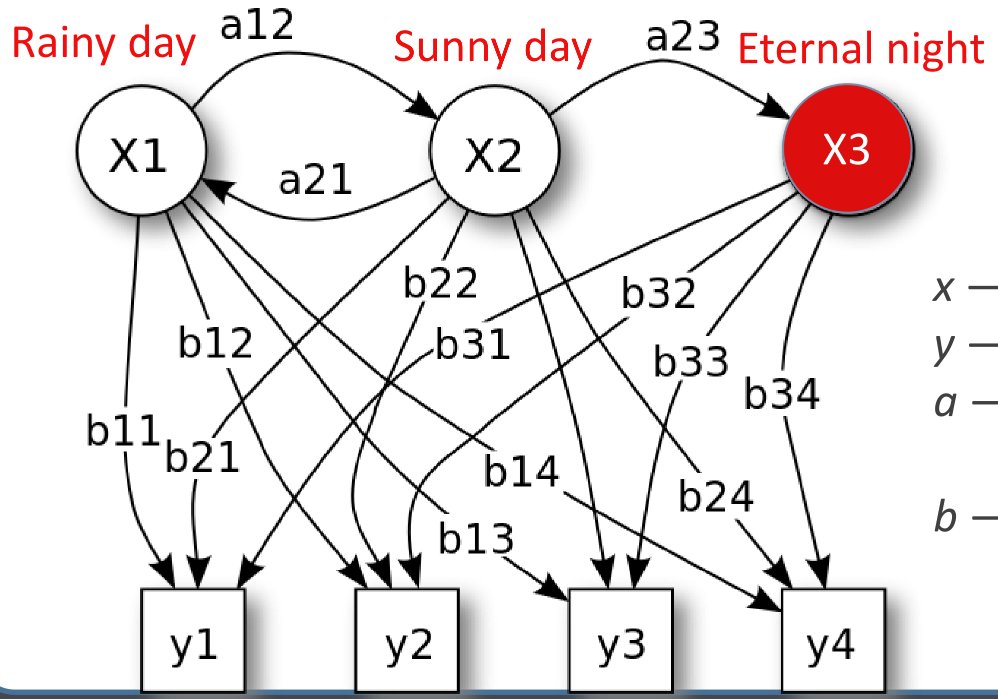


x — states
y — possible observations
a — state transition probabilities
b — output probabilities

Wikipedia

Hidden Markov Models allow annotating chromatin states

- Probabilistic parameters of a Hidden Markov Model (HMM)

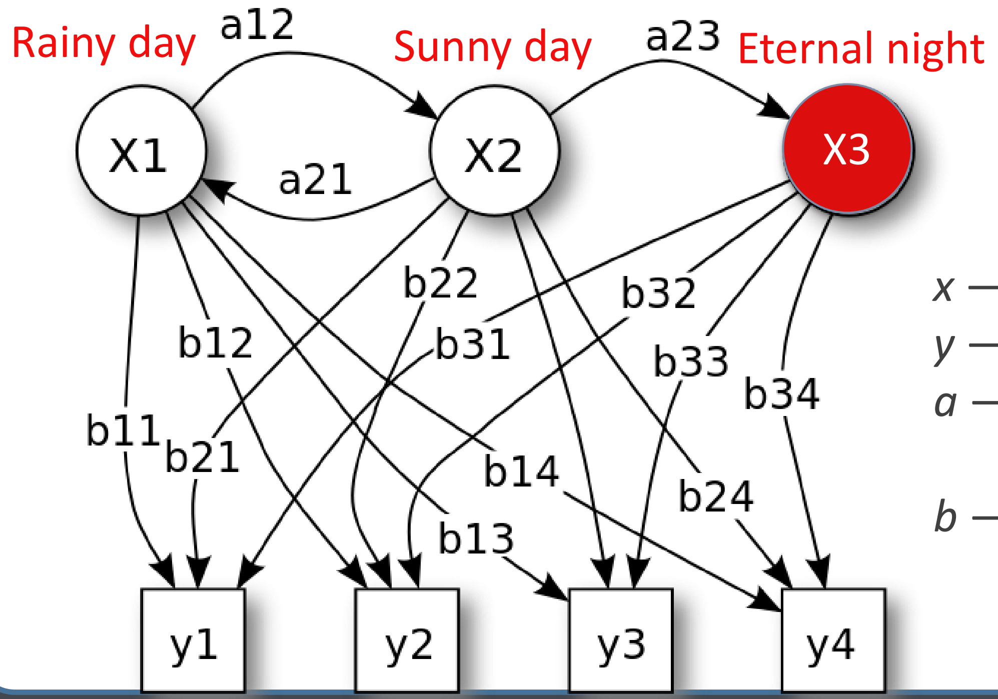


x — states
y — possible observations
a — state transition probabilities
b — output probabilities

Wikipedia

Hidden Markov Models allow annotating chromatin states

- Probabilistic parameters of a Hidden Markov Model (HMM)



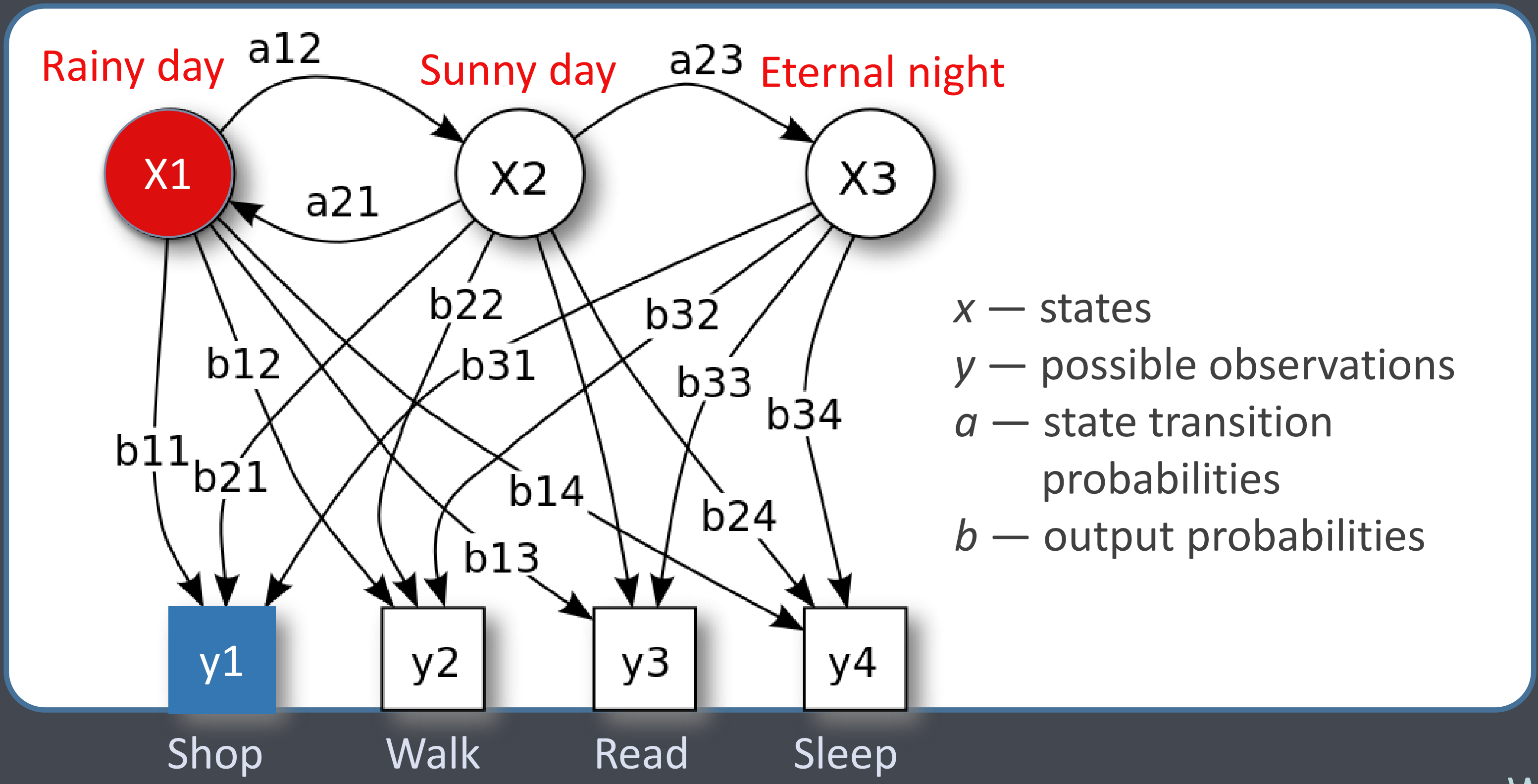
This was Markov chain

x — states
y — possible observations
a — state transition probabilities
b — output probabilities

Wikipedia

Hidden Markov Models allow annotating chromatin states

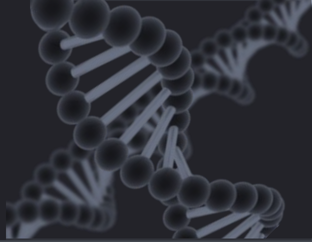
- Probabilistic parameters of a Hidden Markov Model (HMM)



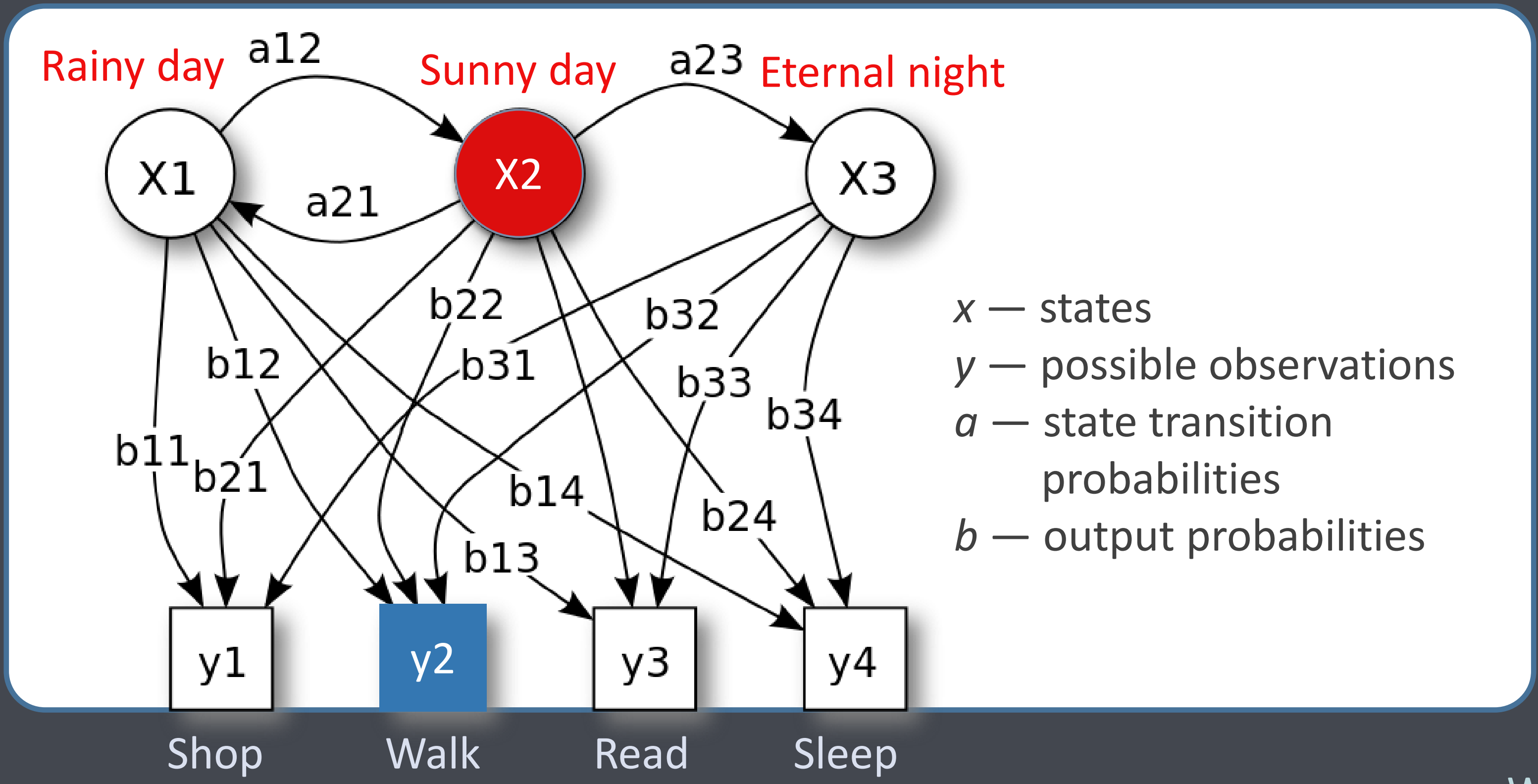
x — states
y — possible observations
a — state transition probabilities
b — output probabilities

Wikipedia

Hidden Markov Models allow annotating chromatin states



- Probabilistic parameters of a Hidden Markov Model (HMM)



x — states

y — possible observations

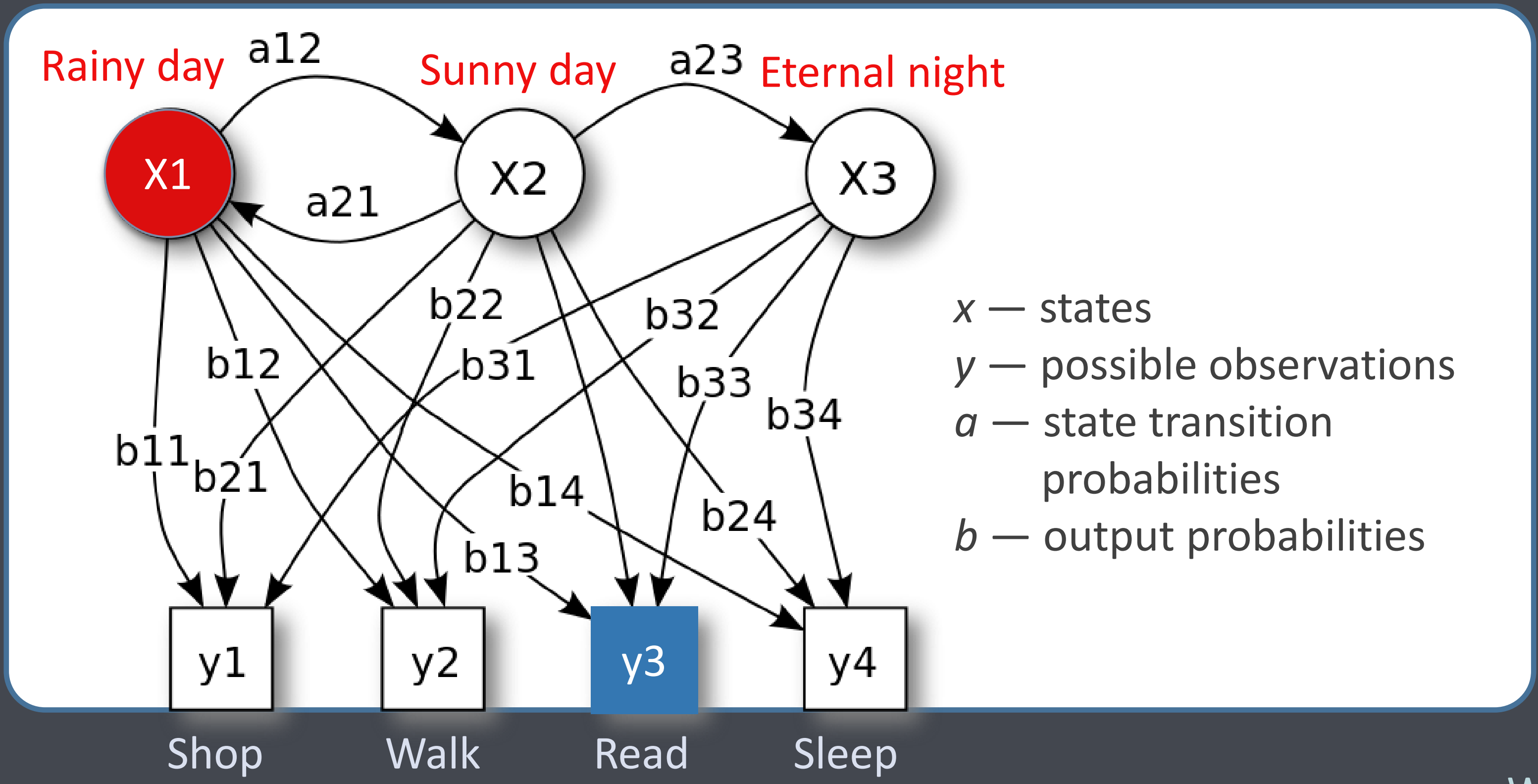
a — state transition probabilities

b — output probabilities

Wikipedia

Hidden Markov Models allow annotating chromatin states

- Probabilistic parameters of a Hidden Markov Model (HMM)

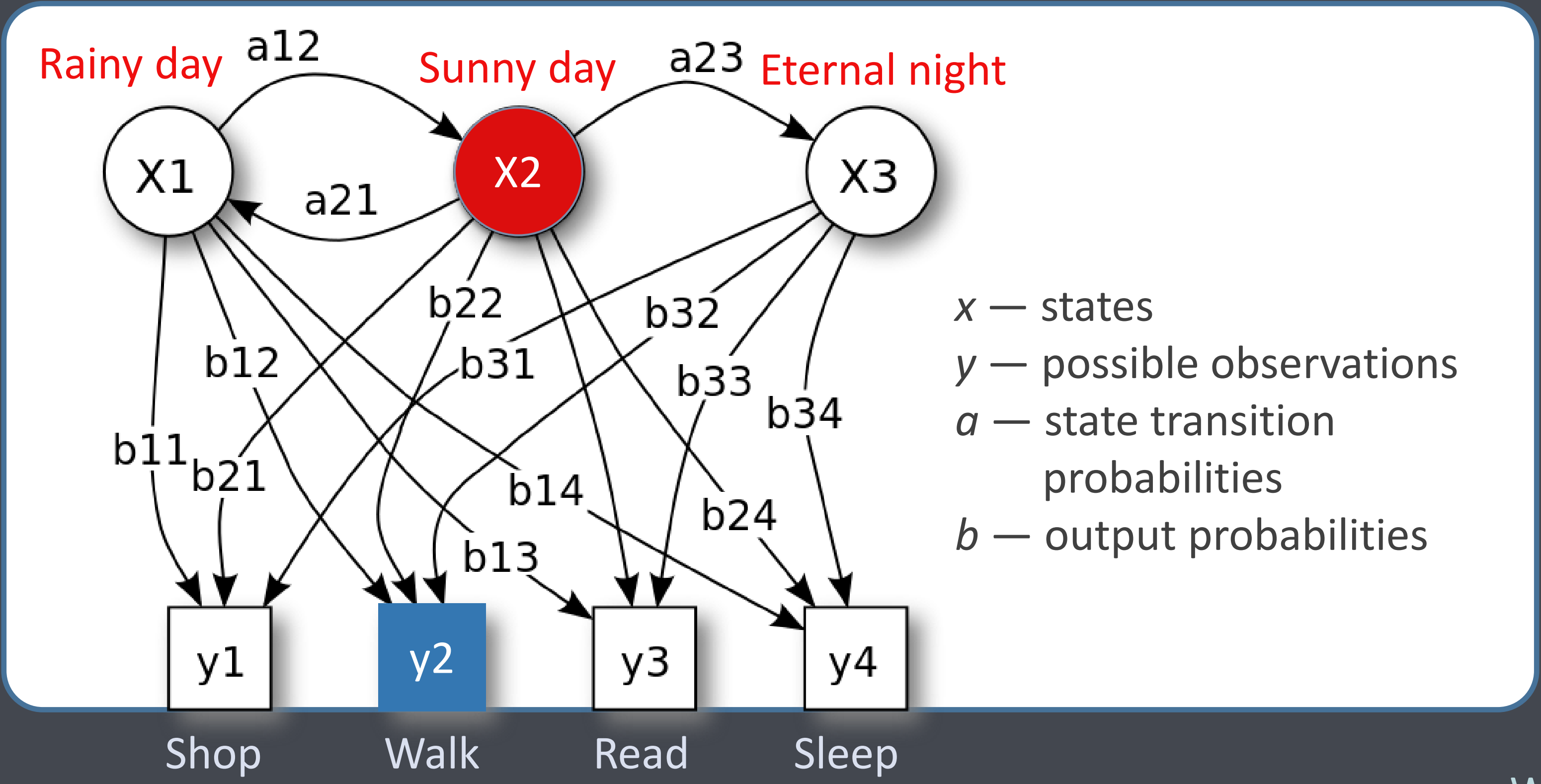


x — states
y — possible observations
a — state transition probabilities
b — output probabilities

Wikipedia

Hidden Markov Models allow annotating chromatin states

- Probabilistic parameters of a Hidden Markov Model (HMM)

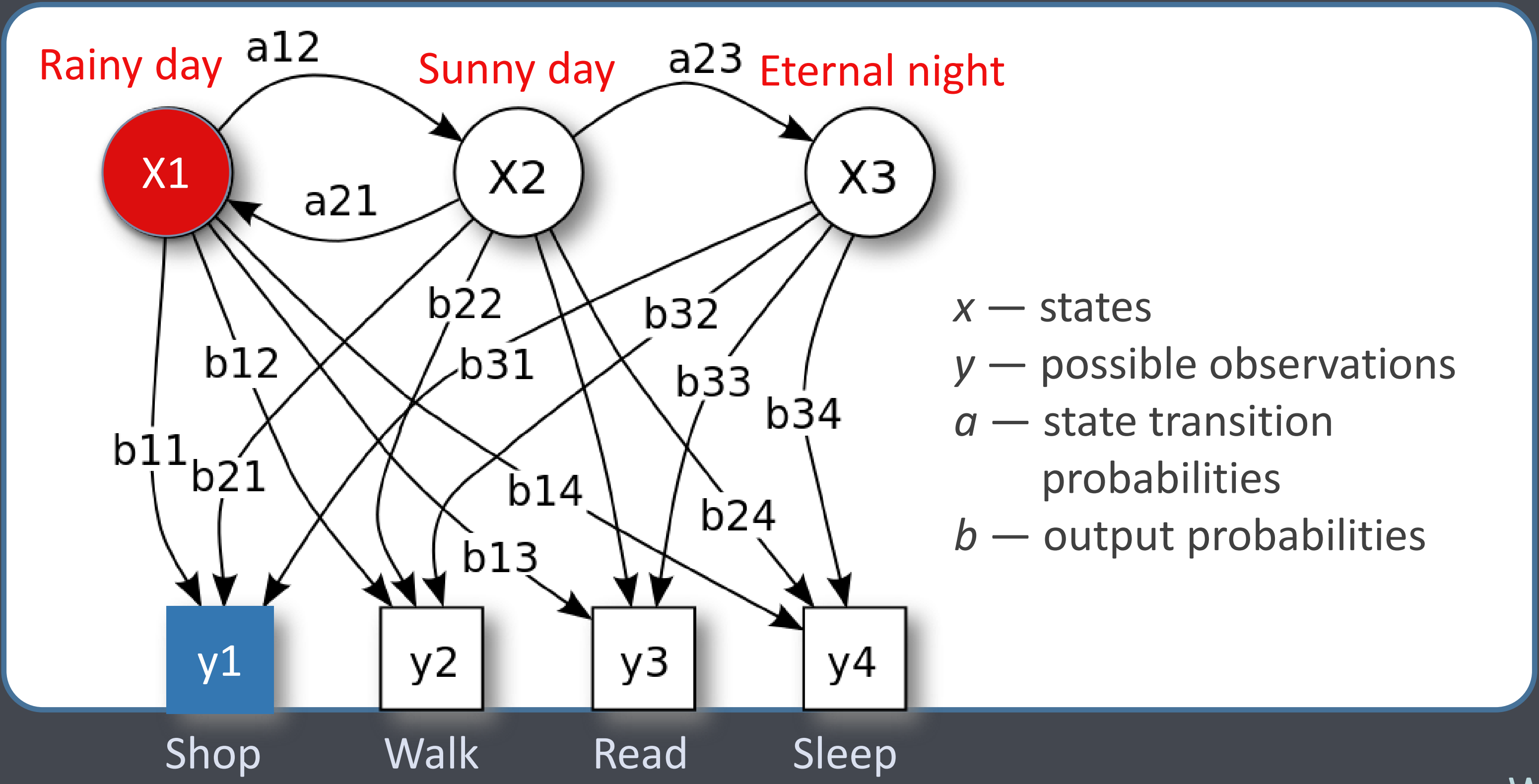


x — states
y — possible observations
a — state transition probabilities
b — output probabilities

Wikipedia

Hidden Markov Models allow annotating chromatin states

- Probabilistic parameters of a Hidden Markov Model (HMM)

