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.
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

Li et al, Cell, 2007
Histone modifications have direct effect on gene transcription

Correlation of different histone marks with gene expression

- H3K27me3
- H3K36me3
- H3K79me2
- H3K9me3
- H3K9ac
- H4K20me1

TSS

Density of histone modification signal

Gene Expression Level
- high
- moderate
- low
- low
- silent

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
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

Histone modifications form groups and indicate distinct chromatin states

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

---

**Predicted regions:**
- **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

---

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
Hidden Markov Models allow annotating chromatin states

- Probabilistic parameters of a Hidden Markov Model (HMM)
Hidden Markov Models allow annotating chromatin states

- Probabilistic parameters of a Hidden Markov Model (HMM)

Diagram:

- States: X1, X2, X3
- Observations: y1, y2, y3, y4
- State transition probabilities: a12, a21, a23
- Output probabilities: b11, b12, b13, b14, b21, b22, b24, b31, b32, b33, b34

\[ x \] — states
\[ y \] — possible observations
\[ a \] — state transition probabilities
\[ b \] — output probabilities
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
Hidden Markov Models allow annotating chromatin states

- Probabilistic parameters of a Hidden Markov Model (HMM)

- States: Rainy day, Sunny day, Eternal night
- Observations: y1, y2, y3, y4
- Transitions: a12, a21, a23
- Output probabilities: b11, b12, b13, b14, b21, b22, b23, b24, b31, b32, b33, b34

\[ x \rightarrow \text{states} \]
\[ y \rightarrow \text{possible observations} \]
\[ a \rightarrow \text{state transition probabilities} \]
\[ b \rightarrow \text{output probabilities} \]
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
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

\[ x_1 \rightarrow a_{12} \rightarrow x_2 \rightarrow a_{23} \rightarrow x_3 \]
\[ y_1 \rightarrow b_{11} \rightarrow y_2 \rightarrow b_{21} \rightarrow y_3 \rightarrow b_{13} \rightarrow y_4 \]

- Rainy day
- Sunny day
- Eternal night
Hidden Markov Models allow annotating chromatin states

- Probabilistic parameters of a Hidden Markov Model (HMM)

\[\begin{align*}
\text{Rainy day} & \xrightarrow{a_{12}} \text{Sunny day} & \xrightarrow{a_{23}} \text{Eternal night} \\
\text{X1} & \xrightarrow{a_{21}} \text{X2} & \text{X3} \\
\text{y1} & \text{y2} & \text{y3} & \text{y4}
\end{align*}\]

- \(x\) — states
- \(y\) — possible observations
- \(a\) — state transition probabilities
- \(b\) — output probabilities
Hidden Markov Models allow annotating chromatin states

- Probabilistic parameters of a Hidden Markov Model (HMM)

\[
\begin{align*}
X_1 & \xrightarrow{a_{12}} X_2 \\
X_2 & \xrightarrow{a_{23}} X_3 \\
\end{align*}
\]

This was Markov chain

- \(x\) — states
- \(y\) — possible observations
- \(a\) — state transition probabilities
- \(b\) — output probabilities
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

- Rainy day
- Sunny day
- Eternal night

- Shop
- Walk
- Read
- Sleep

\[ X_1 \]
\[ X_2 \]
\[ X_3 \]
Hidden Markov Models allow annotating chromatin states

- Probabilistic parameters of a Hidden Markov Model (HMM)

\[ x \rightarrow y \rightarrow a \rightarrow b \]

\[ x \rightarrow y \rightarrow a \rightarrow b \]

- States: \( X_1, X_2, X_3 \)
- Observations: \( Y_{1-4} \)
- Transition probabilities: \( a_{ij} \)
- Output probabilities: \( b_{ij} \)

\( x \) — states
\( y \) — possible observations
\( a \) — state transition probabilities
\( b \) — output probabilities
Hidden Markov Models allow annotating chromatin states

- Probabilistic parameters of a Hidden Markov Model (HMM)

\[ X_1 \xrightarrow{a_{12}} X_2 \xrightarrow{a_{21}} X_3 \]

\( X \) — states
\( y \) — possible observations
\( a \) — state transition probabilities
\( b \) — output probabilities