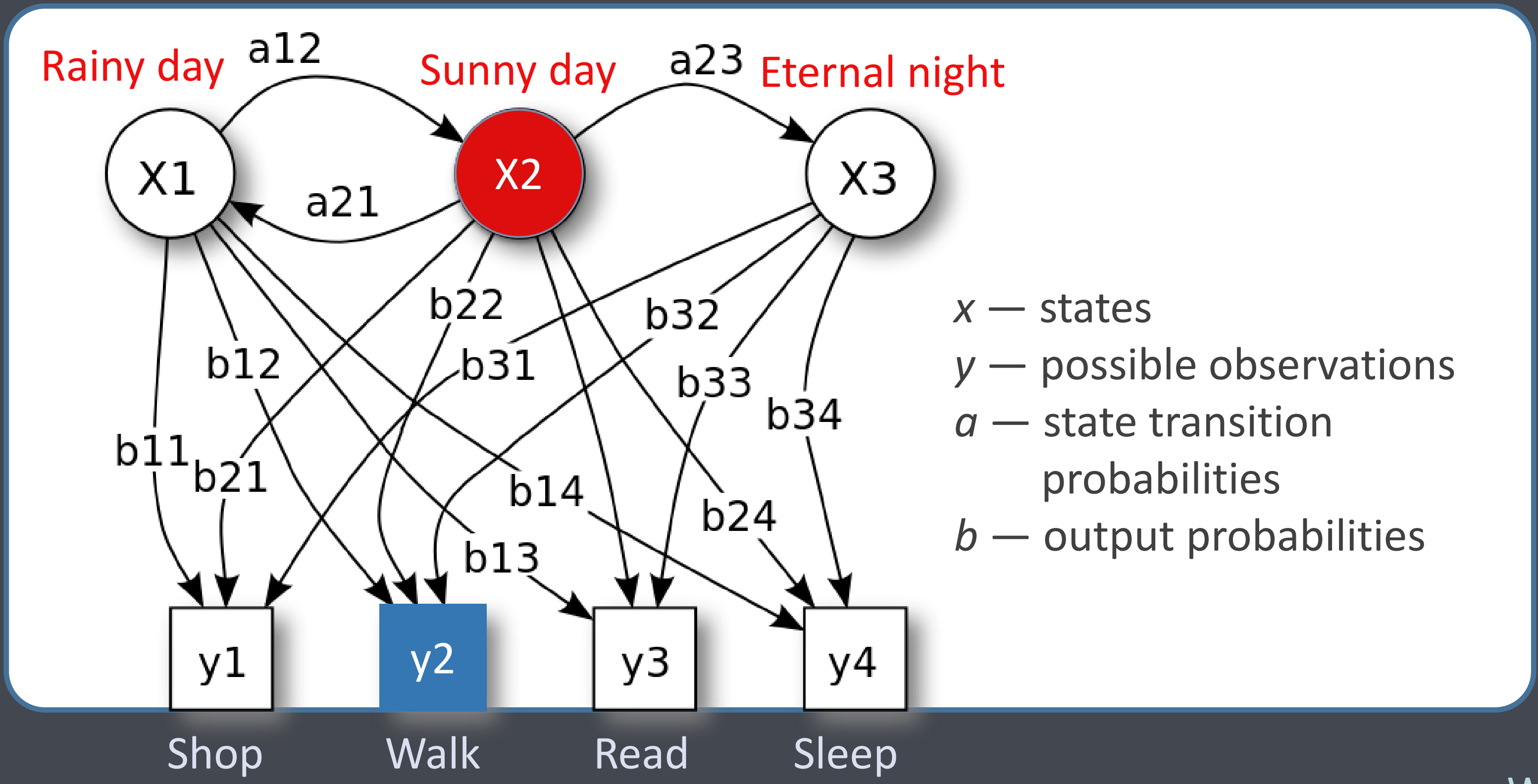


x — states
y — possible observations
a — state transition probabilities
b — output probabilities

Wikipedia

Hidden Markov Models allow annotating chromatin states

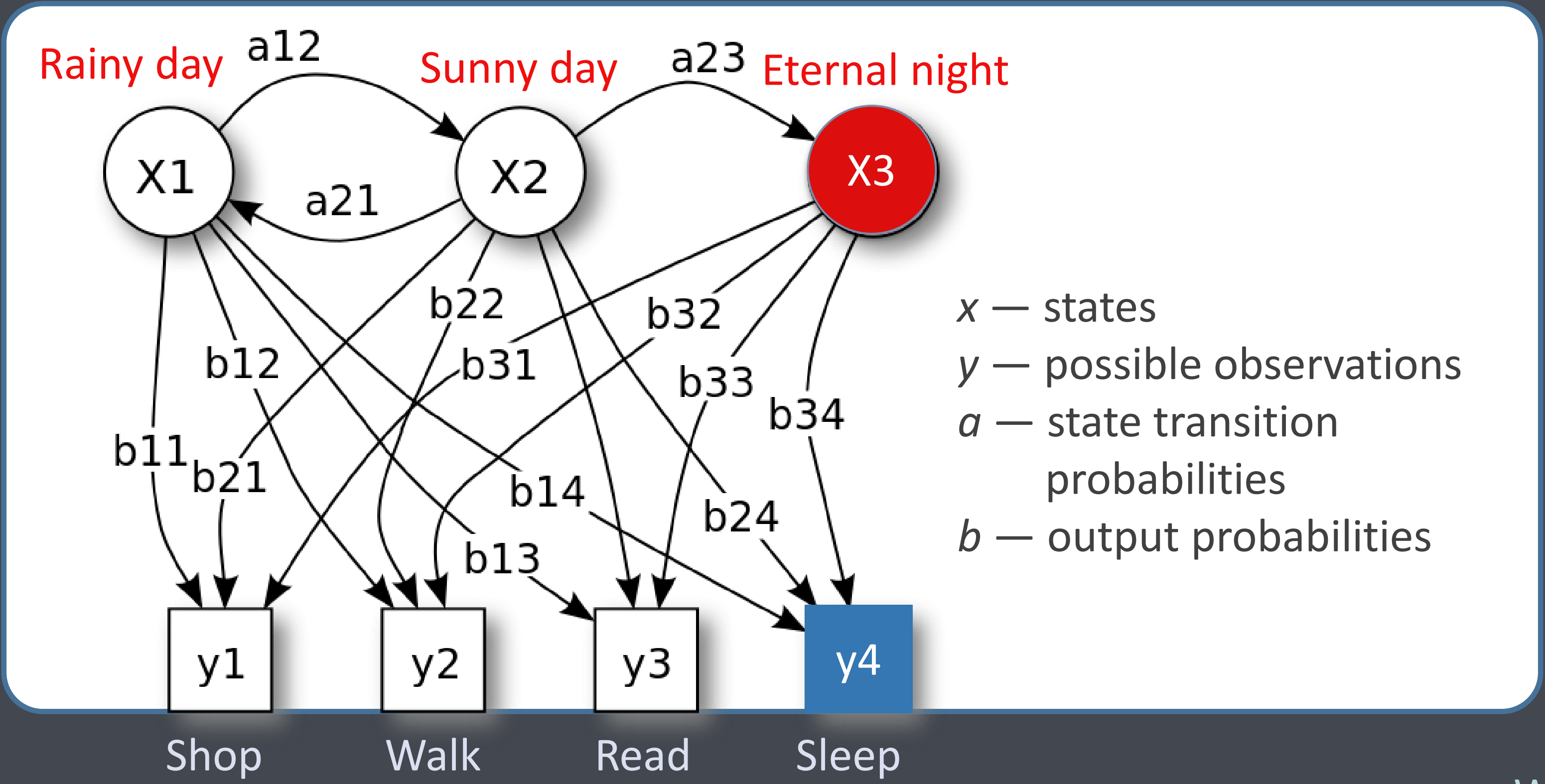
- Probabilistic parameters of a Hidden Markov Model (HMM)



Wikipedia

Hidden Markov Models allow annotating chromatin states

- Probabilistic parameters of a Hidden Markov Model (HMM)

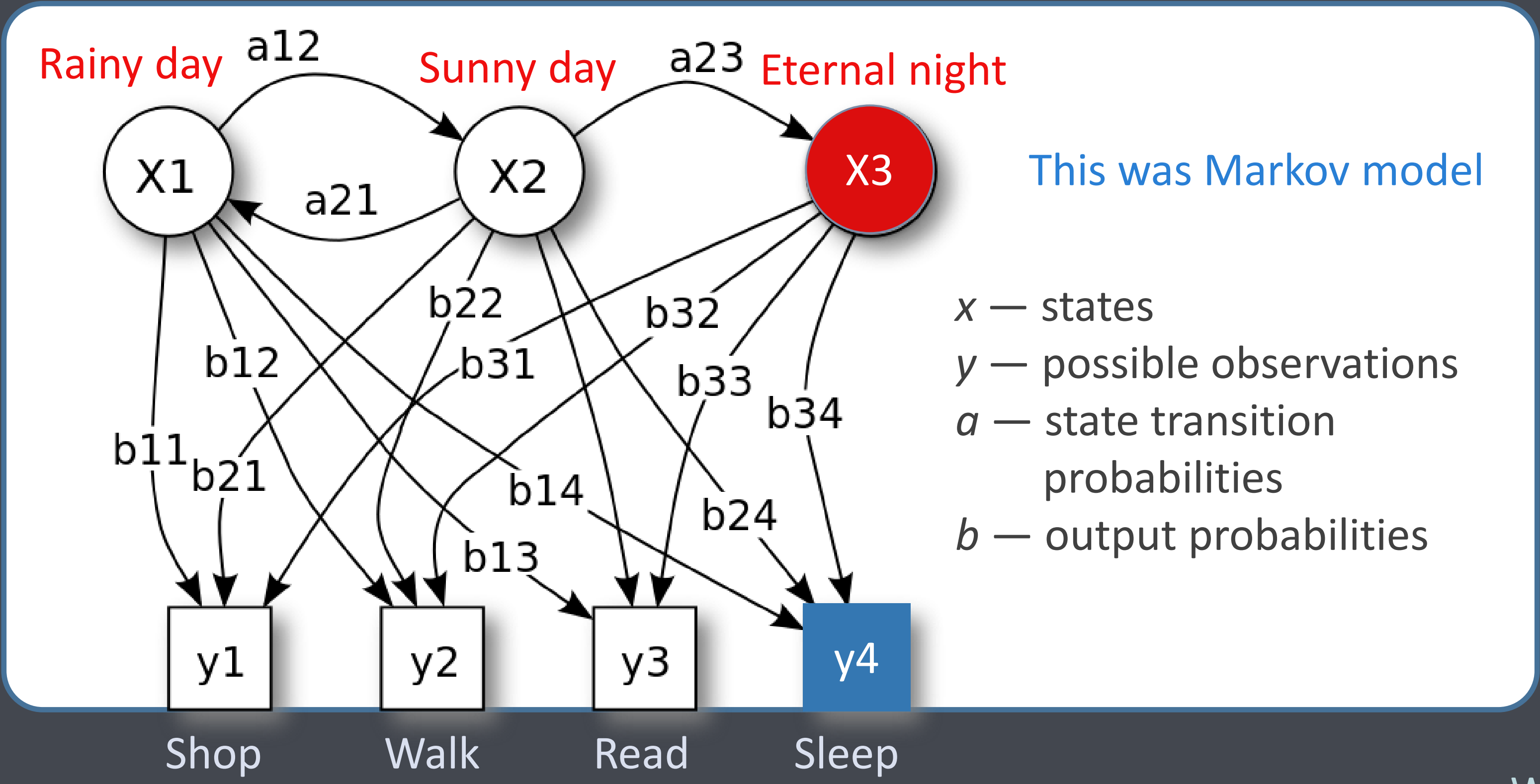


x — states
y — possible observations
a — state transition probabilities
b — output probabilities

Wikipedia

Hidden Markov Models allow annotating chromatin states

- Probabilistic parameters of a Hidden Markov Model (HMM)



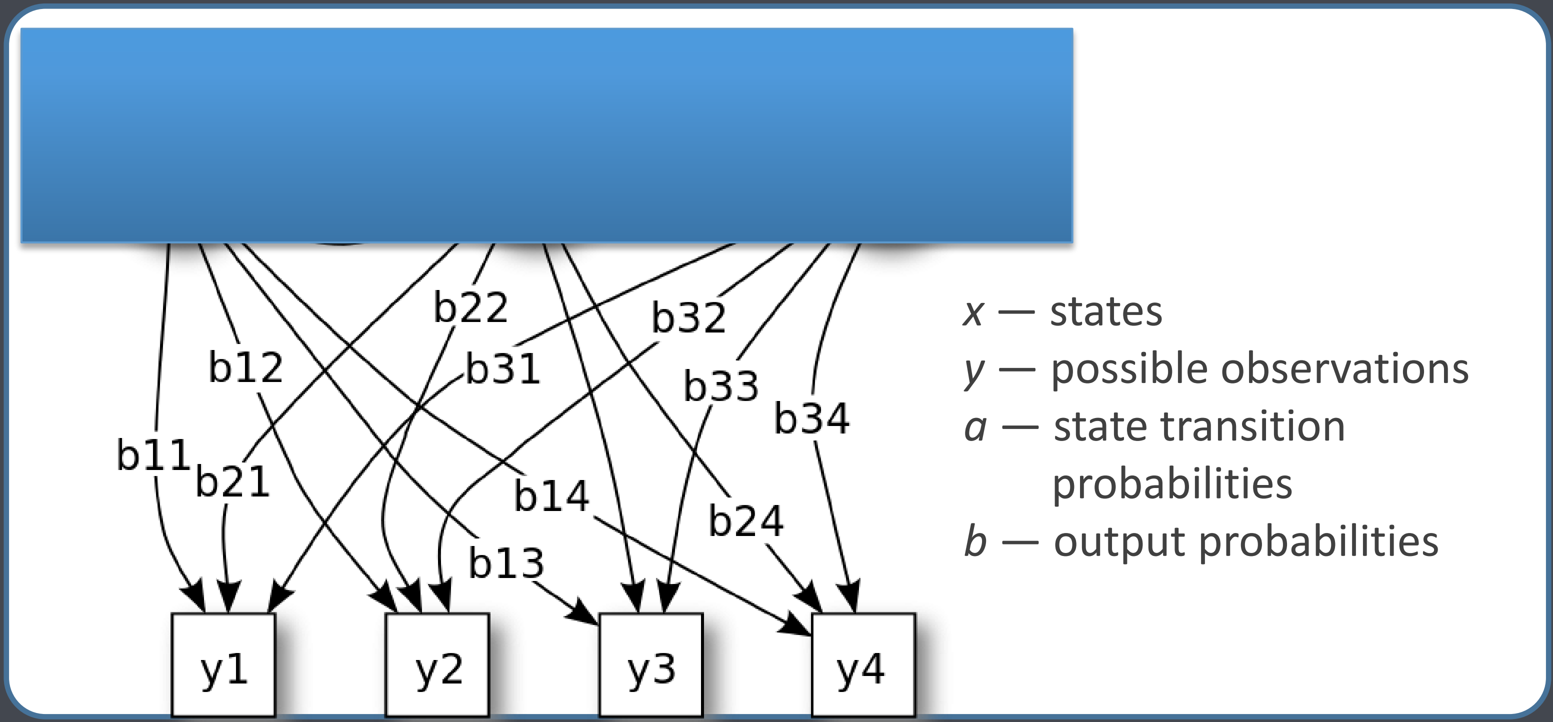
This was Markov model

x — states
y — possible observations
a — state transition probabilities
b — output probabilities

Wikipedia

Hidden Markov Models allow annotating chromatin states

- Hidden Markov Model = we cannot observe the states



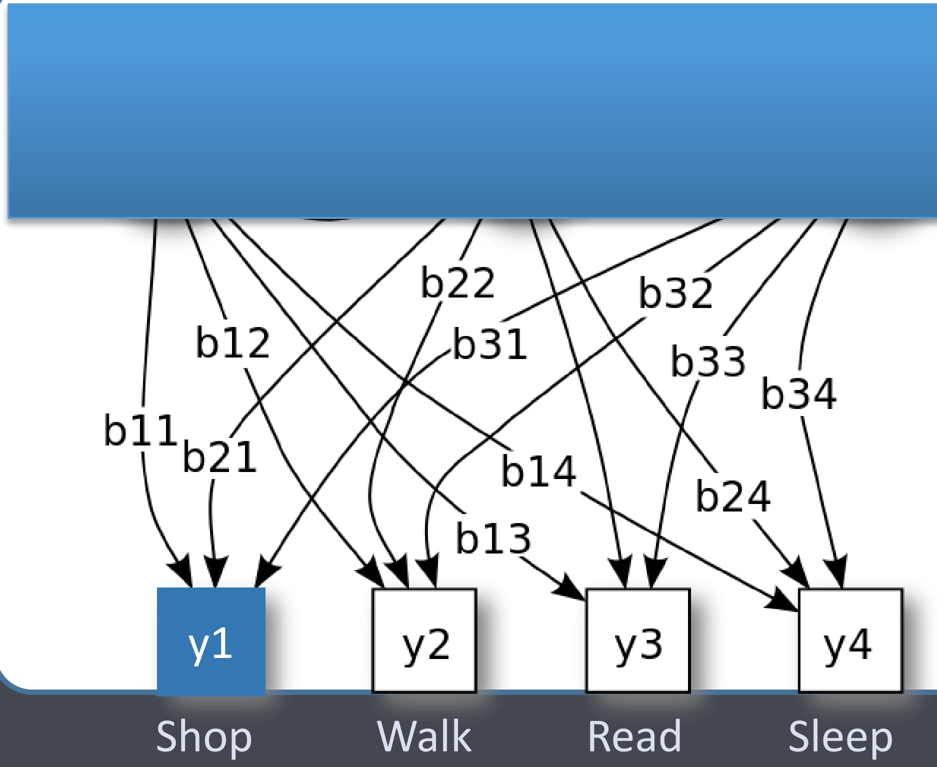
x — states
y — possible observations
a — state transition probabilities
b — output probabilities

Shop Walk Read Sleep

Wikipedia

Hidden Markov Models allow annotating chromatin states

- Probabilistic parameters of a Hidden Markov Model (HMM)

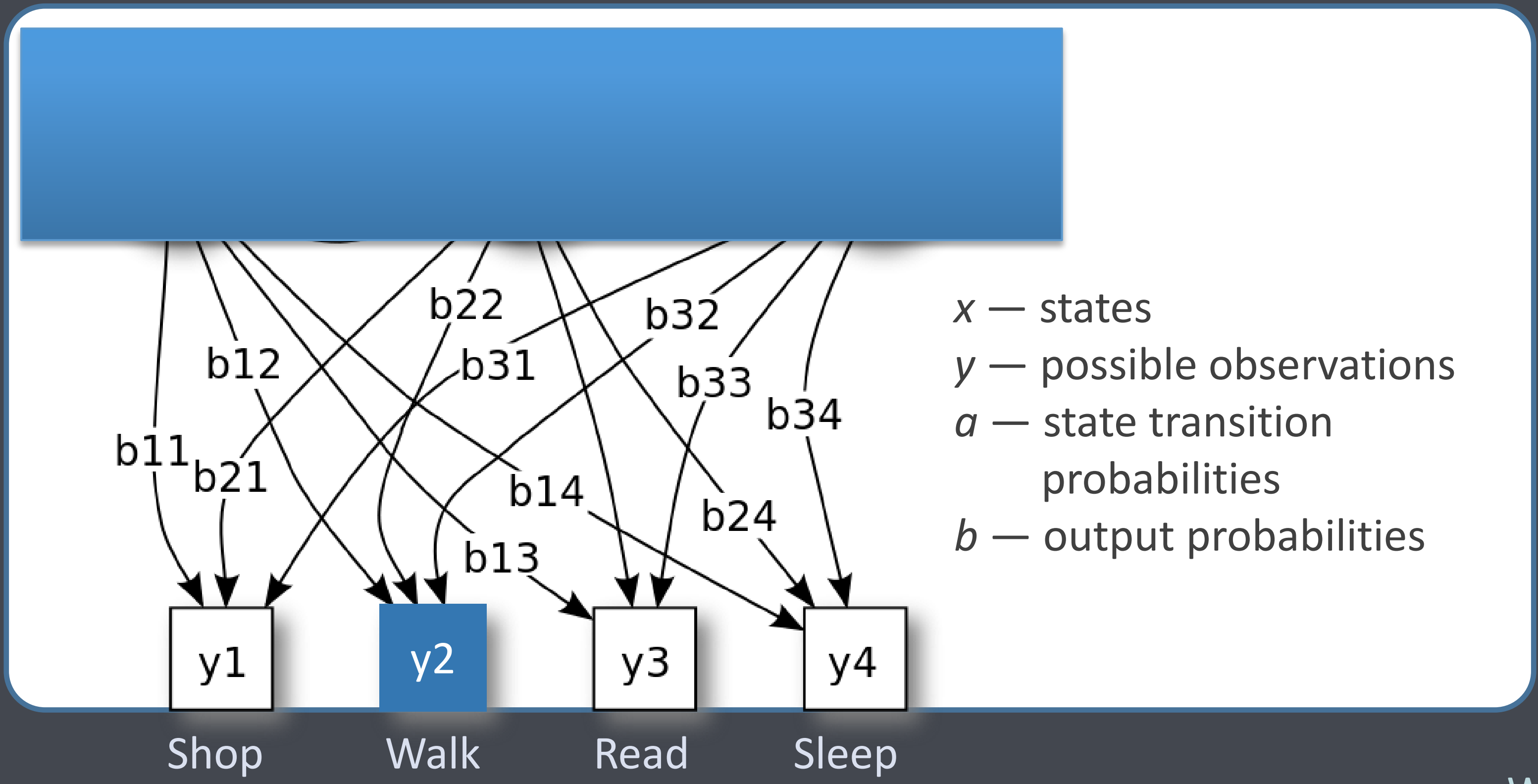


x — states
y — possible observations
a — state transition probabilities
b — output probabilities

Wikipedia

Hidden Markov Models allow annotating chromatin states

- Probabilistic parameters of a Hidden Markov Model (HMM)

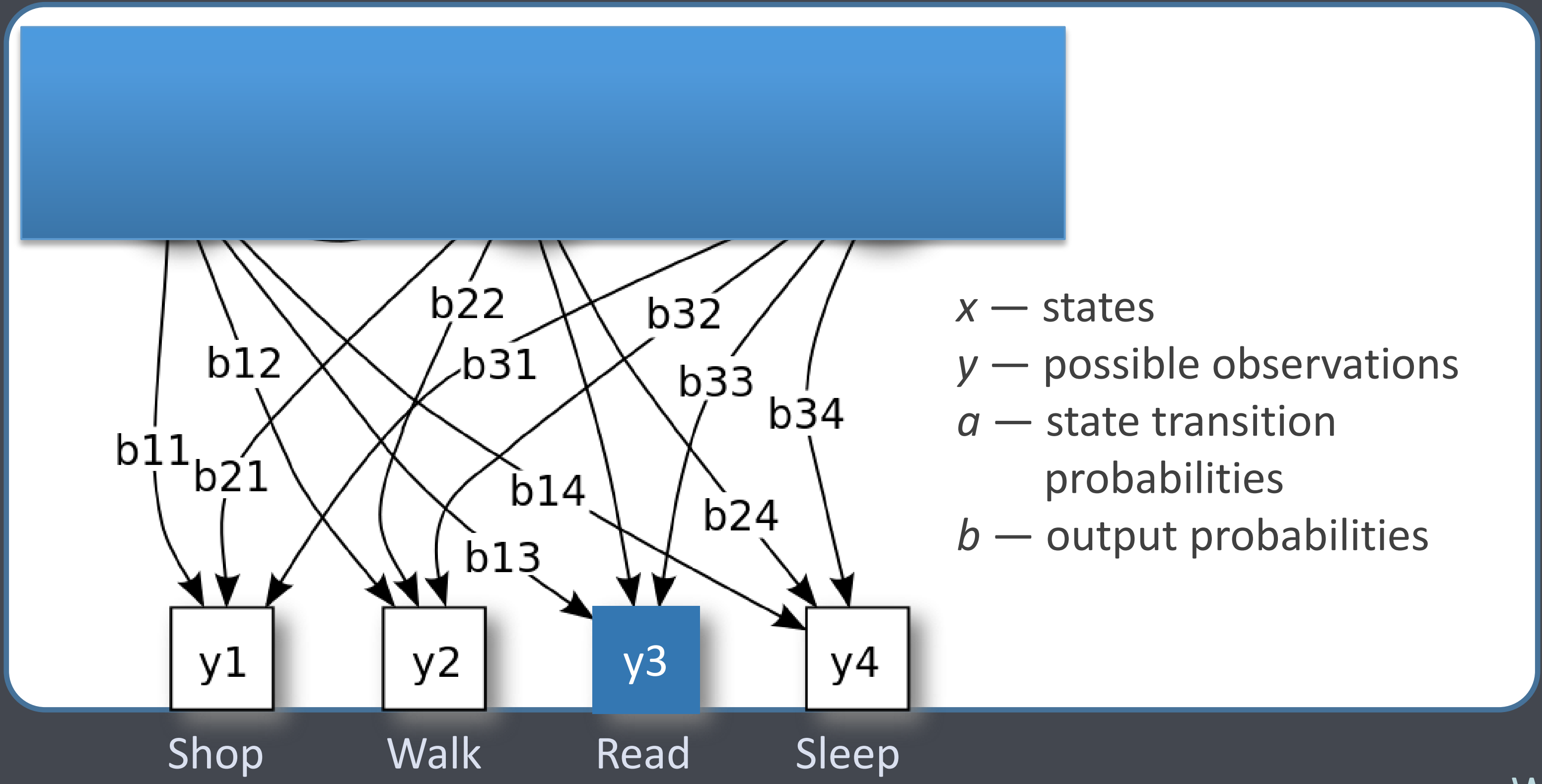


x — states
y — possible observations
a — state transition probabilities
b — output probabilities

Wikipedia

Hidden Markov Models allow annotating chromatin states

- Probabilistic parameters of a Hidden Markov Model (HMM)



x — states
y — possible observations
a — state transition probabilities
b — output probabilities