\( X_1 \):
- \( b_{11} \): Shop
- \( b_{12} \): Walk
- \( b_{13} \): Read
- \( b_{14} \): Sleep

\( X_2 \):
- \( b_{21} \): Shop
- \( b_{22} \): Walk
- \( b_{23} \): Read
- \( b_{24} \): Sleep

\( X_3 \):
- \( b_{31} \): Shop
- \( b_{32} \): Read
- \( b_{33} \): Sleep
- \( b_{34} \): Sleep

Wikipedia
Hidden Markov Models allow annotating chromatin states

- Probabilistic parameters of a Hidden Markov Model (HMM)

$X_1 \xrightarrow{a_{12}} X_2 \xrightarrow{a_{21}} X_3$  

$X_1 \xrightarrow{b_{12}} y_1 \quad X_2 \xrightarrow{b_{22}} y_2 \quad X_3 \xrightarrow{b_{32}} y_3 \quad X_3 \xrightarrow{b_{33}} y_4$

$x \rightarrow \text{states} \quad y \rightarrow \text{possible observations} \quad a \rightarrow \text{state transition probabilities} \quad b \rightarrow \text{output probabilities}$

- Rainy day $a_{12}$  
- Sunny day $a_{21}$  
- Eternal night $a_{23}$

- Shop $y_1$  
- Walk $y_2$  
- Read $y_3$  
- Sleep $y_4$
Hidden Markov Models allow annotating chromatin states

- Probabilistic parameters of a Hidden Markov Model (HMM)

\[ X_1 \quad \text{Rainy day} \quad a_{12} \quad \text{Sunny day} \quad a_{23} \quad \text{Eternal night} \]

\[ \begin{array}{l}
X_1 \quad b_{11} \quad b_{12} \quad b_{13} \\
X_2 \quad b_{21} \quad b_{22} \quad b_{23} \\
X_3 \quad b_{31} \quad b_{32} \quad b_{33} \quad b_{34}
\end{array} \]

\[ y_1 \quad \text{Shop} \quad y_2 \quad \text{Walk} \quad y_3 \quad \text{Read} \quad y_4 \quad \text{Sleep} \]

- \( x \) — states
- \( y \) — possible observations
- \( a \) — state transition probabilities
- \( b \) — output probabilities
Hidden Markov Models allow annotating chromatin states

- Probabilistic parameters of a Hidden Markov Model (HMM)

\[ x \quad \text{— states} \]
\[ y \quad \text{— possible observations} \]
\[ a \quad \text{— state transition probabilities} \]
\[ b \quad \text{— output probabilities} \]
Hidden Markov Models allow annotating chromatin states

- Probabilistic parameters of a Hidden Markov Model (HMM)

\[ x \rightarrow y \rightarrow a \rightarrow b \]

- \( x \) — states
- \( y \) — possible observations
- \( a \) — state transition probabilities
- \( b \) — output probabilities

Events:
- Rainy day
- Sunny day
- Eternal night

Actions:
- Shop
- Walk
- Read
- Sleep

\( y_1, y_2, y_3, y_4 \)
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
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

Rainy day
Sunny day
Eternal night
Shop
Walk
Read
Sleep
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

Example:

- States: Rainy day, Sunny day, Eternal night
- Observations: Shop, Walk, Read, Sleep
Hidden Markov Models allow annotating chromatin states

- Probabilistic parameters of a Hidden Markov Model (HMM)

\[ x \rightarrow y \rightarrow a \rightarrow b \]

- \( x \) — states
- \( y \) — possible observations
- \( a \) — state transition probabilities
- \( b \) — output probabilities

Example:

- \( x \): Rainy day, Sunny day, Eternal night
- \( y \): Shop, Walk, Read, Sleep
- \( b \): Transition probabilities from one state to another
Hidden Markov Models allow annotating chromatin states

- Probabilistic parameters of a Hidden Markov Model (HMM)

\[ x \rightarrow y \rightarrow a \rightarrow b \]

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

Example:

- States: Rainy day, Sunny day, Eternal night
- Observations: Shop, Walk, Read, Sleep