Wikipedia

Hidden Markov Models allow annotating chromatin states

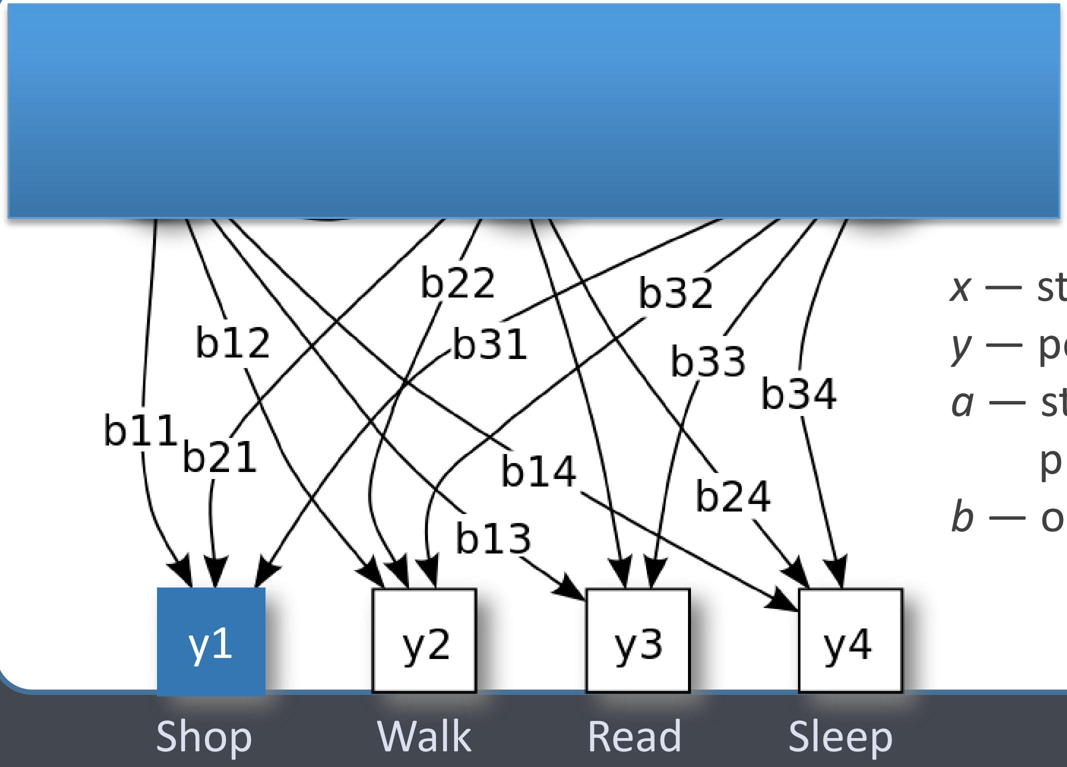
- Probabilistic parameters of a Hidden Markov Model (HMM)

x — states
y — possible observations
a — state transition probabilities
b — output probabilities

Wikipedia

Hidden Markov Models allow annotating chromatin states

- Probabilistic parameters of a Hidden Markov Model (HMM)

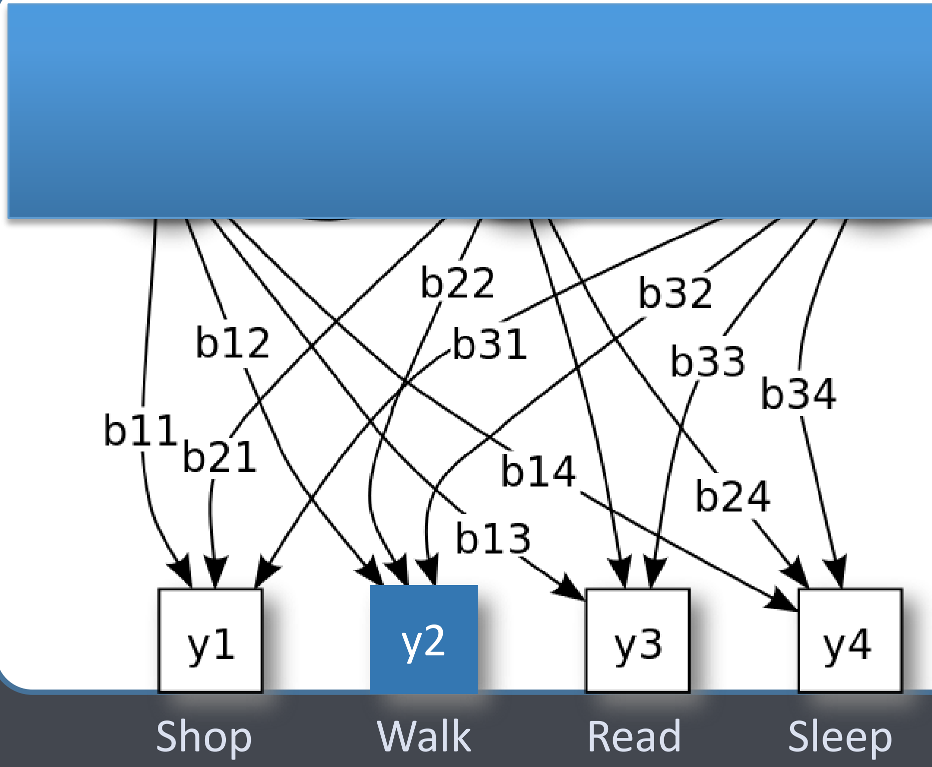


x — states
y — possible observations
a — state transition probabilities
b — output probabilities

Wikipedia

Hidden Markov Models allow annotating chromatin states

- Probabilistic parameters of a Hidden Markov Model (HMM)

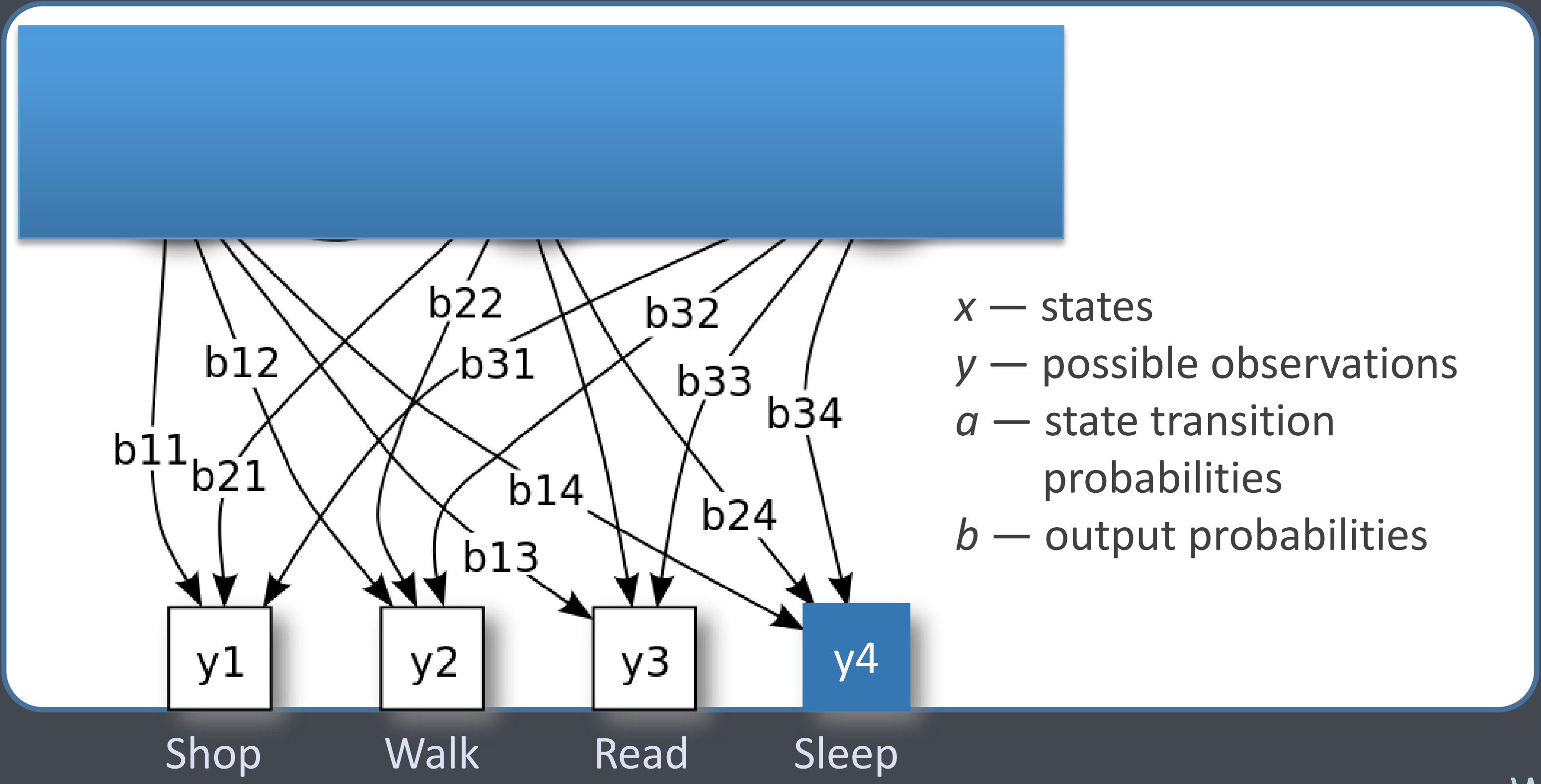


x — states
y — possible observations
a — state transition probabilities
b — output probabilities

Wikipedia

Hidden Markov Models allow annotating chromatin states

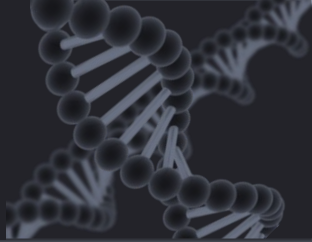
- Probabilistic parameters of a Hidden Markov Model (HMM)



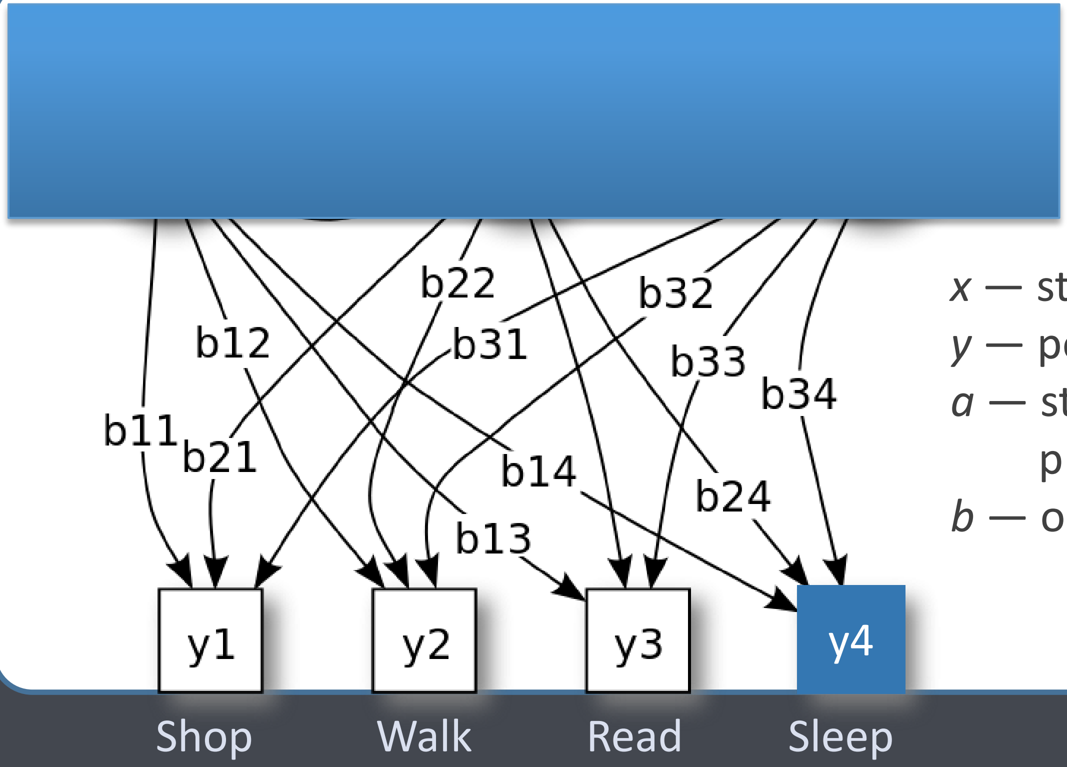
x — states
y — possible observations
a — state transition probabilities
b — output probabilities

Wikipedia

Hidden Markov Models allow annotating chromatin states



- Probabilistic parameters of a Hidden Markov Model (HMM)

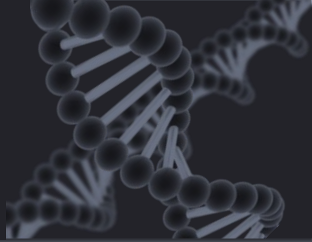


This was hidden Markov model

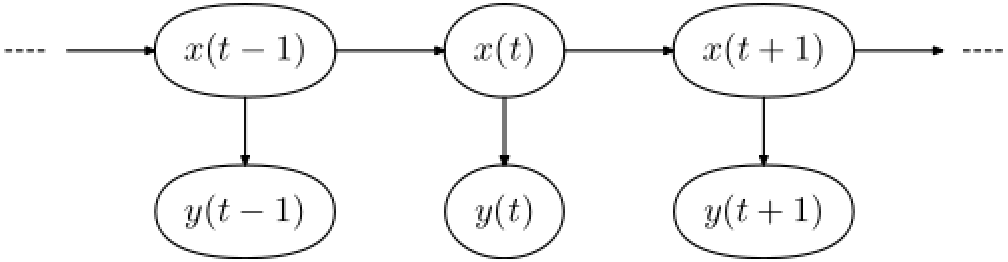
x — states
y — possible observations
a — state transition probabilities
b — output probabilities

Wikipedia

Hidden Markov Models allow annotating chromatin states



- States change in time, but we do not observe them
- States define our observations
- First order HMM: state (t+1) depends only on the previous state (t)



Decoding question in HMM:

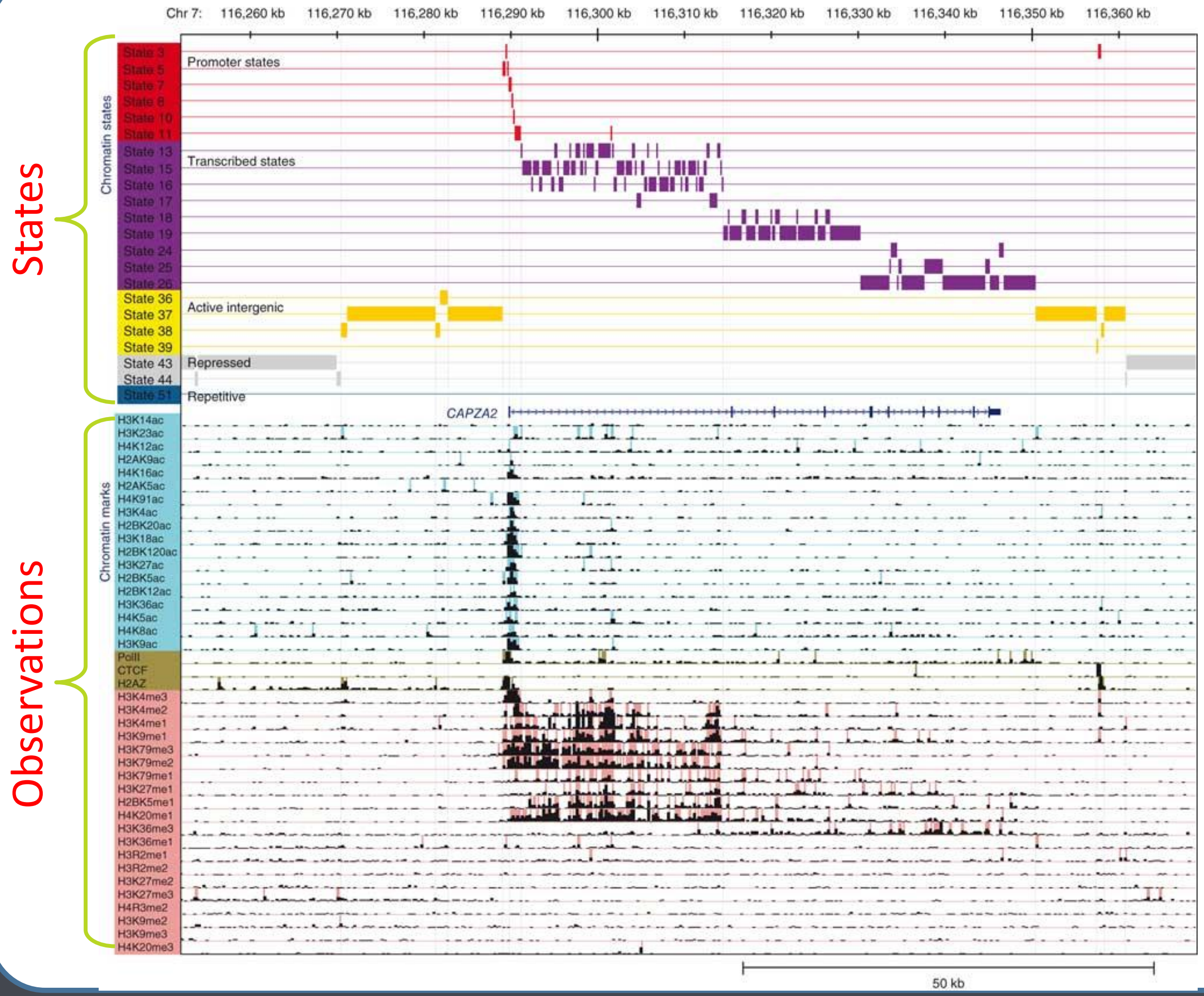
Solution: Viterbi Algorithm

x — states to be defined!

known $\left\{ \begin{array}{l} y \text{ — possible observations} \\ a \text{ — state transition probabilities} \\ b \text{ — output probabilities} \end{array} \right.$

Wikipedia

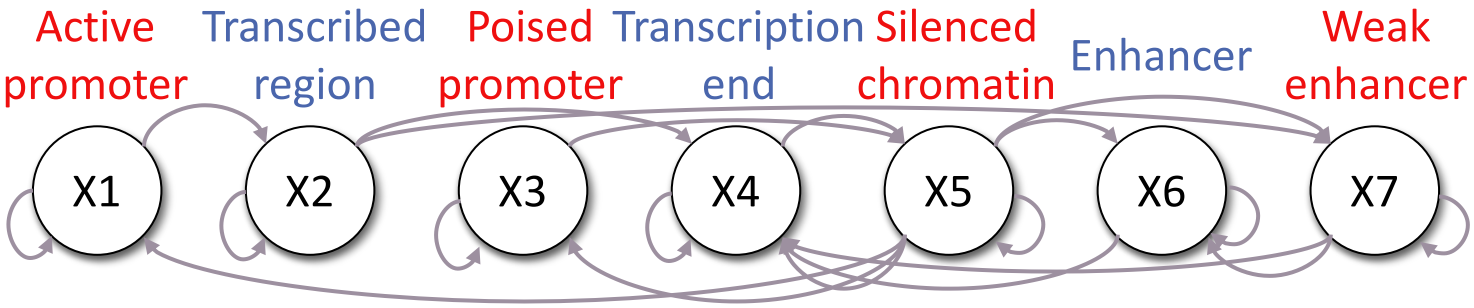
Histone modifications = observations
Functional regions = states



Input chromatin mark information and resulting chromatin state annotation for a 120-kb region of human chromosome 7 surrounding the *CAPZA2* gene

Ernst & Kellis, *Nature Biotechnology,* 2010

Hidden Markov Models allow annotating chromatin states



Active promoter

Transcribed region

Poised promoter

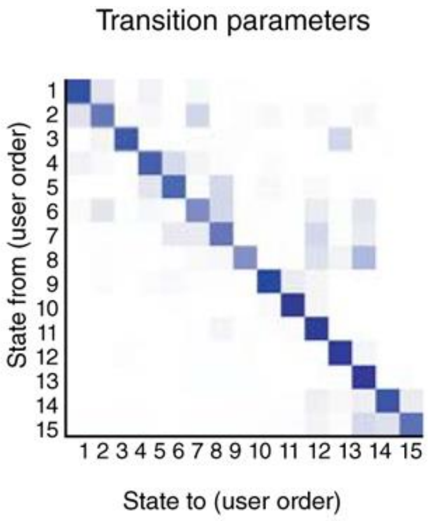
Transcription end

Silenced chromatin

Enhancer

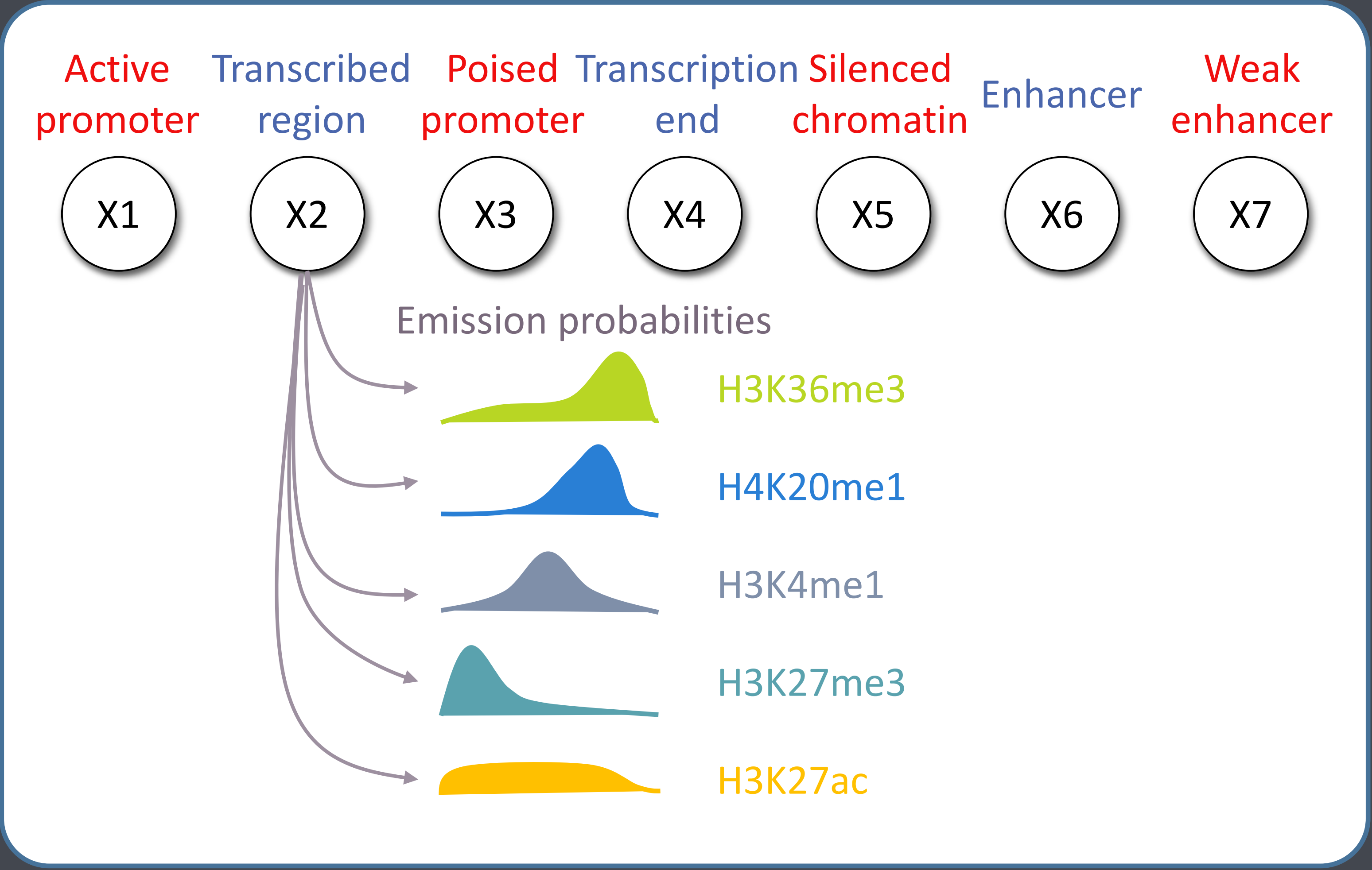
Weak enhancer

Transition probabilities

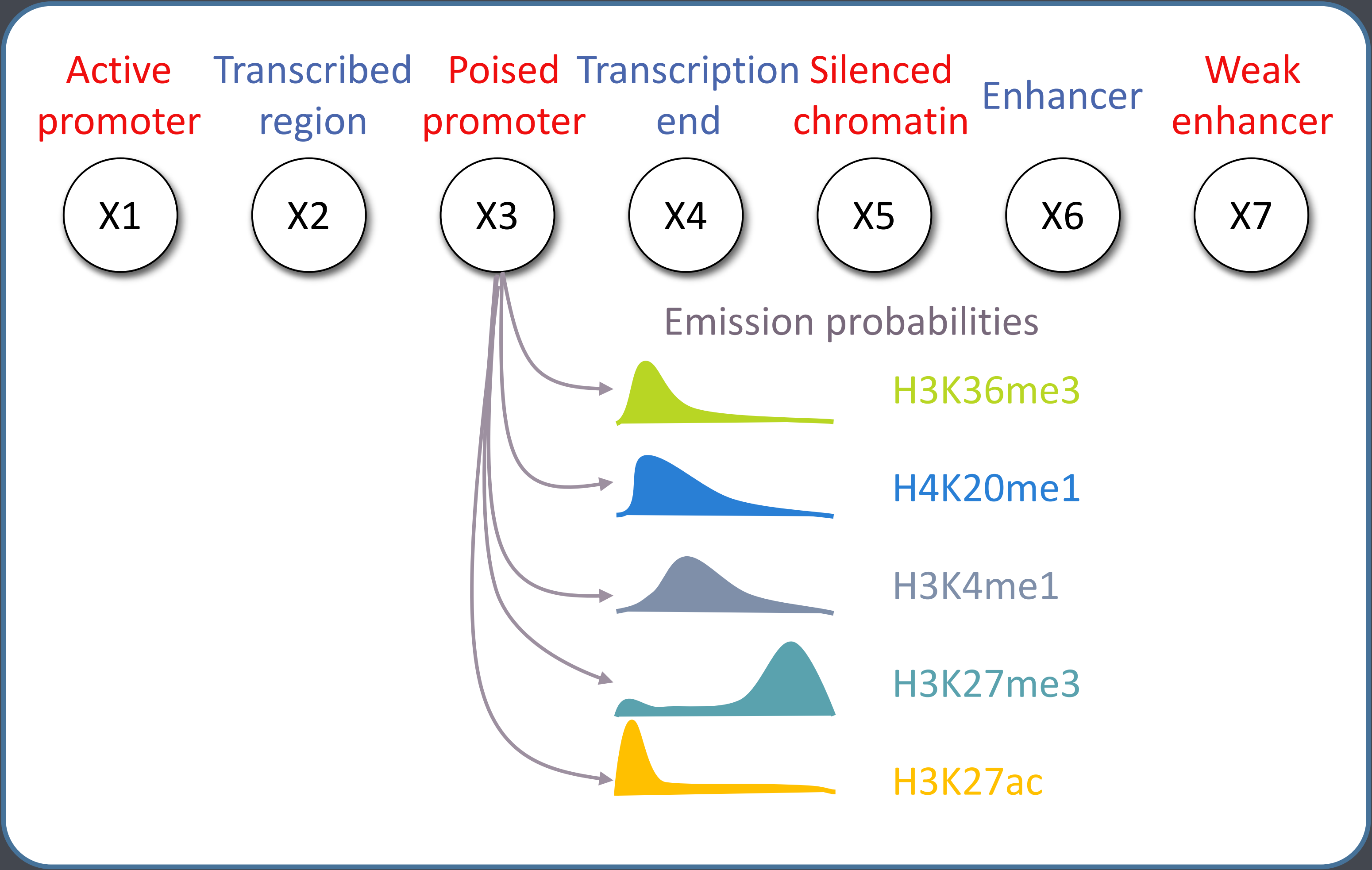


Ernst & Kellis, *Nature Methods*, 2012

Hidden Markov Models allow annotating chromatin states



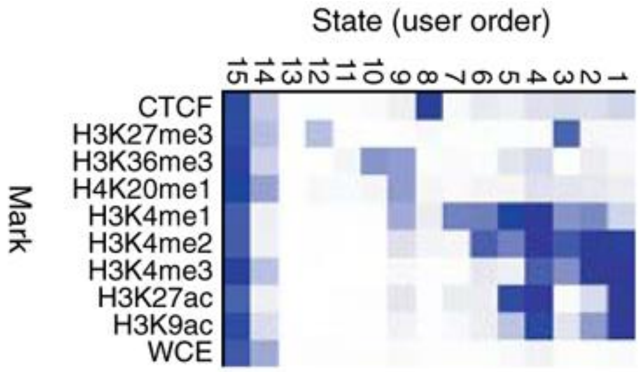
Hidden Markov Models allow annotating chromatin states



Hidden Markov Models allow annotating chromatin states

Active promoter	Transcribed region	Poised promoter	Transcription end	Silenced chromatin	Enhancer	Weak enhancer
X1	X2	X3	X4	X5	X6	X7

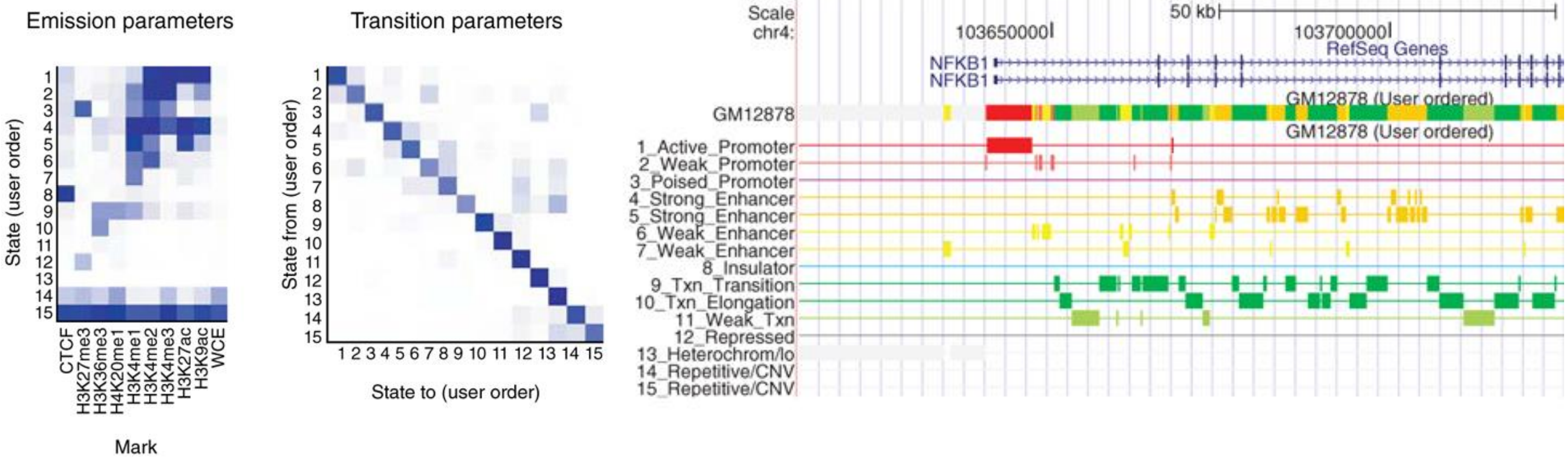
Emission probabilities



Ernst & Kellis, *Nature Methods*, 2012

Hidden Markov Models allow annotating chromatin states

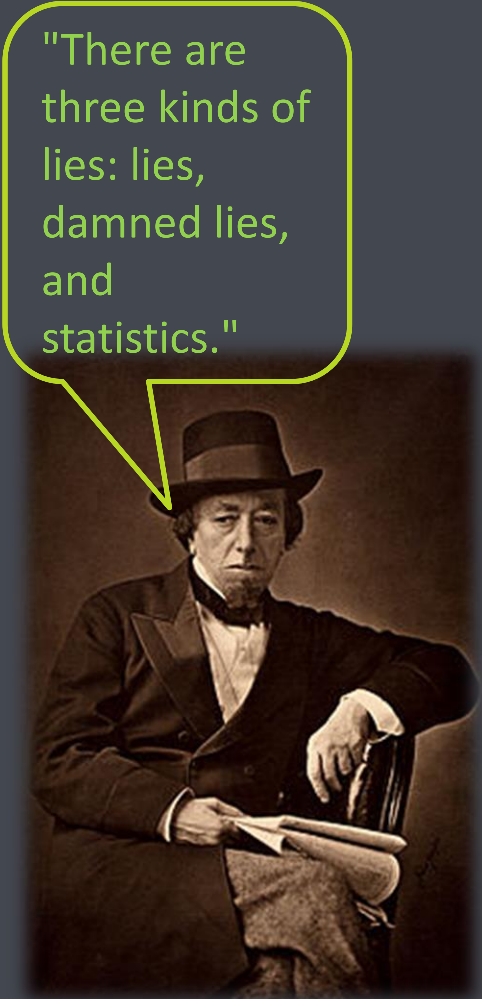
- ChromHMM: automating chromatin-state discovery and characterization



Ernst & Kellis, *Nature Methods*, 2012

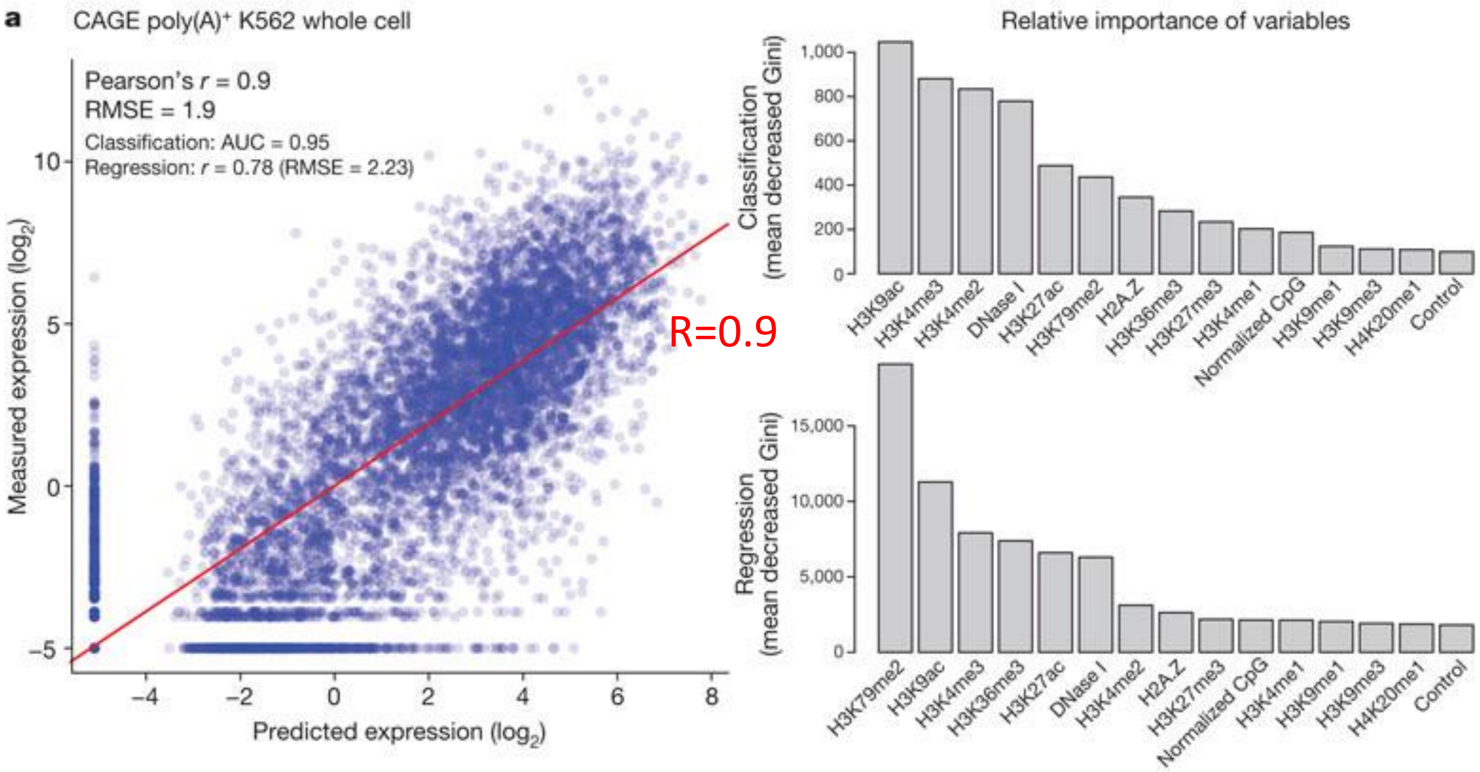
How many states to select?

- Use your biological intuition
- Score each model based on the log likelihood of the model minus a penalization on the model complexity determined by the Bayesian Information Criterion (BIC) of one-half the number of parameters times the natural log of the number of intervals



Benjamin Disraeli

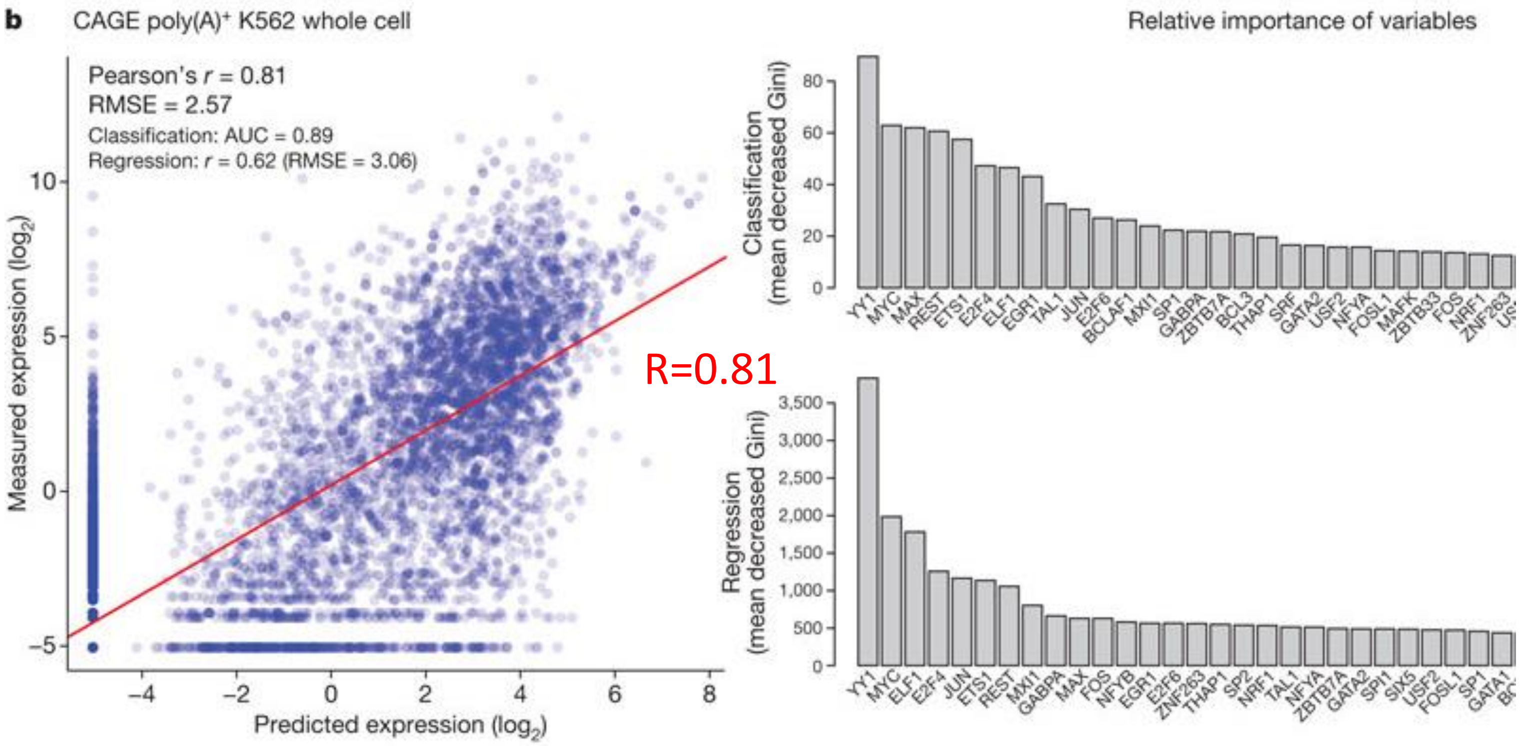
With chromatine states, one can predict gene expression



R=0.9

ENCODE Project Consortium, *Nature,* 2012

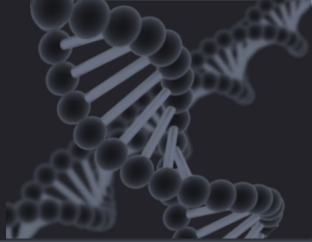
With chromatine states, one can predict gene expression better than with TFBSs



R=0.81

ENCODE Project Consortium, *Nature*, 2012

The prediction of expression is done using linear regression model



- We try to predict:

  $\log_2(Yi)$, where Y_i is expression value (based on CAGE poly-A experiment)

- using

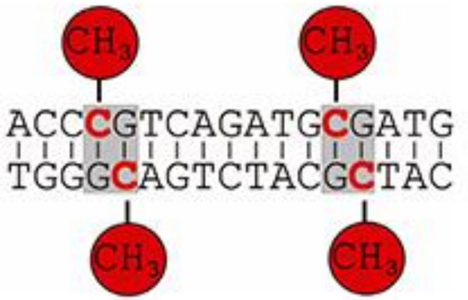
  $\log_2(X_{ij} + a_j)$, where X_{ij} is normalized tag counts for chromatin feature *j* at gene *i* (a_j = pseudocount)

$$\log_2(Yi) \approx \sum_{chromatin\ feature\ j} \alpha_j \log_2(X_{ij} + a_j) + \beta$$

Alternatively, it is possible to use Random Forests or multivariate adaptive regression splines (MARS).

Dong et al, *Genome Biol.*, 2012

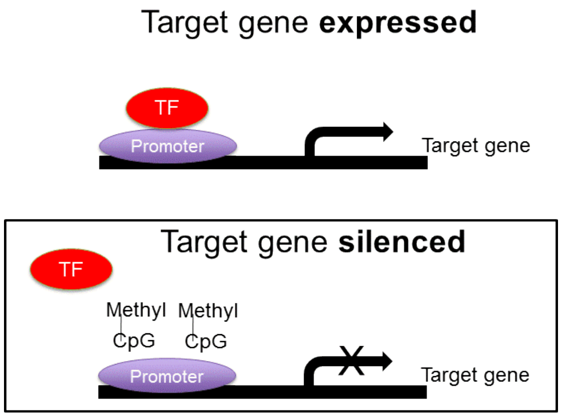
Methylation of CpG sites on DNA is also related to transcription



Addition of a methyl group to a cytosine within CpG di-nucleotide which are frequently located in the regulatory regions of genes

A mechanism for gene silencing:

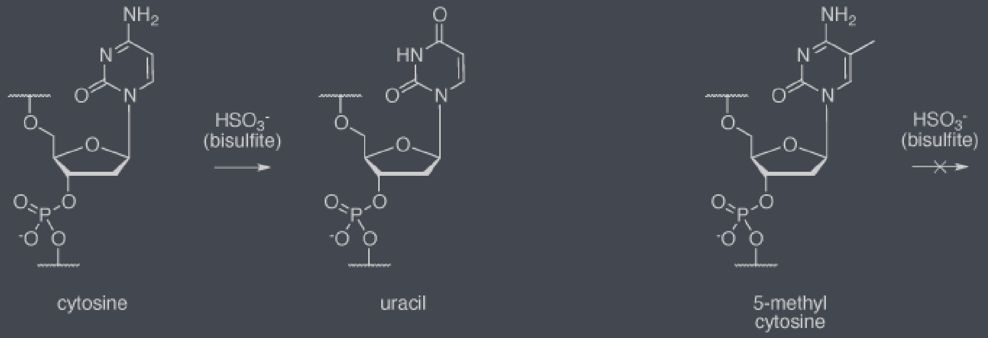
- Preventing binding of regulatory factors
- Affecting chromatin status

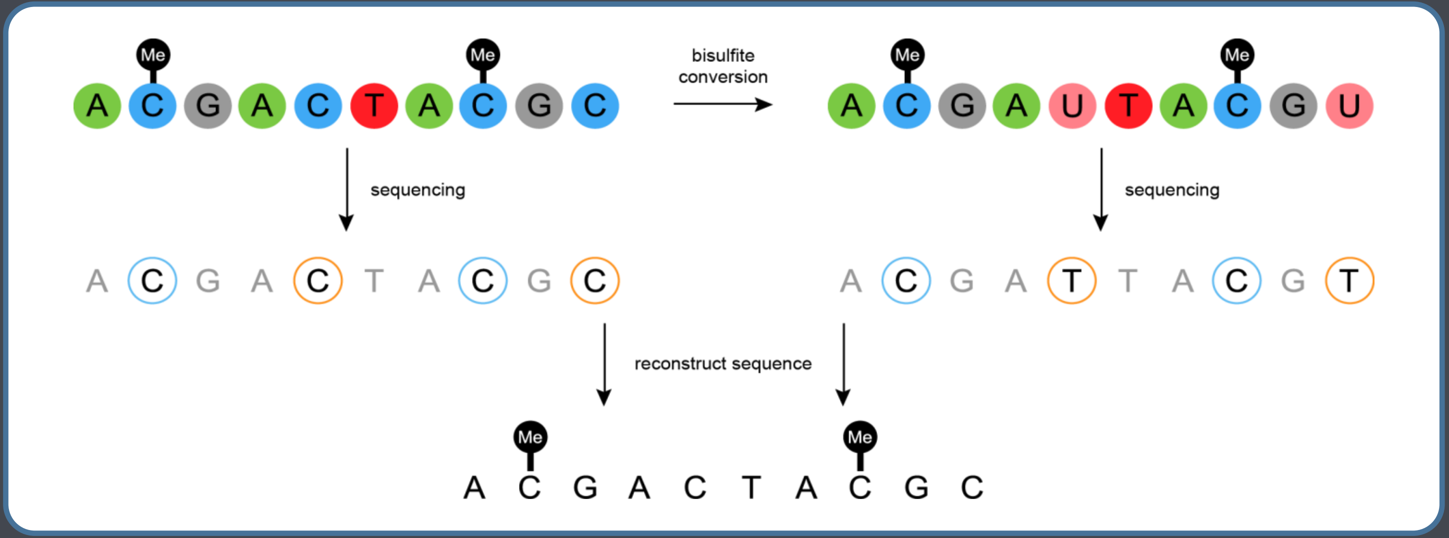


(Fry, 2011)

Bisulfite sequencing employed to detect methylation status of Cytosine

- Bisulfite treatment transforms unmethylated cytosine in uracil





MAKING A GENOME MANUAL

Scientists in the Encyclopedia of DNA Elements Consortium have applied 24 experiment types (across) to more than 150 cell lines (down) to assign functions to as many DNA regions as possible — but the project is still far from complete.

EXPERIMENTAL TARGETS

DNA methylation: regions layered with chemical methyl groups, which regulate gene expression.

Open chromatin: areas in which the DNA and proteins that make up chromatin are accessible to regulatory proteins.

RNA binding: positions where regulatory proteins attach to RNA.

RNA sequences: regions that are transcribed into RNA.

ChIP-seq: technique that reveals where proteins bind to DNA.

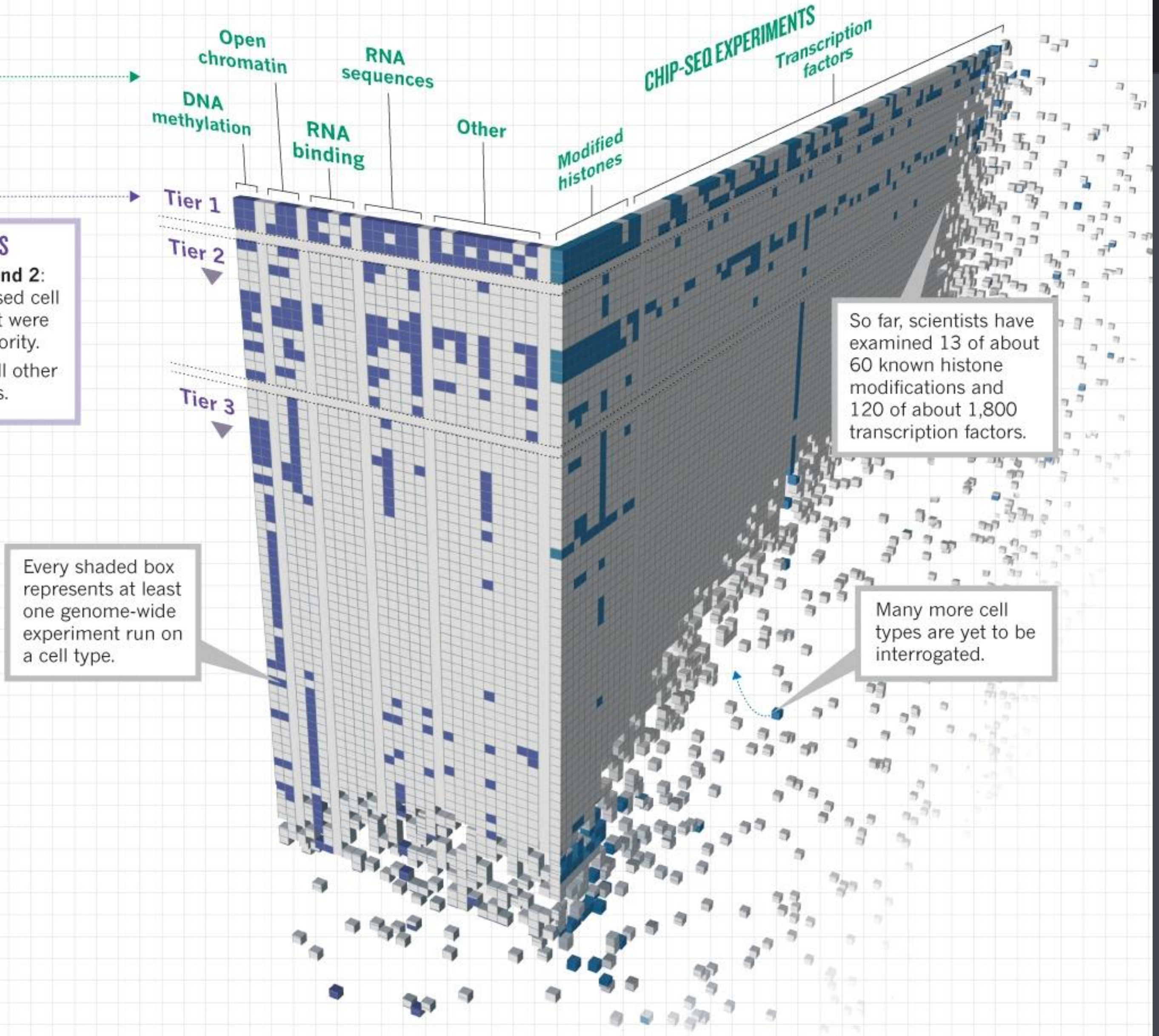
Modified histones: histone proteins, which package DNA into chromosomes, modified by chemical marks.