Wikipedia
Hidden Markov Models allow annotating chromatin states

- Probabilistic parameters of a Hidden Markov Model (HMM)

\[ x \rightarrow states \\
\begin{align*}
  & y \rightarrow possible \ observations \\
  a \rightarrow state \ transition \ probabilities \\
  b \rightarrow output \ probabilities
\end{align*}\]
Hidden Markov Models allow annotating chromatin states

- Probabilistic parameters of a Hidden Markov Model (HMM)

\[
\begin{align*}
X & \rightarrow y1 \rightarrow y2 \rightarrow y3 \rightarrow y4 \\
& x \rightarrow b11 \rightarrow b21 \rightarrow b31 \rightarrow b22 \rightarrow b32 \\
& y \rightarrow a1 \rightarrow a2 \rightarrow a3 \rightarrow a4 \\
& a \rightarrow b12 \rightarrow b22 \rightarrow b32 \rightarrow b42 \\
& b \rightarrow b13 \rightarrow b24 \rightarrow b34 \rightarrow b44
\end{align*}
\]

- \( x \) — states
- \( y \) — possible observations
- \( a \) — state transition probabilities
- \( b \) — output probabilities
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
Hidden Markov Models allow annotating chromatin states

- Probabilistic parameters of a Hidden Markov Model (HMM)

\[ x \rightarrow y \rightarrow b \rightarrow a \]

\[ x \rightarrow \text{Rainy day} \]
\[ \text{Sunny day} \rightarrow b_{21} \]
\[ \text{Eternal night} \rightarrow b_{11} \]
\[ \text{Shop} \rightarrow b_{12} \]
\[ \text{Walk} \rightarrow b_{22} \]
\[ \text{Read} \rightarrow b_{31} \]
\[ \text{Sleep} \rightarrow b_{13} \]
\[ y \rightarrow \text{Shop} \]
\[ \text{Walk} \rightarrow y_2 \]
\[ \text{Read} \rightarrow y_3 \]
\[ \text{Sleep} \rightarrow y_4 \]

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

This was hidden Markov model
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

\[ x \leftarrow \text{states to be defined!} \]

\[ y \leftarrow \text{possible observations} \]

\[ a \leftarrow \text{state transition probabilities} \]

\[ b \leftarrow \text{output probabilities} \]
Histone modifications = observations
Functional regions = states

Ernst & Kellis, *Nature Biotechnology*, 2010
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

Emission probabilities:
- H3K36me3
- H4K20me1
- H3K4me1
- H3K27me3
- H3K27ac
Hidden Markov Models allow annotating chromatin states

- Active promoter
- Transcribed region
- Poised promoter
- Transcription end
- Silenced chromatin
- Enhancer
- Weak enhancer

Emission probabilities:
- H3K36me3
- H4K20me1
- H3K4me1
- H3K27me3
- H3K27ac
Hidden Markov Models allow annotating chromatin states

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

Emission probabilities

Mark
- CTCF
- H3K27me3
- H3K36me3
- H4K20me1
- H3K4me1
- H3K4me2
- H3K4me3
- H3K27ac
- H3K9ac
- WCE

Hidden Markov Models allow annotating chromatin states

- **ChromHMM**: automating chromatin-state discovery and characterization

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

"There are three kinds of lies: lies, damned lies, and statistics."

Benjamin Disraeli
With chromatin states, one can predict gene expression

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

![Graph showing correlation between measured and predicted expression](image)

**Pearson's $r = 0.81$**

The prediction of expression is done using linear regression model

- We try to predict:
  \[ \log_2(Y_i) \], 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)

\[
\log_2(Y_i) \approx \sum_{\text{chromatin feature } j} \alpha_j \log_2(X_{ij} + a_j) + \beta
\]

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

Target gene **expressed**

![Diagram](image)

Target gene **silenced**

![Diagram](image)

[Fry, 2011]
Bisulfite sequencing employed to detect methylation status of Cytosine

- Bisulfite treatment transforms unmethylated cytosine in uracil
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.

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.
ENCODE project includes data for normal and cancer cell lines
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?
Histone and CpG-methyl modifying proteins are often mutated or deleted in cancer.