Transcription factors: proteins that bind to DNA and regulate transcription.

CELL LINES

Tiers 1 and 2: widely used cell lines that were given priority.

Tier 3: all other cell types.



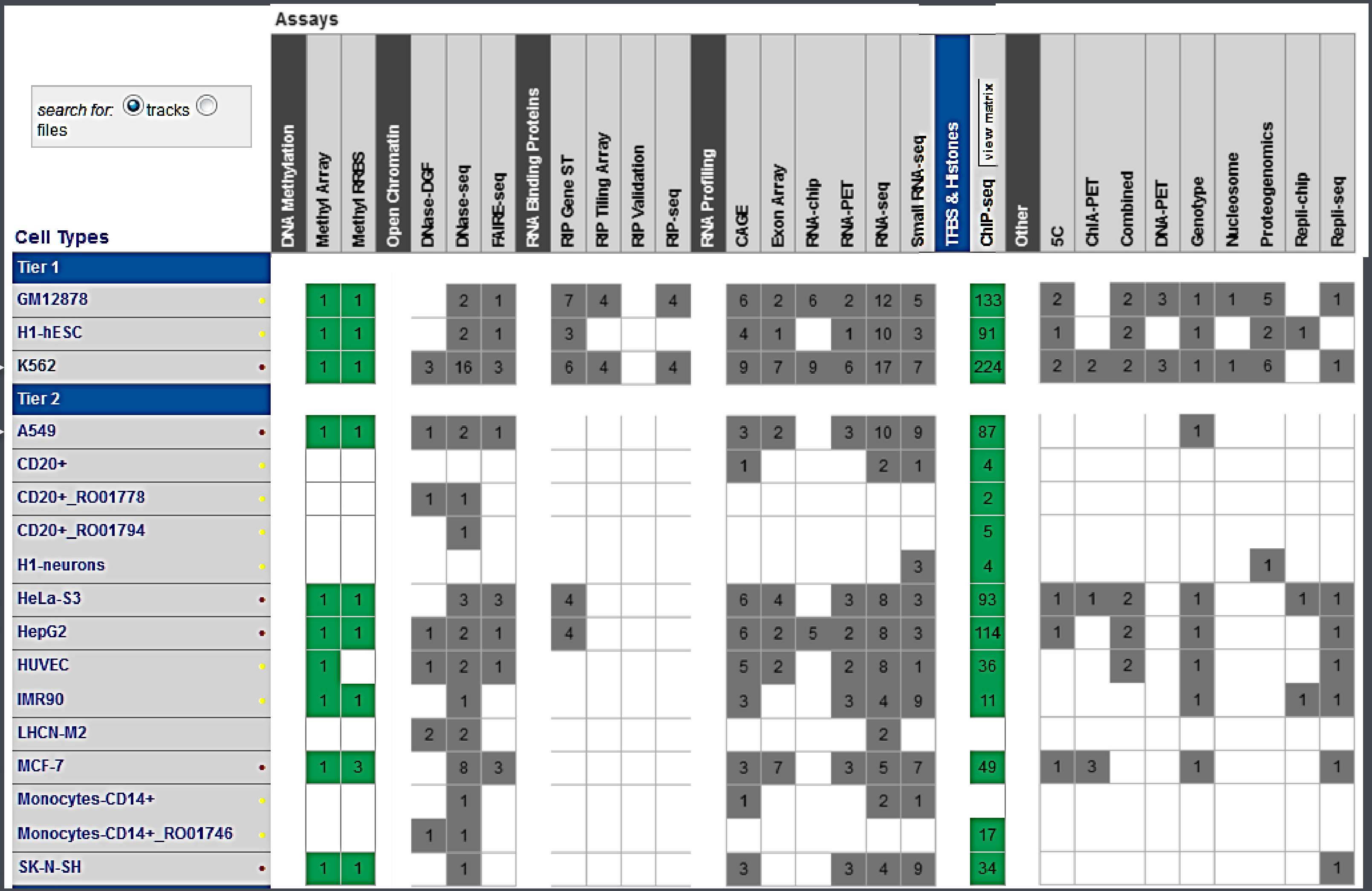
Every shaded box represents at least one genome-wide experiment run on a cell type.

So far, scientists have examined 13 of about 60 known histone modifications and 120 of about 1,800 transcription factors.

Many more cell types are yet to be interrogated.

B. Maher, Nature, 2012

ENCODE project includes data for normal and cancer cell lines



Assays

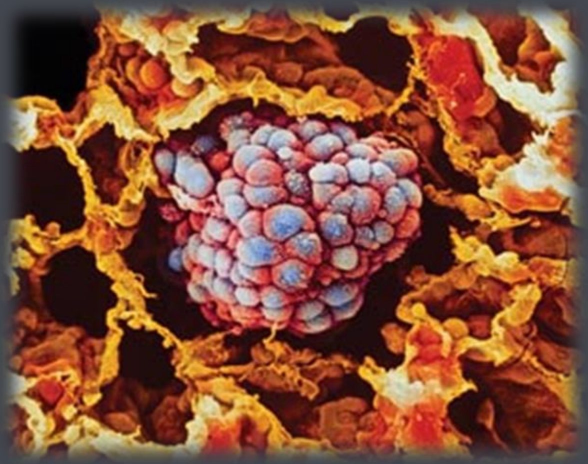
search for: ● tracks ○ files

Cell Types	DNA Methylation	Methyl Array	Methyl RRBS	Open Chromatin	DNase-DGF	DNase-seq	FAIRE-seq	RNA Binding Proteins	RIP Gene ST	RIP Tiling Array	RIP Validation	RIP-seq	RNA Profiling	CAGE	Exon Array	RNA-chip	RNA-PET	RNA-seq	Small RNA-seq	TFBS & Histones	ChIP-seq view matrix	Other	5C	ChIA-PET	Combined	DNA-PET	Genotype	Nucleosome	Proteogenomics	Repli-chip	Repli-seq
Tier 1																															
GM12878		1	1			2	1		7	4		4		6	2	6	2	12	5		133		2		2	3	1	1	5		1
H1-hESC		1	1			2	1		3					4	1		1	10	3		91		1		2		1		2	1	
K562		1	1		3	16	3		6	4		4		9	7	9	6	17	7		224		2	2	2	3	1	1	6		1
Tier 2																															
A549		1	1		1	2	1							3	2		3	10	9		87						1				
CD20+														1				2	1		4										
CD20+_RO01778					1	1															2										
CD20+_RO01794						1															5										
H1-neurons																			3		4								1		
HeLa-S3		1	1			3	3		4					6	4		3	8	3		93		1	1	2		1			1	1
HepG2		1	1		1	2	1		4					6	2	5	2	8	3		114		1		2		1				1
HUVEC		1			1	2	1							5	2		2	8	1		36				2		1				1
IMR90		1	1			1								3			3	4	9		11						1			1	1
LHCN-M2					2	2												2													
MCF-7		1	3			8	3							3	7		3	5	7		49		1	3			1				1
Monocytes-CD14+						1								1				2	1												
Monocytes-CD14+_RO01746					1	1															17										
SK-N-SH		1	1			1								3			3	4	9		34										1

Cancer cell line

What is going on with epigenetic profiles in cancer?

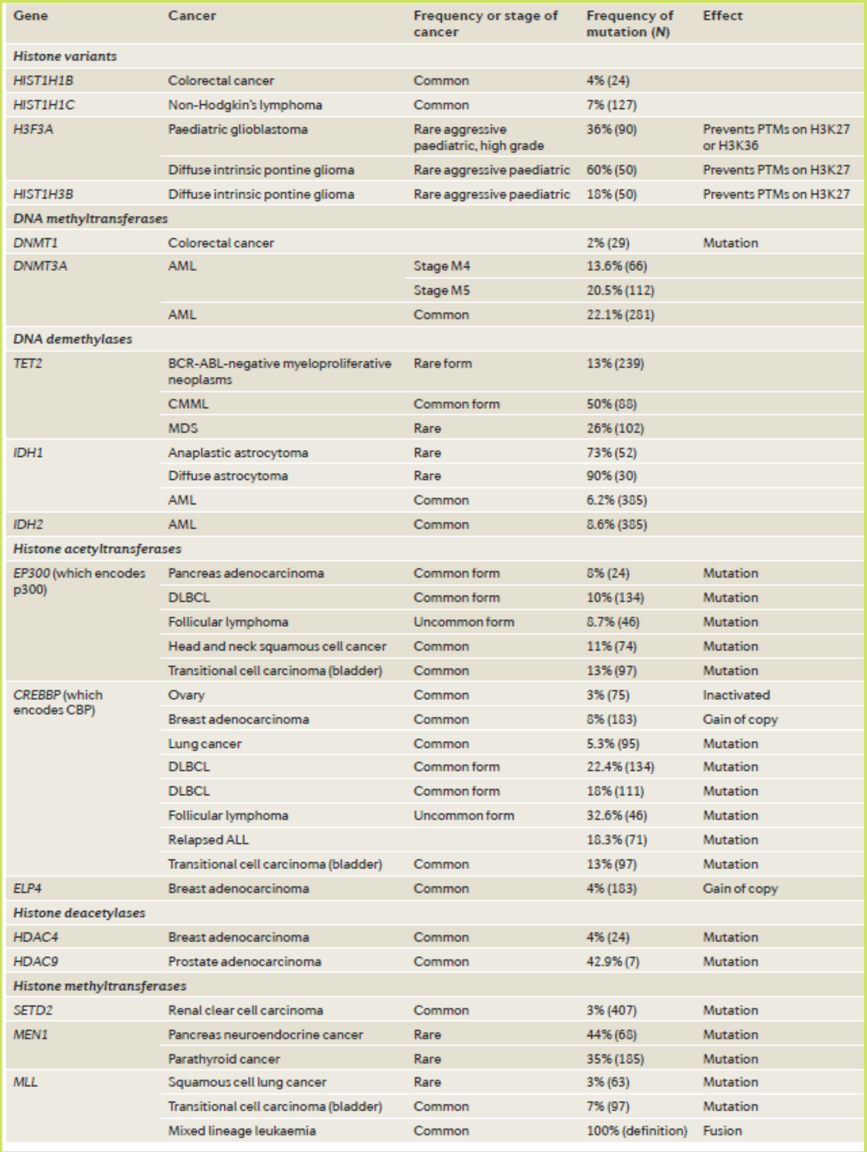
- Do epigenetic states change compared to normal ancestral cells?
- Is there any global phenomenon related to epigenetics inherent to cancer cells?



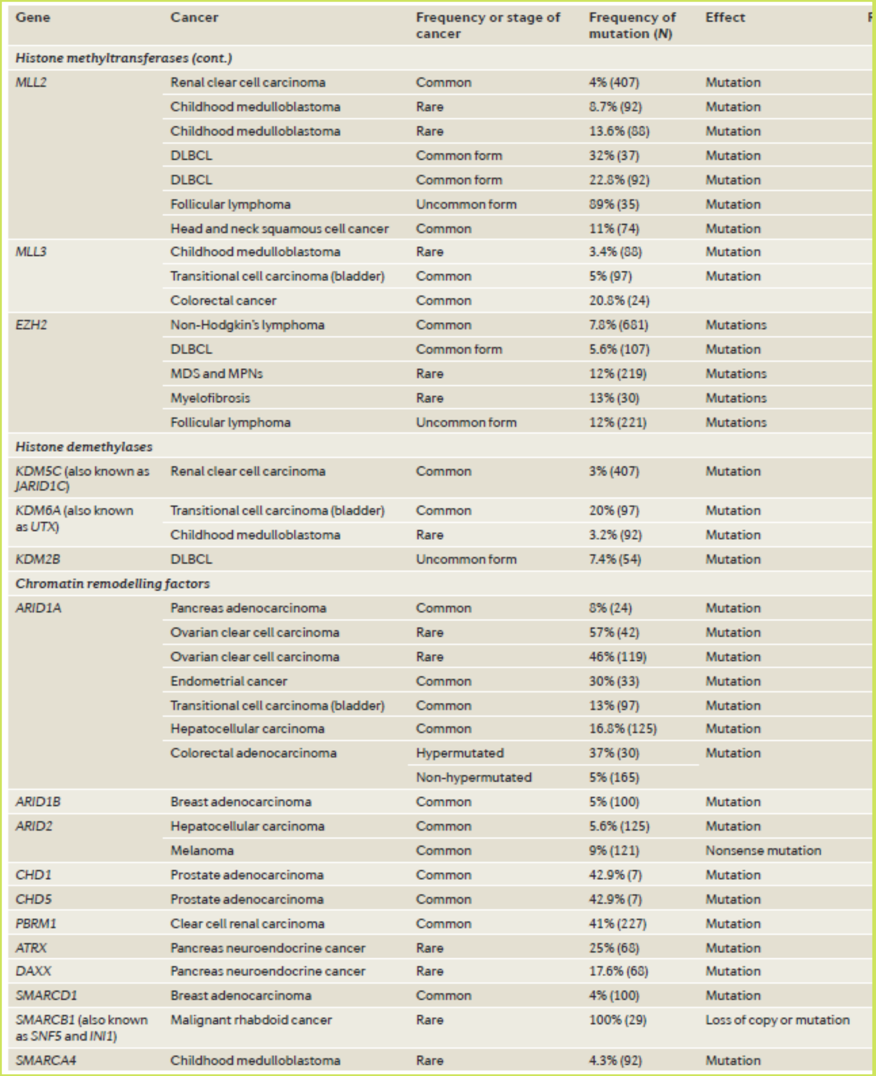
Lung cancer close-up. MOREDUN ANIMAL HEALTH LTD/SPL / Gettyimages

Histone and CpG-methyl modifying proteins are often mutated or deleted in cancer

Epigenome-modifying gene mutations in human cancer



Gene	Cancer	Frequency or stage of cancer	Frequency of mutation (*N*)	Effect
Histone variants				
HIST1H1B	Colorectal cancer	Common	4% (24)	
HIST1H1C	Non-Hodgkin's lymphoma	Common	7% (127)	
H3F3A	Paediatric glioblastoma	Rare aggressive paediatric, high grade	36% (90)	Prevents PTMs on H3K27 or H3K36
	Diffuse intrinsic pontine glioma	Rare aggressive paediatric	60% (50)	Prevents PTMs on H3K27
HIST1H3B	Diffuse intrinsic pontine glioma	Rare aggressive paediatric	18% (50)	Prevents PTMs on H3K27
DNA methyltransferases				
DNMT1	Colorectal cancer		2% (29)	Mutation
DNMT3A	AML	Stage M4	13.6% (66)	
		Stage M5	20.5% (112)	
	AML	Common	22.1% (281)	
DNA demethylases				
TET2	BCR-ABL-negative myeloproliferative neoplasms	Rare form	13% (239)	
	CMML	Common form	50% (88)	
	MDS	Rare	26% (102)	
IDH1	Anaplastic astrocytoma	Rare	73% (52)	
	Diffuse astrocytoma	Rare	90% (30)	
	AML	Common	6.2% (385)	
IDH2	AML	Common	8.6% (385)	
Histone acetyltransferases				
EP300 (which encodes p300)	Pancreas adenocarcinoma	Common form	8% (24)	Mutation
	DLBCL	Common form	10% (134)	Mutation
	Follicular lymphoma	Uncommon form	8.7% (46)	Mutation
	Head and neck squamous cell cancer	Common	11% (74)	Mutation
	Transitional cell carcinoma (bladder)	Common	13% (97)	Mutation
CREBBP (which encodes CBP)	Ovary	Common	3% (75)	Inactivated
	Breast adenocarcinoma	Common	8% (183)	Gain of copy
	Lung cancer	Common	5.3% (95)	Mutation
	DLBCL	Common form	22.4% (134)	Mutation
	DLBCL	Common form	18% (111)	Mutation
	Follicular lymphoma	Uncommon form	32.6% (46)	Mutation
	Relapsed ALL		18.3% (71)	Mutation
	Transitional cell carcinoma (bladder)	Common	13% (97)	Mutation
ELP4	Breast adenocarcinoma	Common	4% (183)	Gain of copy
Histone deacetylases				
HDAC4	Breast adenocarcinoma	Common	4% (24)	Mutation
HDAC9	Prostate adenocarcinoma	Common	42.9% (7)	Mutation
Histone methyltransferases				
SETD2	Renal clear cell carcinoma	Common	3% (407)	Mutation
MEN1	Pancreas neuroendocrine cancer	Rare	44% (68)	Mutation
	Parathyroid cancer	Rare	35% (185)	Mutation
MLL	Squamous cell lung cancer	Rare	3% (63)	Mutation
	Transitional cell carcinoma (bladder)	Common	7% (97)	Mutation
	Mixed lineage leukaemia	Common	100% (definition)	Fusion

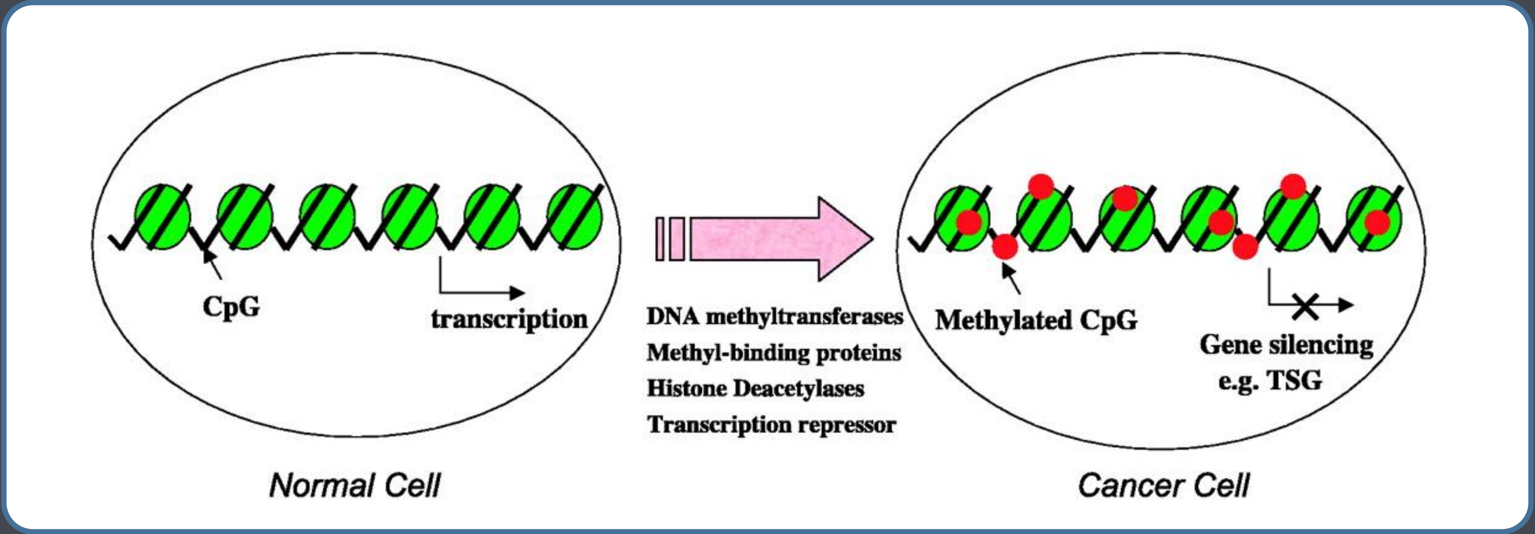


Gene	Cancer	Frequency or stage of cancer	Frequency of mutation (*N*)	Effect
Histone methyltransferases (cont.)				
MLL2	Renal clear cell carcinoma	Common	4% (407)	Mutation
	Childhood medulloblastoma	Rare	8.7% (92)	Mutation
	Childhood medulloblastoma	Rare	13.6% (88)	Mutation
	DLBCL	Common form	32% (37)	Mutation
	DLBCL	Common form	22.8% (92)	Mutation
	Follicular lymphoma	Uncommon form	89% (35)	Mutation
	Head and neck squamous cell cancer	Common	11% (74)	Mutation
MLL3	Childhood medulloblastoma	Rare	3.4% (88)	Mutation
	Transitional cell carcinoma (bladder)	Common	5% (97)	Mutation
	Colorectal cancer	Common	20.8% (24)	
EZH2	Non-Hodgkin's lymphoma	Common	7.8% (681)	Mutations
	DLBCL	Common form	5.6% (107)	Mutation
	MDS and MPNs	Rare	12% (219)	Mutations
	Myelofibrosis	Rare	13% (30)	Mutations
	Follicular lymphoma	Uncommon form	12% (221)	Mutations
Histone demethylases				
KDM5C (also known as *JARID1C*)	Renal clear cell carcinoma	Common	3% (407)	Mutation
KDM6A (also known as *UTX*)	Transitional cell carcinoma (bladder)	Common	20% (97)	Mutation
	Childhood medulloblastoma	Rare	3.2% (92)	Mutation
KDM2B	DLBCL	Uncommon form	7.4% (54)	Mutation
Chromatin remodelling factors				
ARID1A	Pancreas adenocarcinoma	Common	8% (24)	Mutation
	Ovarian clear cell carcinoma	Rare	57% (42)	Mutation
	Ovarian clear cell carcinoma	Rare	46% (119)	Mutation
	Endometrial cancer	Common	30% (33)	Mutation
	Transitional cell carcinoma (bladder)	Common	13% (97)	Mutation
	Hepatocellular carcinoma	Common	16.8% (125)	Mutation
	Colorectal adenocarcinoma	Hypermutated	37% (30)	Mutation
		Non-hypermutated	5% (165)	
ARID1B	Breast adenocarcinoma	Common	5% (100)	Mutation
ARID2	Hepatocellular carcinoma	Common	5.6% (125)	Mutation
	Melanoma	Common	9% (121)	Nonsense mutation
CHD1	Prostate adenocarcinoma	Common	42.9% (7)	Mutation
CHD5	Prostate adenocarcinoma	Common	42.9% (7)	Mutation
PBRM1	Clear cell renal carcinoma	Common	41% (227)	Mutation
ATRX	Pancreas neuroendocrine cancer	Rare	25% (68)	Mutation
DAXX	Pancreas neuroendocrine cancer	Rare	17.6% (68)	Mutation
SMARCD1	Breast adenocarcinoma	Common	4% (100)	Mutation
SMARCB1 (also known as *SNF5* and *INI1*)	Malignant rhabdoid cancer	Rare	100% (29)	Loss of copy or mutation
SMARCA4	Childhood medulloblastoma	Rare	4.3% (92)	Mutation

Timp & Feinberg, *Nature Rev. Cancer*, 2013

Changes in CpG methylation are common in cancer

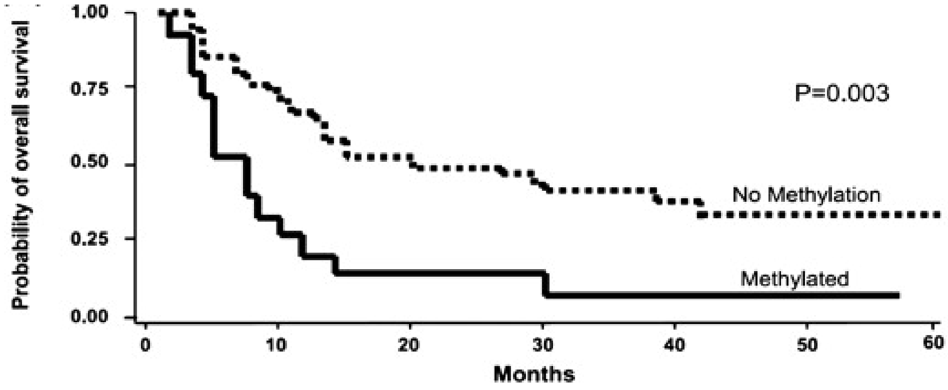
- Loss of imprinting (e.g. of *IGF2*)
- Hypermethylation of CpG islands of tumor suppressor genes
- Genome-wide DNA hypomethylation



DNA methylation status can be associated with tumor aggressiveness

ER-α methylation

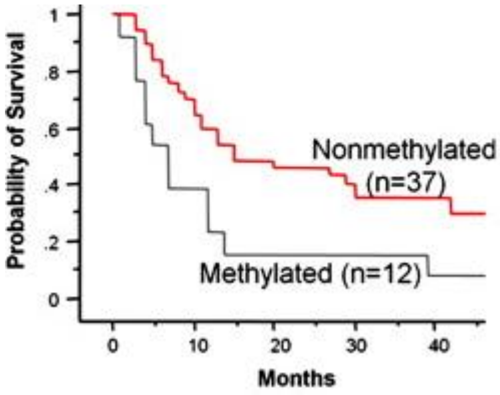
Skin cancer



Kaplan–Meier curves showing the correlation of pre-biochemotherapy serum ER-α methylation status with OS (p = 0.003)

RASSF1A methylation

Skin cancer

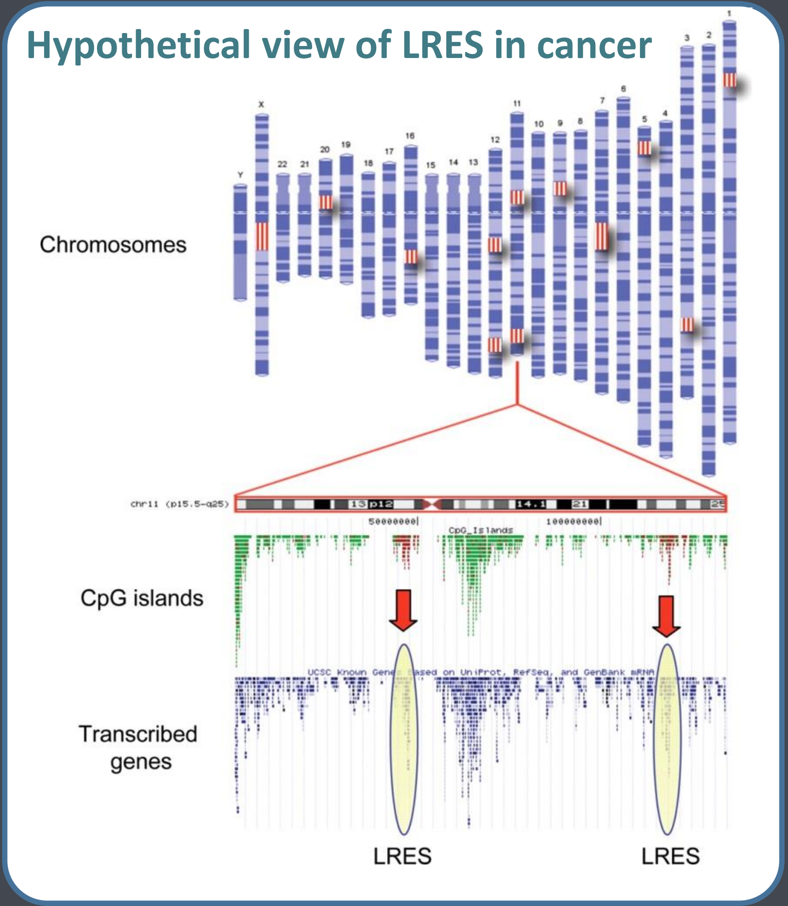


Kaplan–Meier survival curves of biochemotherapy patients: Correlation of pre-BC serum RASSF1A methylation BM with overall survival (p = .013).

Mori et al, 2006; From Mori et al, 2005

LRES & LOCKs: Global changes in epigenetic patterns in cancer

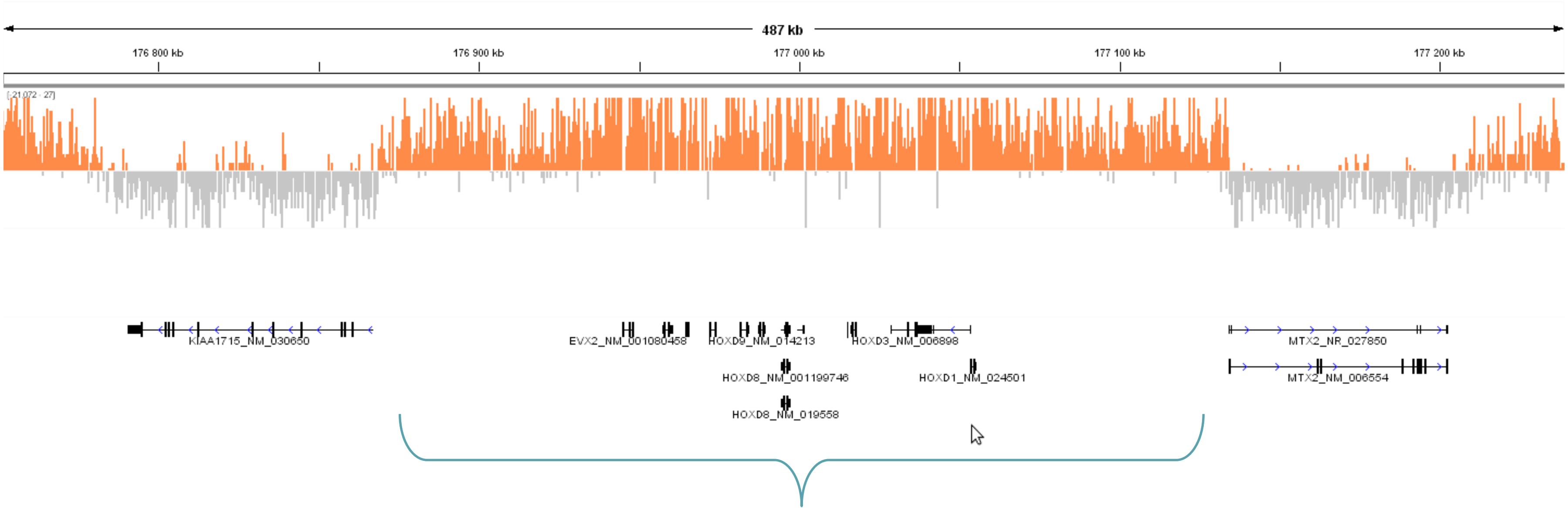
- Histone modification patterns are altered in human tumors
  - Long Range Epigenetic Silencing (LRES)
  - Large organized chromatin-lysine-(K) modifications (LOCKs)


Hypothetical view of LRES in cancer

S.J. Clark, *Hum. Mol. Genet.*, 2007

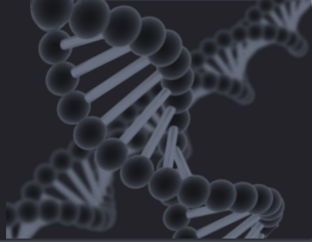
Example of LRES/LOCK of *HOXD* gene cluster in bladder cancer

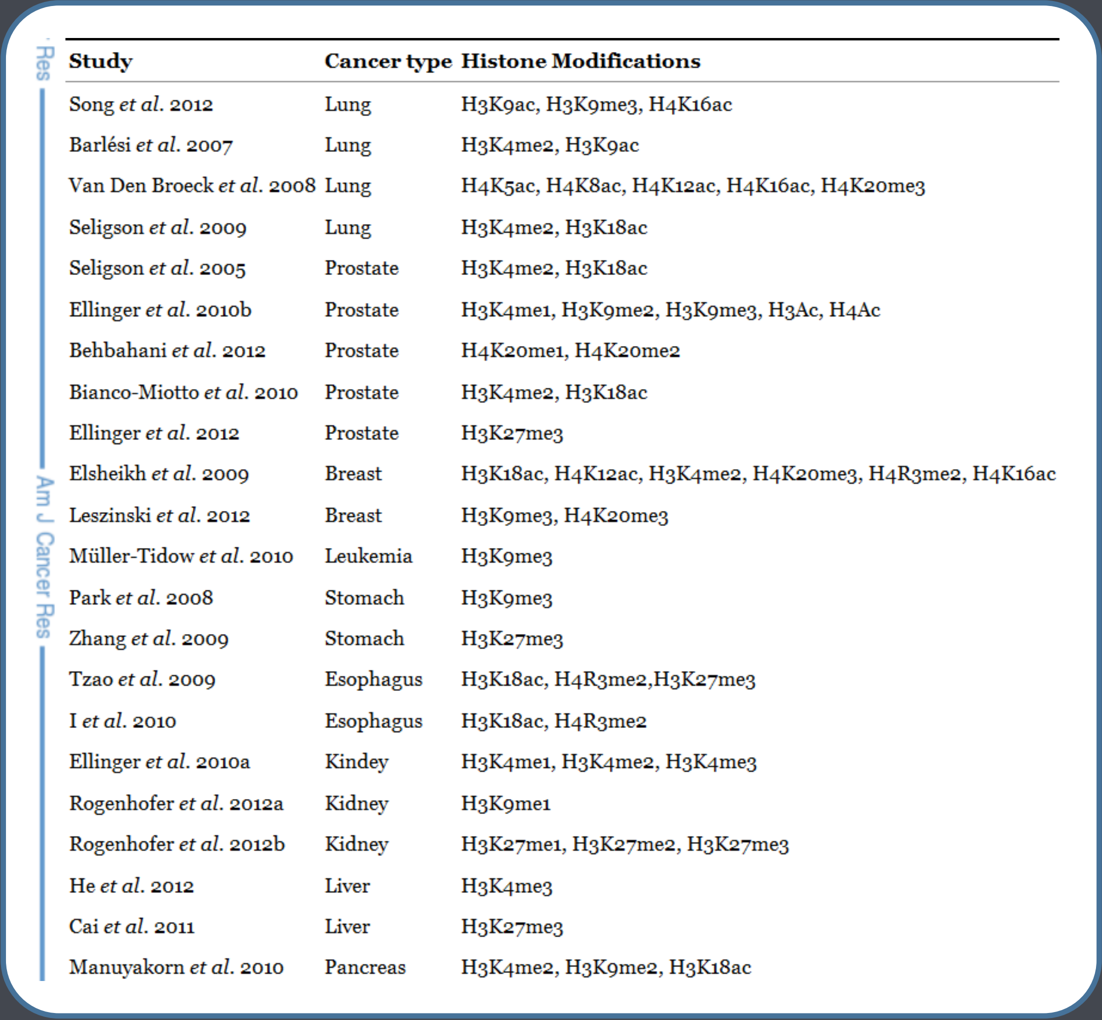
Enrichment in repressive histone mark H3K27me3



Cluster of HOXD genes repressed by epigenetic mechanisms

Histone modifications can predict the prognosis of various cancers



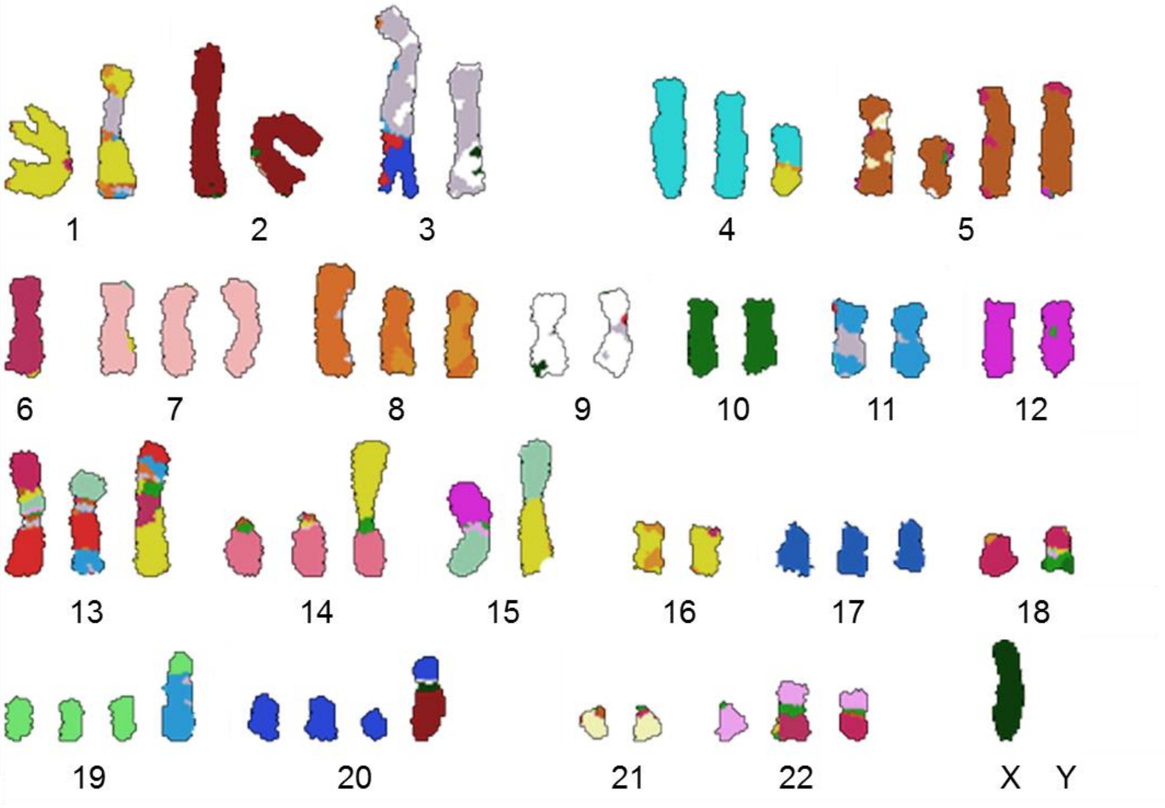


Study	Cancer type	Histone Modifications
Song *et al*. 2012	Lung	H3K9ac, H3K9me3, H4K16ac
Barlési *et al*. 2007	Lung	H3K4me2, H3K9ac
Van Den Broeck *et al*. 2008	Lung	H4K5ac, H4K8ac, H4K12ac, H4K16ac, H4K20me3
Seligson *et al*. 2009	Lung	H3K4me2, H3K18ac
Seligson *et al*. 2005	Prostate	H3K4me2, H3K18ac
Ellinger *et al*. 2010b	Prostate	H3K4me1, H3K9me2, H3K9me3, H3Ac, H4Ac
Behbahani *et al*. 2012	Prostate	H4K20me1, H4K20me2
Bianco-Miotto *et al*. 2010	Prostate	H3K4me2, H3K18ac
Ellinger *et al*. 2012	Prostate	H3K27me3
Elsheikh *et al*. 2009	Breast	H3K18ac, H4K12ac, H3K4me2, H4K20me3, H4R3me2, H4K16ac
Leszinski *et al*. 2012	Breast	H3K9me3, H4K20me3
Müller-Tidow *et al*. 2010	Leukemia	H3K9me3
Park *et al*. 2008	Stomach	H3K9me3
Zhang *et al*. 2009	Stomach	H3K27me3
Tzao *et al*. 2009	Esophagus	H3K18ac, H4R3me2,H3K27me3
I *et al*. 2010	Esophagus	H3K18ac, H4R3me2
Ellinger *et al*. 2010a	Kindey	H3K4me1, H3K4me2, H3K4me3
Rogenhofer *et al*. 2012a	Kidney	H3K9me1
Rogenhofer *et al*. 2012b	Kidney	H3K27me1, H3K27me2, H3K27me3
He *et al*. 2012	Liver	H3K4me3
Cai *et al*. 2011	Liver	H3K27me3
Manuyakorn *et al*. 2010	Pancreas	H3K4me2, H3K9me2, H3K18ac

Chervona & Costa, *Am. J. Cancer Res.*, 2012

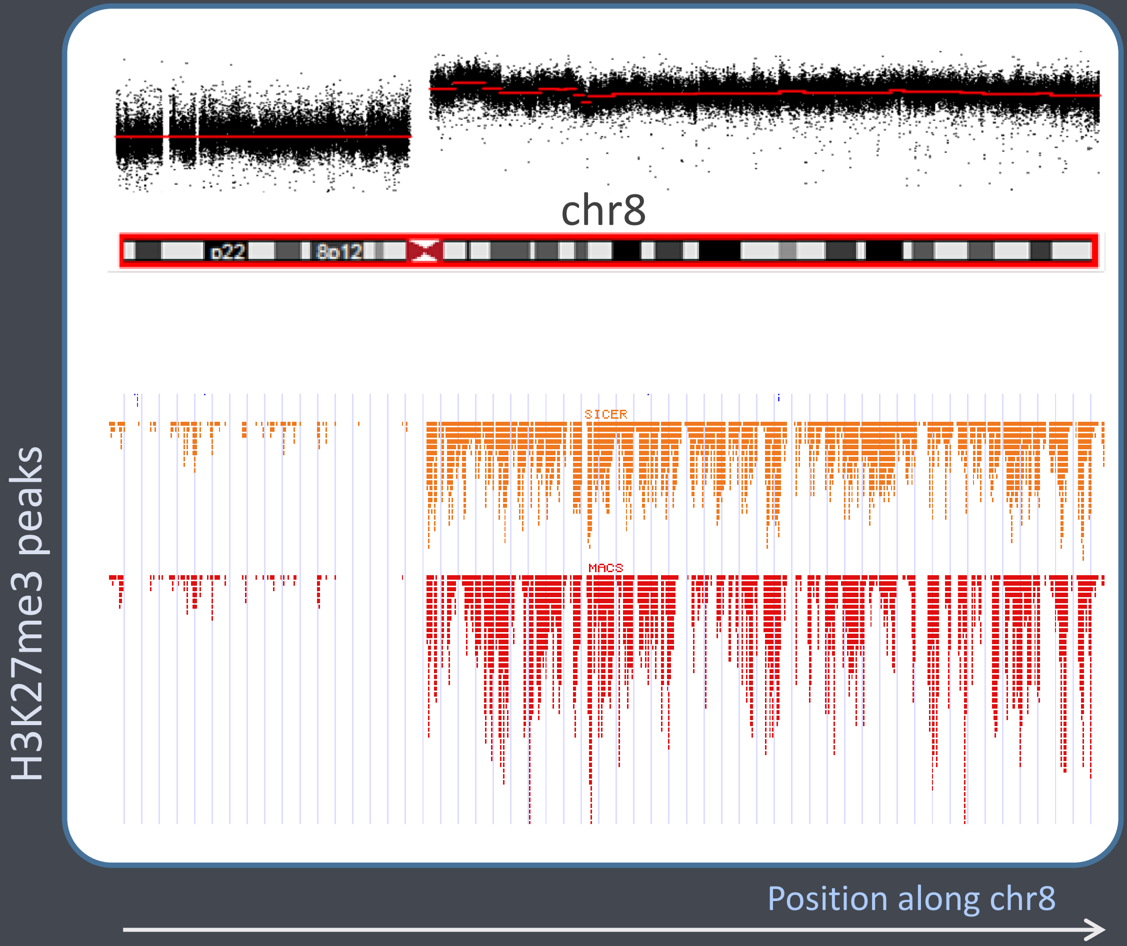
One should apply specific methods to detect histone modifications in cancer

- Specific feature of cancer samples: large copy number changes



Lung Adenocarcinoma

Standard methods for signal detection can miss signal in regions of loss in cancer



Copy number profile

Peaks predicted by tools:

MACS

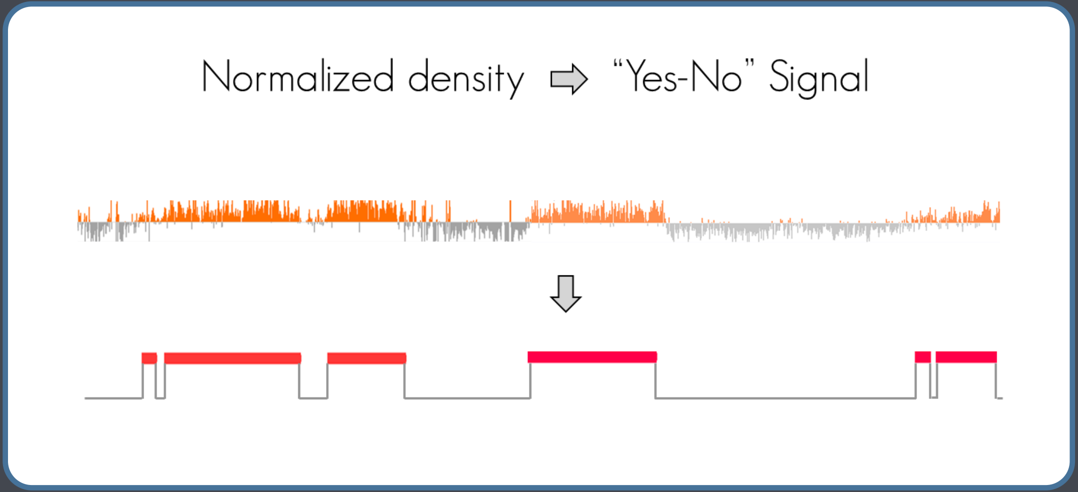
Zhang,Y. *et al.* (2008) *Genome Biol.*, **9**, R137

SICER

Zang,C. *et al.* (2009) *Bioinformatics*, **25**, 1952–1958.

Solution: explicit normalization for copy number status

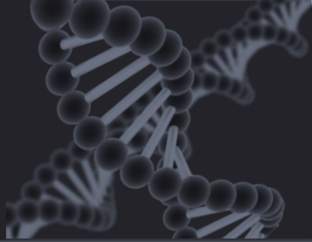
- Hidden Markov model after correction of ChIP-seq signal for copy number and GC-content bias