**Epigenome-modifying gene mutations in human cancer**

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

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

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)

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

<table>
<thead>
<tr>
<th>Study</th>
<th>Cancer type</th>
<th>Histone Modifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Song et al. 2012</td>
<td>Lung</td>
<td>H3K9ac, H3K9me3, H4K16ac</td>
</tr>
<tr>
<td>Barlési et al. 2007</td>
<td>Lung</td>
<td>H3K4me2, H3K9ac</td>
</tr>
<tr>
<td>Van Den Broeck et al. 2008</td>
<td>Lung</td>
<td>H4K5ac, H4K8ac, H4K12ac, H4K16ac, H4K20me3</td>
</tr>
<tr>
<td>Seligson et al. 2009</td>
<td>Lung</td>
<td>H3K4me2, H3K18ac</td>
</tr>
<tr>
<td>Seligson et al. 2005</td>
<td>Prostate</td>
<td>H3K4me2, H3K18ac</td>
</tr>
<tr>
<td>Ellinger et al. 2010b</td>
<td>Prostate</td>
<td>H3K4me2, H3K9me2, H3K9me3, H5Ac, H4Ac</td>
</tr>
<tr>
<td>Behbahani et al. 2012</td>
<td>Prostate</td>
<td>H4K20me1, H4K20me2</td>
</tr>
<tr>
<td>Bianco-Miotto et al. 2010</td>
<td>Prostate</td>
<td>H3K4me2, H3K18ac</td>
</tr>
<tr>
<td>Ellinger et al. 2012</td>
<td>Prostate</td>
<td>H3K27me3</td>
</tr>
<tr>
<td>Elsheikh et al. 2009</td>
<td>Breast</td>
<td>H3K18ac, H4K12ac, H3K4me2, H4K20me3, H4R3me2, H4K16ac</td>
</tr>
<tr>
<td>Leszinski et al. 2012</td>
<td>Breast</td>
<td>H3K9me3, H4K20me3</td>
</tr>
<tr>
<td>Müller-Tidow et al. 2010</td>
<td>Leukemia</td>
<td>H3K9me3</td>
</tr>
<tr>
<td>Park et al. 2008</td>
<td>Stomach</td>
<td>H3K9me3</td>
</tr>
<tr>
<td>Zhang et al. 2009</td>
<td>Stomach</td>
<td>H3K27me3</td>
</tr>
<tr>
<td>Teao et al. 2009</td>
<td>Esophagus</td>
<td>H3K18ac, H4R3me2, H3K27me3</td>
</tr>
<tr>
<td>I et al. 2010</td>
<td>Esophagus</td>
<td>H3K18ac, H4R3me2</td>
</tr>
<tr>
<td>Ellinger et al. 2010a</td>
<td>Kidney</td>
<td>H3K4me3, H3K4me2, H3K9me3</td>
</tr>
<tr>
<td>Rogenhofner et al. 2012a</td>
<td>Kidney</td>
<td>H3K9me1</td>
</tr>
<tr>
<td>Rogenhofner et al. 2012b</td>
<td>Kidney</td>
<td>H3K27me1, H3K27me2, H3K27me3</td>
</tr>
<tr>
<td>He et al. 2012</td>
<td>Liver</td>
<td>H3K4me3</td>
</tr>
<tr>
<td>Cai et al. 2011</td>
<td>Liver</td>
<td>H3K27me3</td>
</tr>
<tr>
<td>Manuyakorn et al. 2010</td>
<td>Pancreas</td>
<td>H3K4me2, H3K9me2, H3K18ac</td>
</tr>
</tbody>
</table>

One should apply specific methods to detect histone modifications in cancer

- Specific feature of cancer samples: large copy number changes
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**
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

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)

Read densities profiles are corrected for copy number

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

HMCan applies Hidden Markov Models to extract the signal

Normalized density $\Rightarrow$ “Yes-No” Signal

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

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

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

V. Boeva, A. Lermine et al, Bioinformatics, 2012
Nebula: web-service for analysis of ChIP-seq data

Statistics for external connections to Curie
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 by Nebula.

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

- Lamina associated domains (LAD) affect gene expression

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

- Scales of genome architecture
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 reconstruct 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.
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

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

Summary

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

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

• Histone modifications form groups and indicate distinct chromatin states

• 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