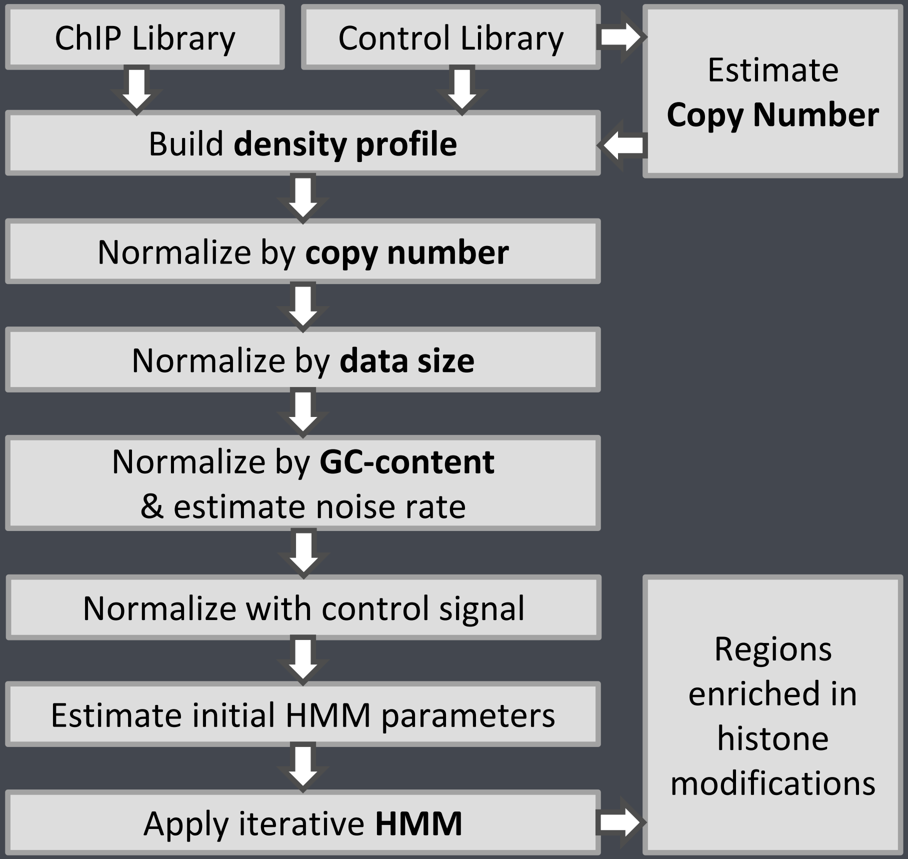


Software:
HMCan
www.cbrc.kaust.edu.sa/hmcan

H. Ashoor et al, *Bioinformatics*, 2013

HMCan explicitly corrects for copy number bias and GC-content





ChIP Library

Control Library

Estimate **Copy Number**

Build **density profile**

Normalize by **copy number**

Normalize by **data size**

Normalize by **GC-content** & estimate noise rate

Normalize with control signal

Estimate initial HMM parameters

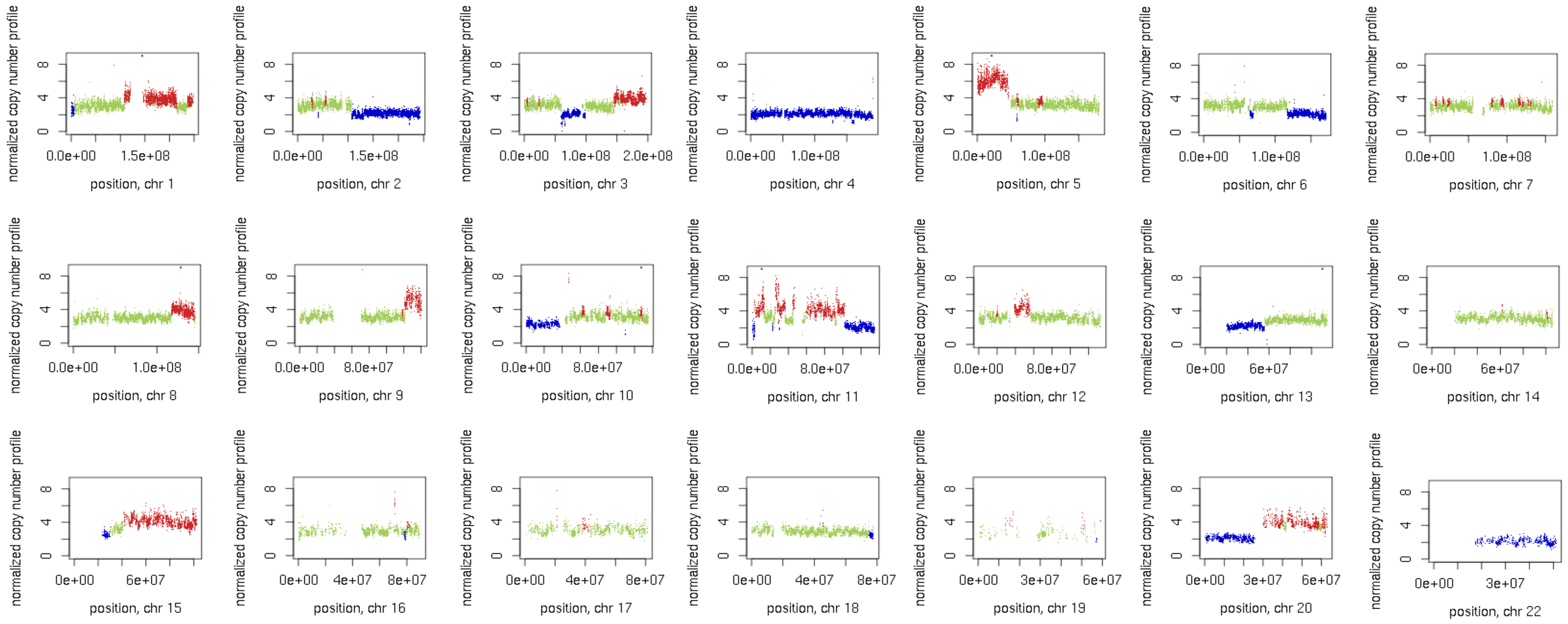
Apply iterative **HMM**

Regions enriched in histone modifications

H. Ashoor et al, *Bioinformatics*, 2013

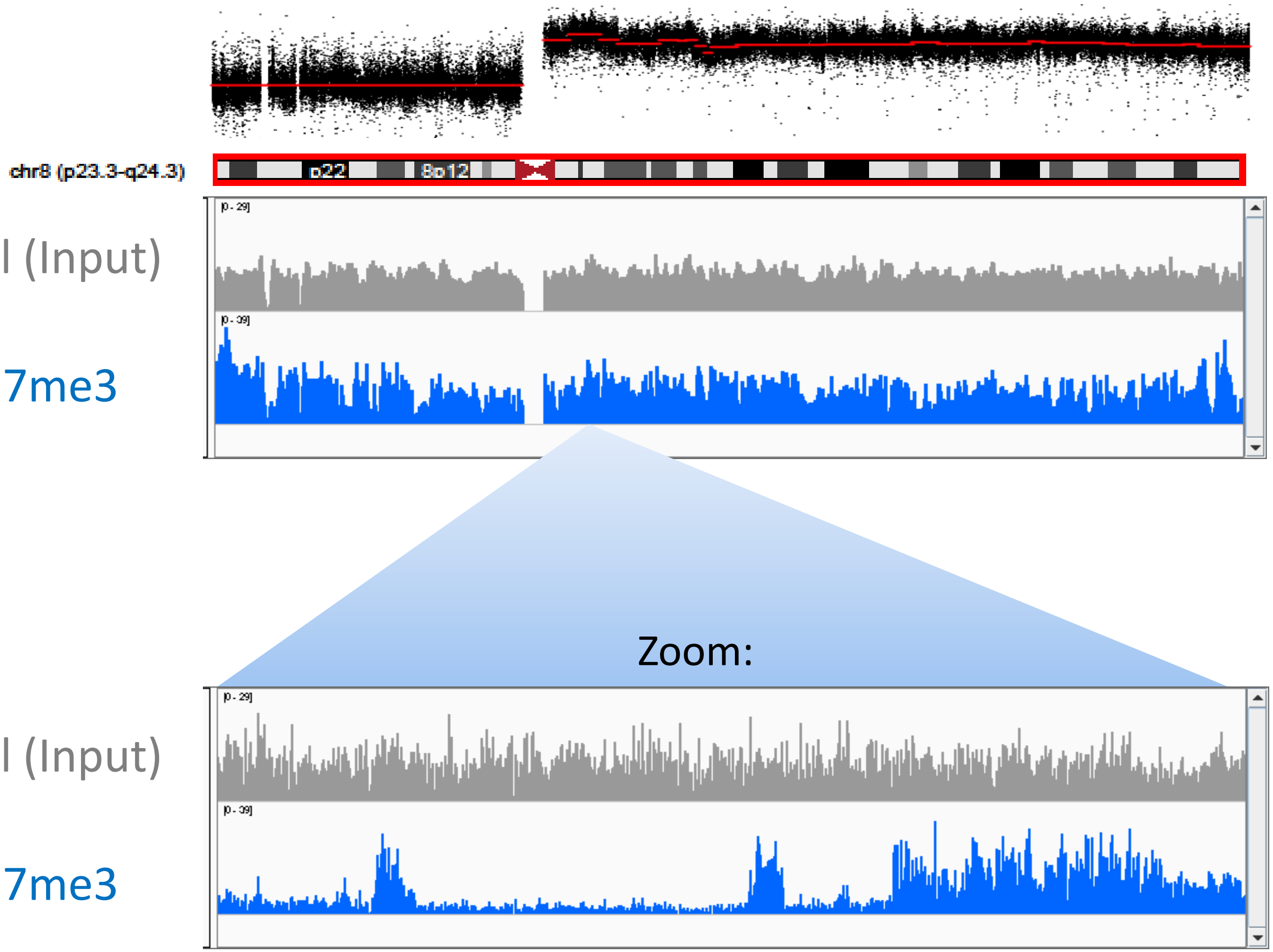
HMCan uses FREEC's algorithm for annotation of copy number alterations

Copy number profile for Hela-S3 cell line obtained using the Input data (ENCODE dataset)



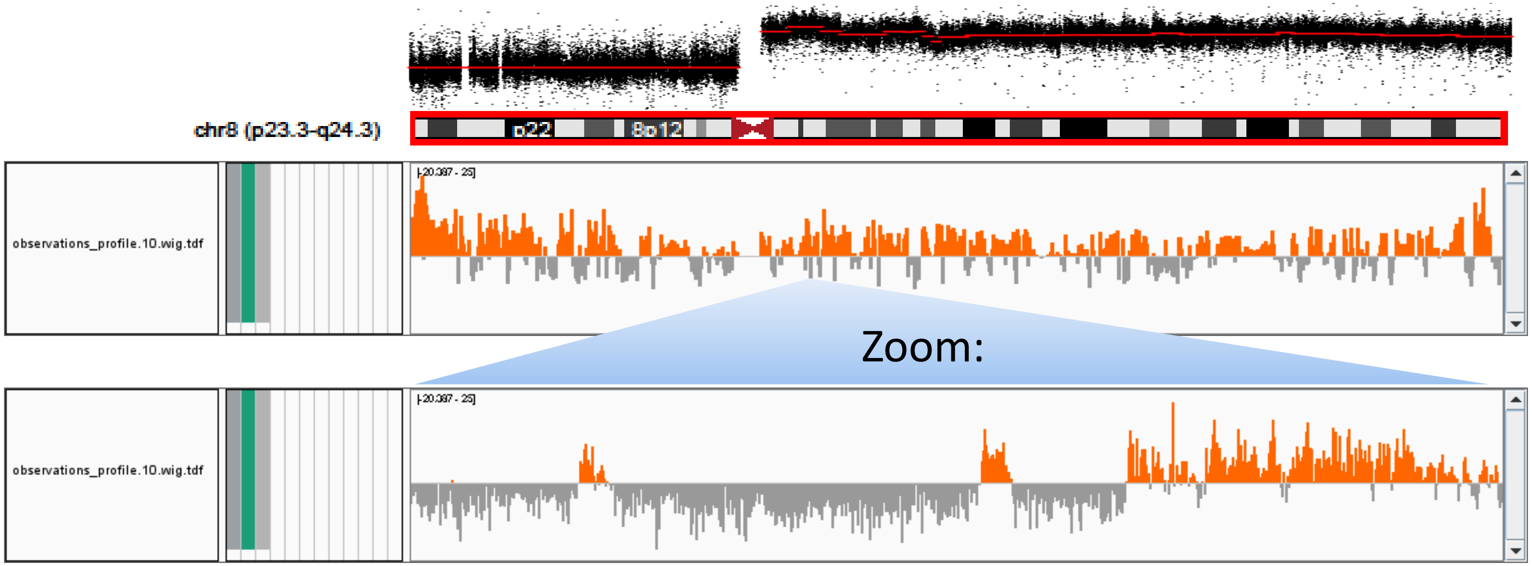
V. Boeva et al, *Bioinformatics*, 2011

Read densities profiles are corrected for copy number



H. Ashoor et al, *Bioinformatics*, 2013

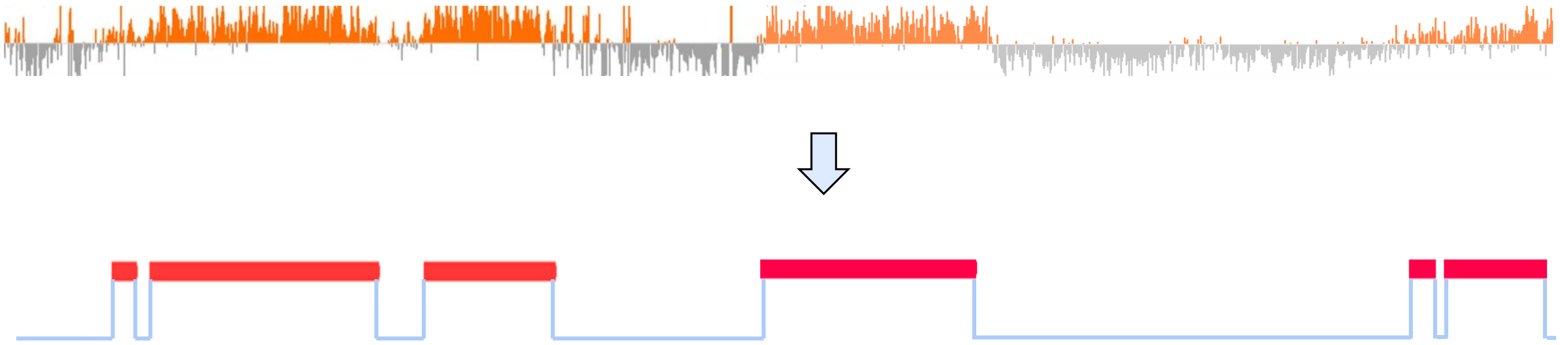
HMCan subtracts the control (input) density from the ChIP density



H. Ashoor et al, *Bioinformatics*, 2013

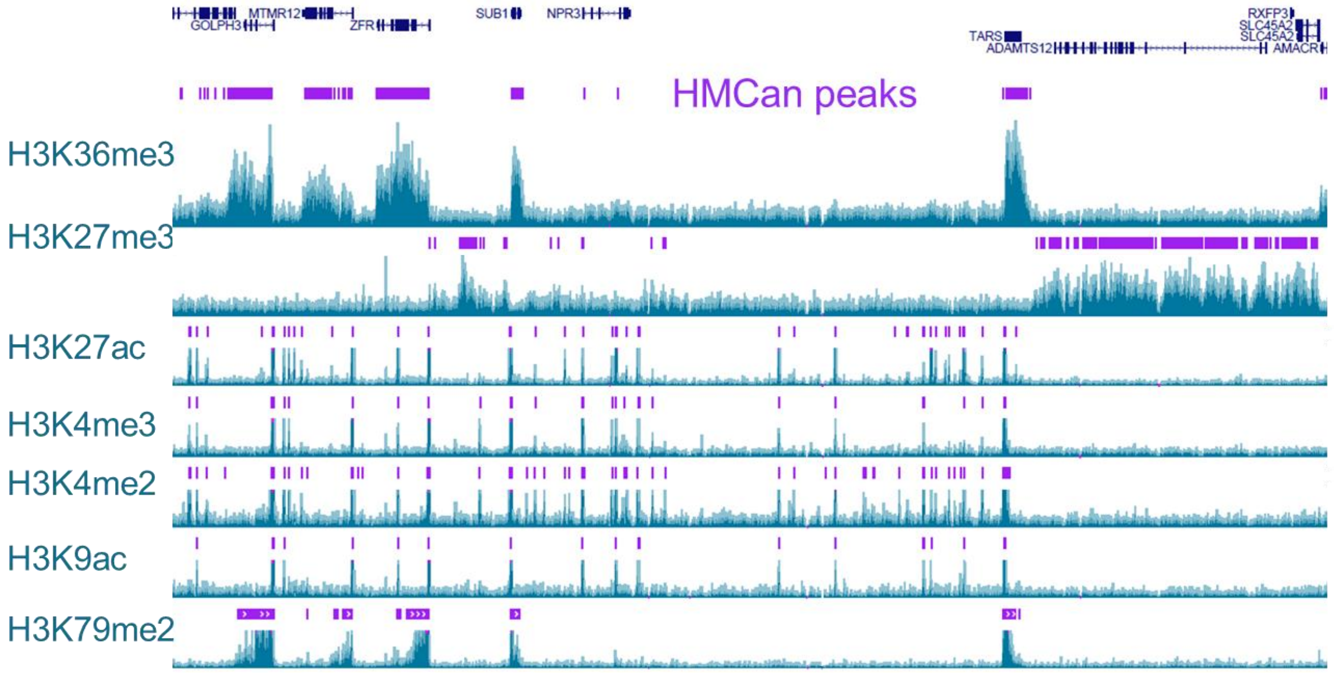
HMCan applies Hidden Markov Models to extract the signal

Normalized density ⇨ "Yes-No" Signal



H. Ashoor et al, *Bioinformatics*, 2013

Peaks predicted by HMCan for ENCODE data do not show copy number bias



H. Ashoor et al, *Bioinformatics*, 2013

Framework for the analysis of histone modification profiles & TFBSs nebula.curie.fr

- Nebula: web-service for analysis of ChIP-seq data

Welcome to Nebula

Nebula is a web service provided by **Institut Curie** and powered by **Galaxy** which allows users (Bioinformaticians as far as Biologists) to analyze their ChIP-seq data.

If you never used Galaxy and/or bioinformatics tools, please read the tutorial before starting your own analysis.

You can also download the Nebula's tools that are not in Galaxy by default and integrate them in your own instance to perform local analysis

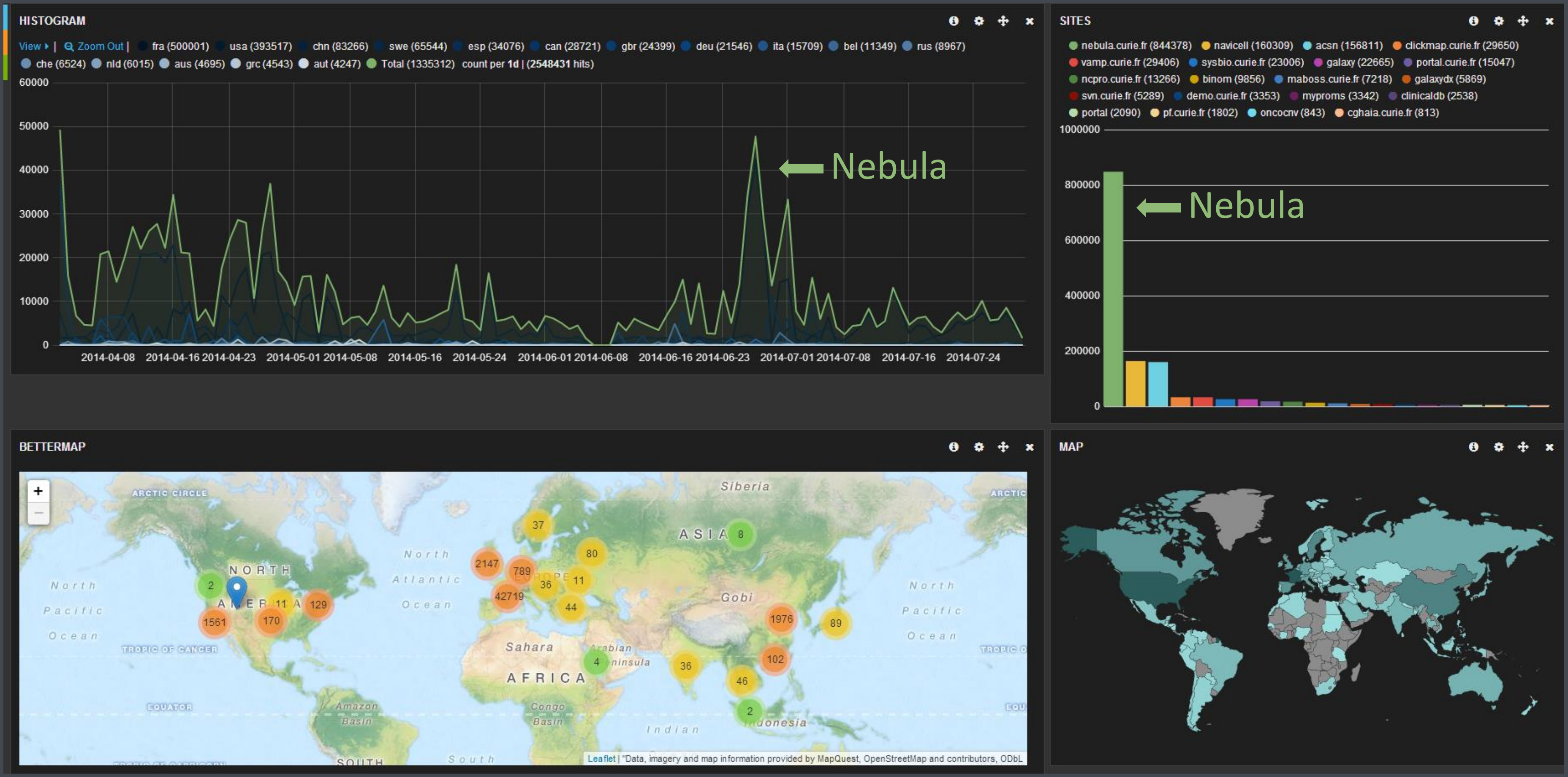
Tools are also available from the main Galaxy toolshed under sequence analysis category

For any troubleshootings, please contact: Valentina Boeva or Alban Lermine

V. Boeva, A. Lermine *et al*, *Bioinformatics*, 2012

Nebula: web-service for analysis of ChIP-seq data

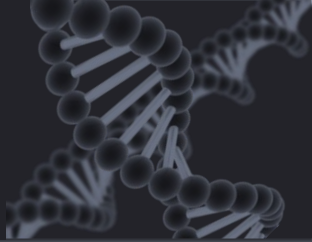
Statistics for external connections to Curie



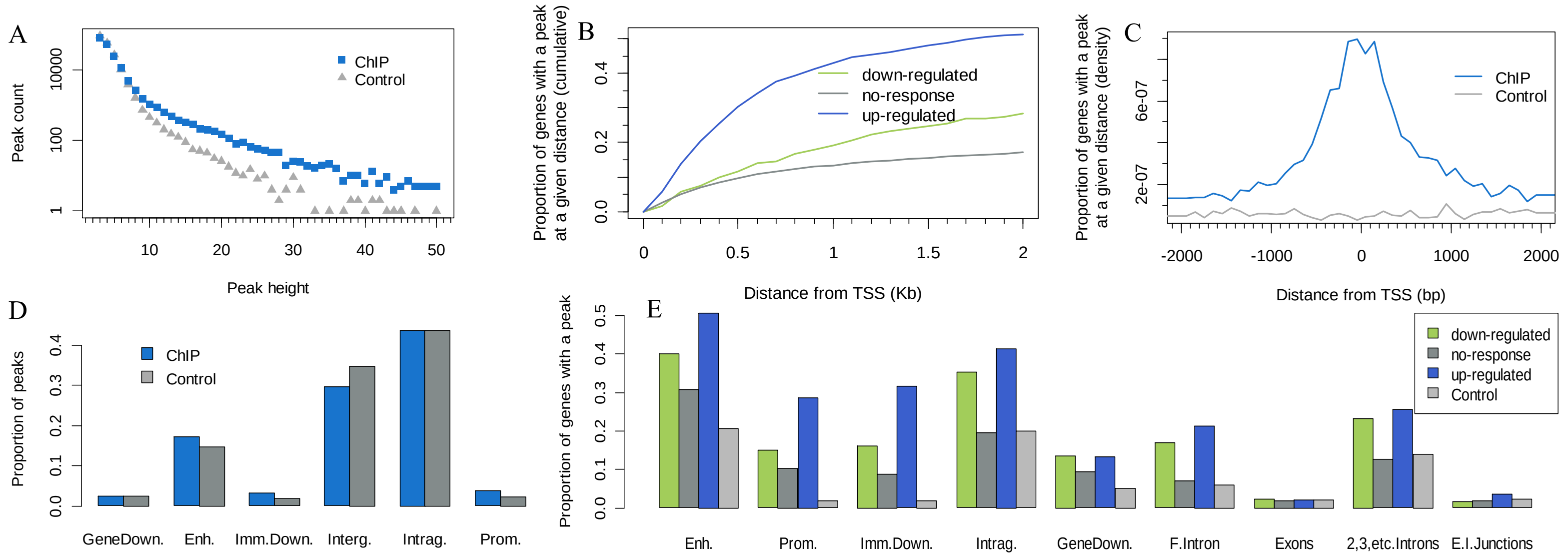
← Nebula

← Nebula

Nebula: web-service for analysis of ChIP-seq data



- Peak calling
- Calculation of the density and cumulative distribution of peak locations relative to gene transcription start sites
- Annotation of peaks with genomic features and genes with peak information

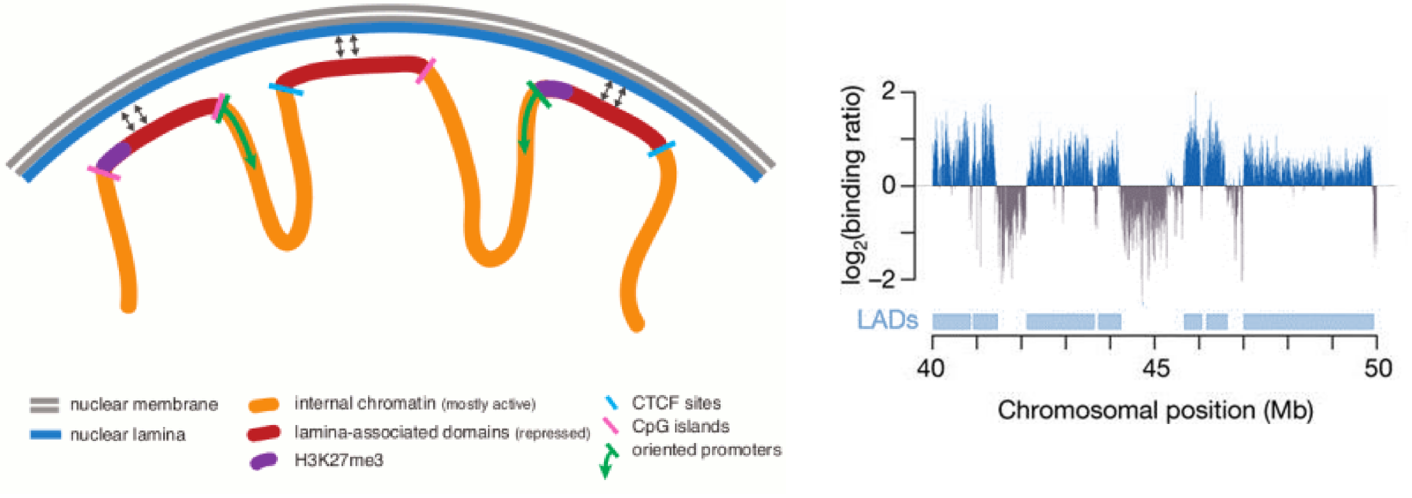


Some graphs produced produced by Nebula.

V. Boeva, A. Lermine *et al*, *Bioinformatics*, 2012

3D structure of the genome is related to the epigenetic profiles

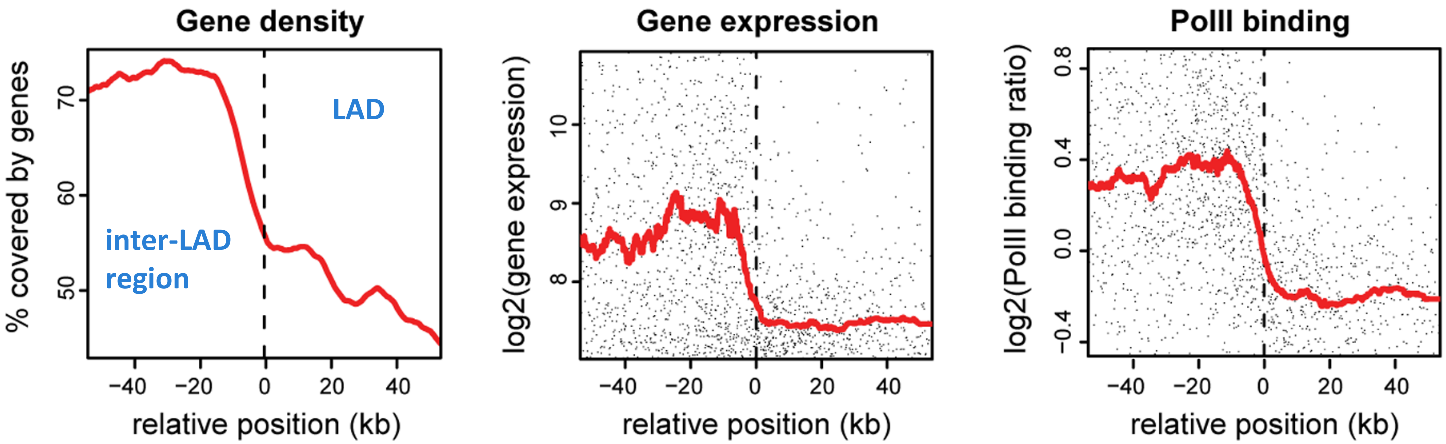
- Lamina associated domains (LAD) affect gene expression


Model of chromosome organization in interphase

Guelen et al., *Nature*, 2008

3D structure of the genome is related to the epigenetic profiles

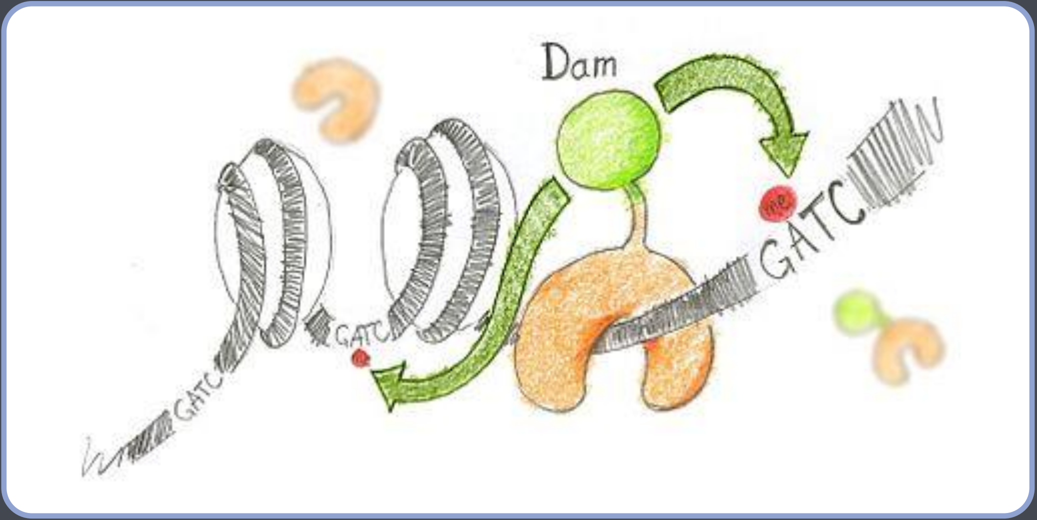
- Lamina associated domains (LAD) affect gene expression
- 82% of the H3K9Me2 LOCKs overlap with the locations of LADs
- H3K9 methyltransferase G9a is a regulator of NL contacts



J.G. van Bemmel et al., *PLoS ONE*, 2010
B. Wen et al., *Nat Genet.*, 2009
J.Kind et al., *Cell*, 2013

The technique DamID can be used to get genome-wide profiles for lamina binding

- The enzyme Dam (green) is fused to the protein of interest (orange - lamin) by expression of a chimeric DNA sequence.
- The protein of interest drags Dam onto its cognate targets.
- The tethering leads to methylation of GATCs in the neighborhood of the binding site (red) but not at a distance.

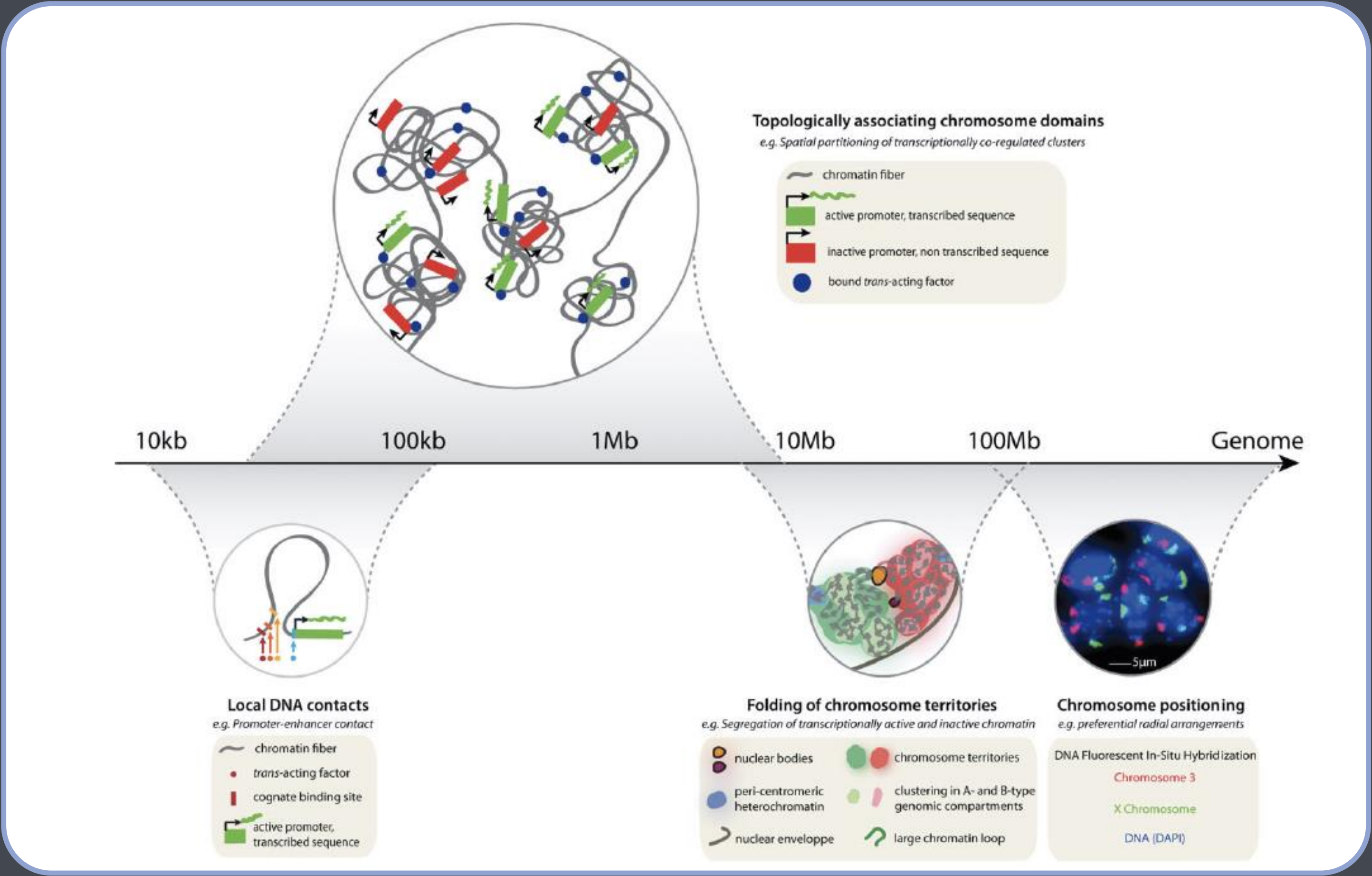


Methylation tag can be detected using restriction enzymes (DpnI and DpnII)

Wikipedia

3D structure of the genome defines regulatory interactions between promoters and enhancers

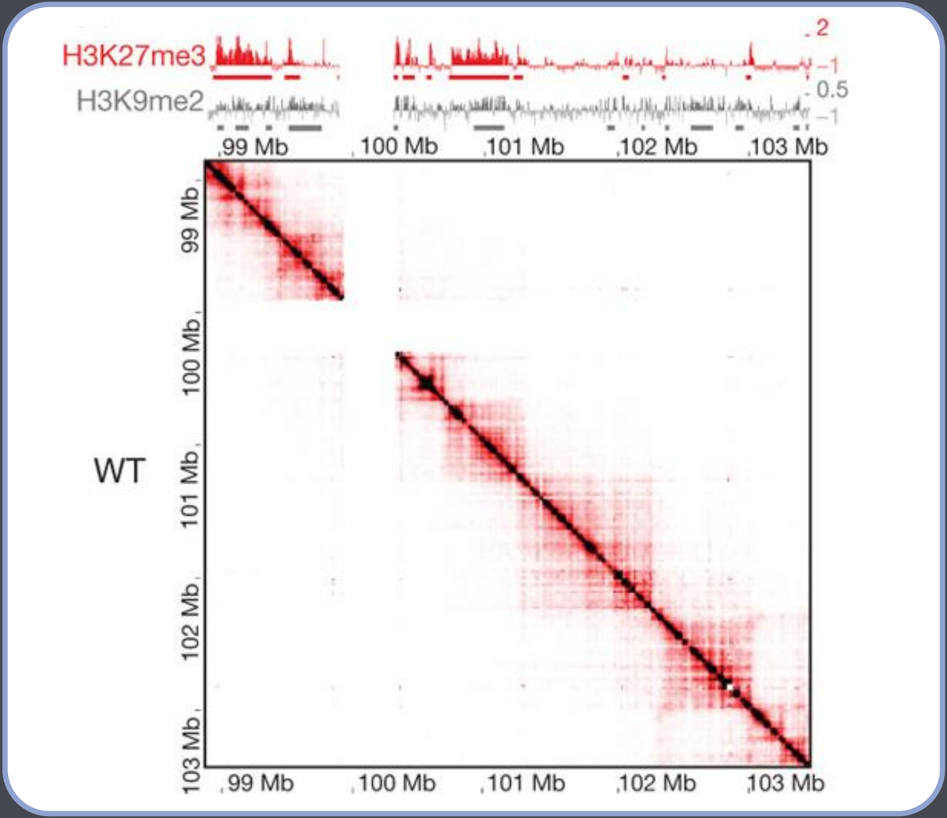
- Scales of genome architecture



E. P. Nora et al., *Bioessays*, 2013

3D structure of the genome defines regulatory interactions between promoters and enhancers

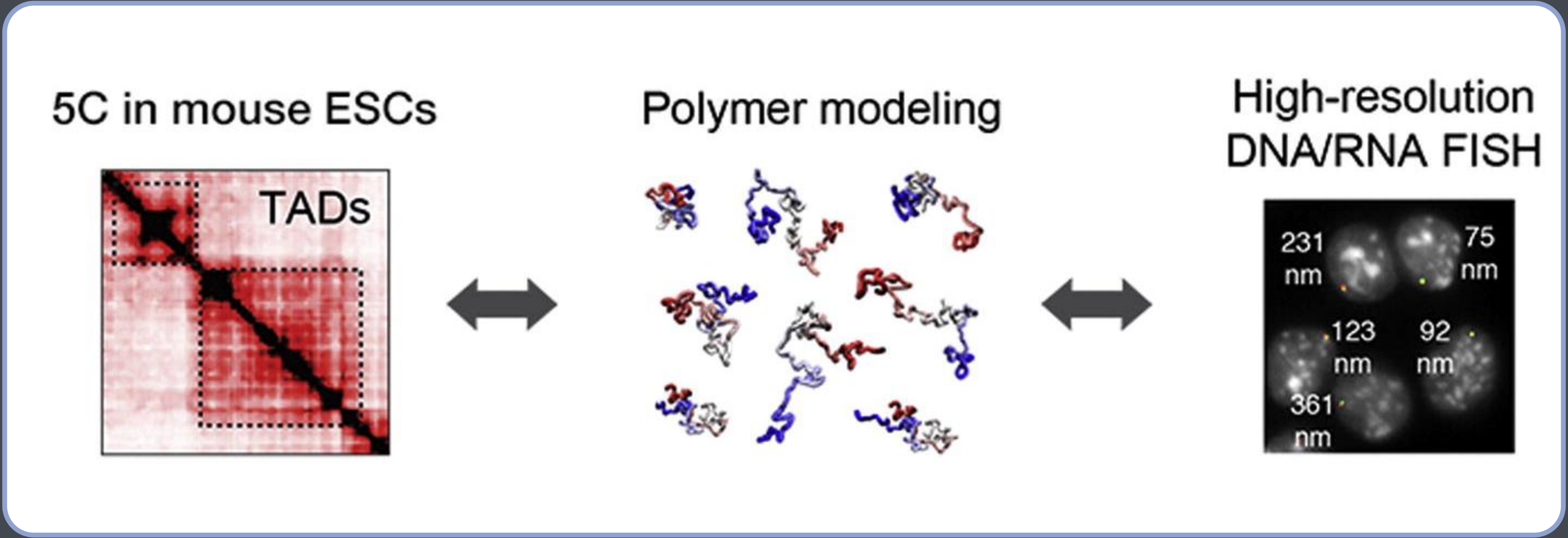
- 3C, 4C, 5-C and Hi-C methods allow mapping 3D interactions of chromatin



E. P. Nora et al., *Nature*, 2012

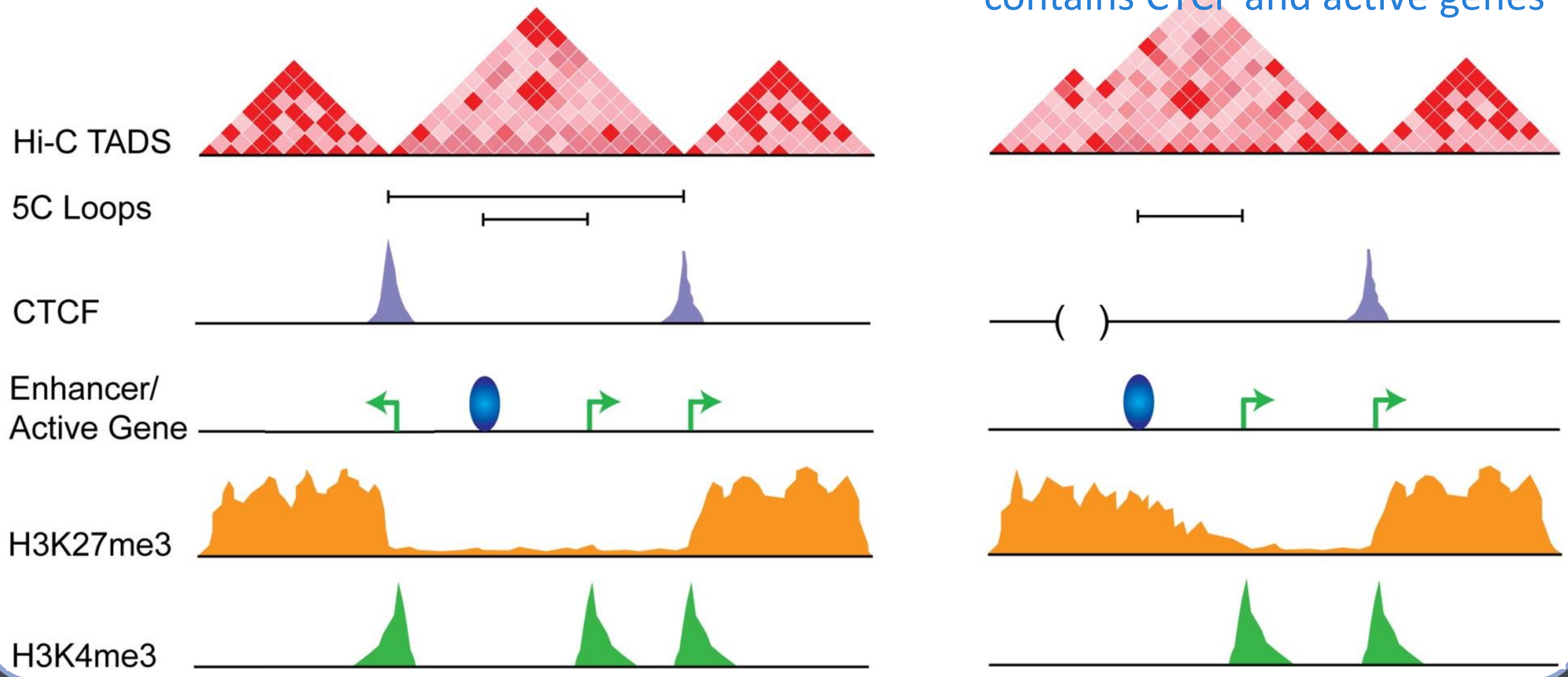
In some cases, it is even possible to reconsruct the 3D structure using hi-C intercations maps

- Physical Modeling of the Chromatin Fiber



Giorgetti et al., *Cell*, 2014

Disruption of genomic sequence can lead to the disruption of TAD and changes in the regulation of gene expression

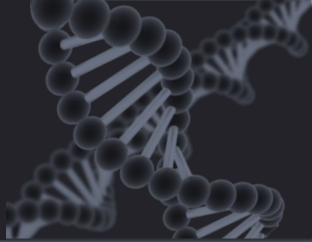


CTCF -mediated TAD boundaries and intra-TAD genomic organization via long-range interactions

Blurring of TAD boundaries after deletion of a 50–80 kb boundary between TADs that contains CTCF and active genes

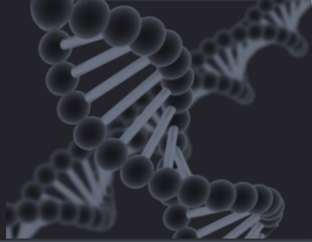
J.E. Phillips-Cremins & V.G. Corcesemail, *Mol. Cell*, 2013

Disruption of the genomic sequence in cancer can affect epigenetic profiles



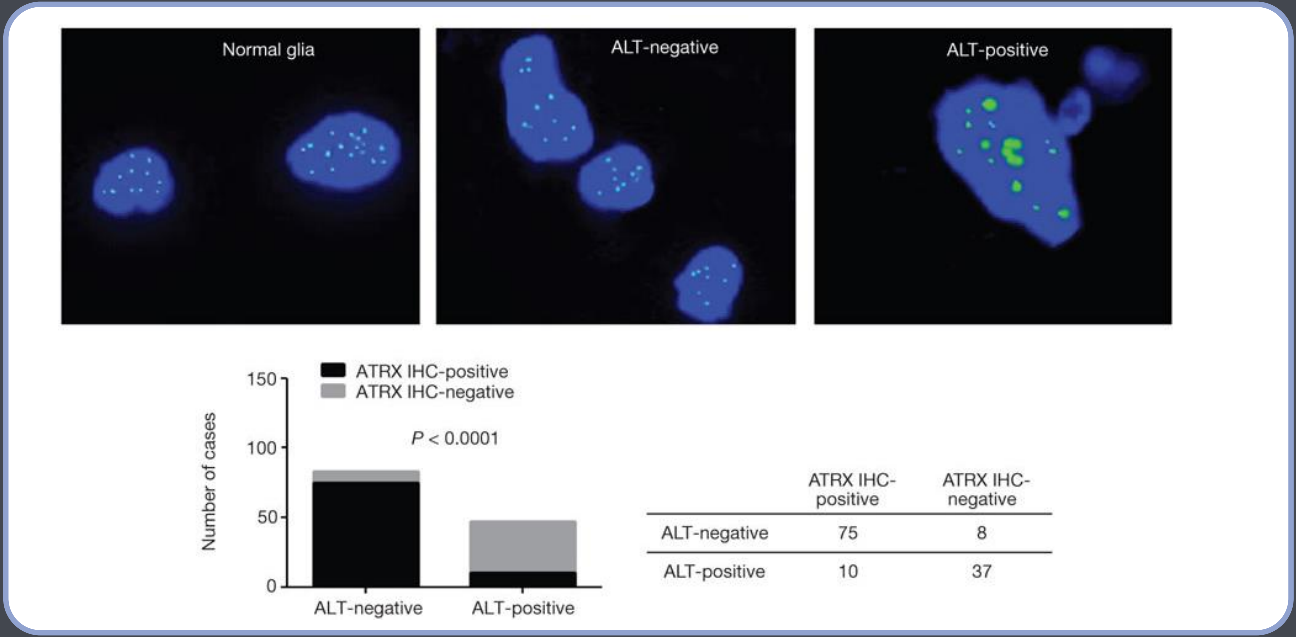
- Mutations and structural variants (SVs) in cancer genomes ➡
  - Disruption of epigenetic profiles by mutation of epigenome-regulatory proteins (readers, writers or erasers)
  - Disruption of regulatory elements
  - Disruption of TAD/LAD structure
  - Disruption of interactions between genes and regulatory elements

Changes in histone variant profiles can also happen in cancer



- Mutations in H3.3-ATRX-DAXX chromatin remodelling pathway were identified in 44% of glioblastomas (mostly high grade paediatric glioblastomas)
- Alternate lengthening of telomere (ALT) is associated with the presence of mutant H3F3A/ATRX

ATRX and *DAXX* encode two subunits of a chromatin remodelling complex required for H3.3 incorporation at pericentric heterochromatin and telomeres



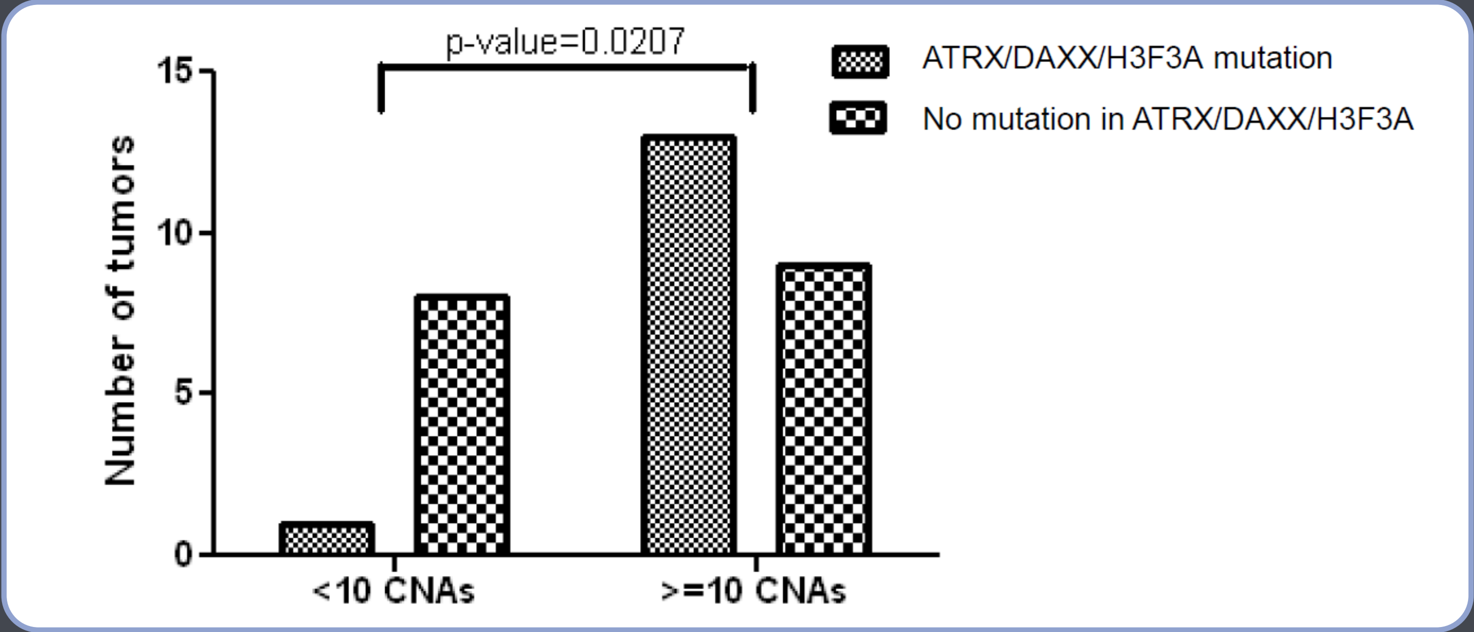
	ATRX IHC-positive	ATRX IHC-negative
ALT-negative	75	8
ALT-positive	10	37

J. Schwartzentruber et al., *Nature*, 2012

Changes in histone variant profiles can also happen in cancer

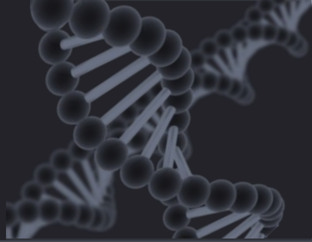
- Mutations in H3.3-ATRX-DAXX chromatin remodelling pathway were identified in 44% of glioblastomas (mostly high grade paediatric glioblastomas)
- The number of CNAs per tumour is higher in samples with *H3F3A/ATRX-DAXX/TP53* mutations)

ATRX and *DAXX* encode two subunits of a chromatin remodelling complex required for H3.3 incorporation at pericentric heterochromatin and telomeres



J. Schwartzentruber et al., *Nature*, 2012

Summary



- Chromatin is organised in LAD and TAD, related to histone modifications

DamID

Hi-C, 5-C, 4-C,…

- Histone modifications/histone variants have direct effect on gene transcription

Regression & Machine learning

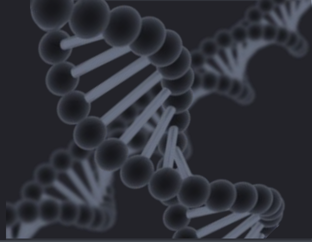
- Histone modifications form groups and indicate distinct chromatin states

ChIP-seq

Hidden Markov Models

- Epigenetic states change in cancer compared to normal ancestral cells (>30 epigenome-modifying proteins can be mutated in different cancers)

Summary



- Epigenetic states change in cancer compared to normal ancestral cells
    - Loss of imprinting
    - Hypermethylation of CpG islands of tumor suppressor genes
    - Genome-wide DNA hypomethylation
    - Long Range Epigenetic Silencing (LRES)
    - Large organized chromatin-lysine-(K) modifications (LOCKs)
    - Disruption of TADs and LADs
    - Depletion of histone variant H3.3 in pericentric heterochromatin and telomeres resulting in alternate lengthening of telomere (ALT) and increased number of copy number alterations