Outline

• Epigenetic phenomenon

• What is epigenetics?

• Epigenetic mechanisms

• Epigenome in health, in disease, and with environment

• Modern epigenomic technology and resources
The classics

• High risk of cancer, type 2 diabetes, obesity, shortened lifespan

Zinc, methionine, betaine, choline, folate, B₁₂

Morgan, Whitelaw, 1999
Waterland, Jirtle, 2004
Epigenetic mutation upregulate oncogenes

Normal:
- TE
- No activity
- Methylated CpG
- Unmethylated CpG
- Canonical promoter

Epigenetic insult or misregulation

Cancer:
- TE
- Active cryptic promoter
- Methylated CpG
- Unmethylated CpG
- Canonical promoter

Oncogene or proto-oncogene

mRNA transcripts

Oncogene expression
50% tumors had at least one oncoexaptation

A total of 145 events involving 117 oncogenes across 4103 tumors (53%). On average, each event was in 49 samples with the highest occurring one in 901 samples.

Jang, Shah, et al, Nat Genetics, 2019
Oncoexaptation: from prognostic to cancer evolution

Jang, Shah, et al, Nat Genetics, 2019
Conservation versus innovation – a complex model of evolution

Modzelewski and Shao, Cell 2021

Figure 5. The MT2B2-driven Cdk2ap1 D N isoform is evolutionarily conserved in human

(A and B) Preimplantation-specific Cdk2ap1 D N isoforms are derived from species-specific promoters (A) but exhibit evolutionary conservation in protein sequences (B).

(A) Mouse Cdk2ap1 D N originates from the MT2B2 promoter; human CDK2AP1 D N originates from a promoter region containing an L2a and a Charlie4z hAT transposon element. Blue, canonical exons; red, alternative exons.

(B) Canonical Cdk2ap1 and Cdk2ap1 D N isoforms are 97.4% and 98.8% identical, respectively, between mouse and human.

(C and D) Ectopic expression of CDK2AP1 D N, but not CDK2AP1 CAN, rescues defective cell proliferation in Cdk2ap1 MT2B2/ D MT2B2 morulae (C) and blastocysts (D), as demonstrated by BrdU incorporation and total cell number.

(C) Representative confocal images of BrdU staining (left), quantification of BrdU incorporation (middle), and total cell number (right) are shown for 3.0 dpc embryos.

(legend continued on next page)
Conserved protein isoform, species-specific expression/phenotype

**QUANTIFICATION AND STATISTICAL ANALYSIS**

- Quantification and statistical analysis for experimental data
- Quantification of genes and retrotransposons
- Differential expression analysis on genes and retrotransposons
- Differential expression analysis on retrotransposon-gene junctions
- Differential expression of Cdk2ap1 isoforms

**ADDITIONAL RESOURCES**

**SUPPLEMENTAL INFORMATION**

Supplemental information can be found online at [https://doi.org/10.1016/j.cell.2021.09.021](https://doi.org/10.1016/j.cell.2021.09.021).

Figure 7. A model on the transposon-dependent gene regulation of Cdk2ap1 in mammalian preimplantation embryos
Where is the Queen?
Workers are more methylated
Inhibition of Dnmt3 phenocopy Royal Jelly
The King of Butterflies
The Monarch
Same Genome
Different Epigenome
Different Phenotype
What is Epigenetics/Epigenomics?

- A mitotically or meiotically heritable state of different gene activity and expression (phenotype) that is independent of differences in DNA sequence (genotype) – based on Conrad Waddington, 1942

- The sum of the alterations to the chromatin template that collectively establish and propagate different patterns of gene expression (transcription) and silencing from the same genome.

- Epigenetic changes influence the phenotype without altering the genotype.

- While **epigenetics** often refers to the study of single genes or sets of genes, **epigenomics** refers to more global analyses of epigenetic changes across the entire genome.
The Epigenome

The complete number, location, and types of epigenetic modifications that occur in a given cell.
Epigenetic Landscape

Conrad Hal Waddington
(1905–1975)
Developmental biologist
Paleontologist
Geneticist
Embryologist
Philosopher
Founder for systems biology
Inheritance, broad definition of epigenetics

A

Trans epigenetic signal

B

Cis epigenetic signal
Epigenetics Mechanisms

RNA Interference

Gene Expression

Histone Modifications
Nucleosome Positioning

DNA Methylation
Epigenetic mechanisms

• DNA methylation
  • Normal cells: role in gene expression and chromosome stability
  • Cancer cells: consequences of aberrant hypo- and hyper-methylation

• Histone modification
  • Normal cells – the histone code
  • Cancer cells – consequences of altered histone modifying enzymes

• Interaction between DNA methylation, histone modifications and small RNAs

• Cell/tissue type specificity

• Gene/Environment interaction, disease susceptibility
What is DNA Methylation?
The 5\textsuperscript{th} base

\begin{align*}
\text{G} & \quad \text{C} \\
\text{C} & \quad \text{G}
\end{align*}
## History of DNA 5-mC

<table>
<thead>
<tr>
<th>Year</th>
<th>Event</th>
<th>Scientists</th>
</tr>
</thead>
<tbody>
<tr>
<td>1951</td>
<td>DNA 5-mC first reported</td>
<td>Wyatt</td>
</tr>
<tr>
<td>1968</td>
<td>Activity of a DNA 5-mC writer detected</td>
<td>Kalousek &amp; Morris</td>
</tr>
<tr>
<td>1975</td>
<td>Model for maintaining 5-mC across cell divisions proposed by 2 independent groups</td>
<td>Riggs, Holliday &amp; Pugh</td>
</tr>
<tr>
<td>1980</td>
<td>DNA 5-mC is associated with gene repression using 5-azacytidine</td>
<td>Jones &amp; Taylor</td>
</tr>
<tr>
<td>1982</td>
<td>De novo DNA methylation detected</td>
<td>Jahner et al.</td>
</tr>
<tr>
<td>1983</td>
<td>1st DNA 5-mC writer, Dnmt1, purified</td>
<td>Bestor &amp; Ingram</td>
</tr>
<tr>
<td>1987</td>
<td>DNA methylation of promoters associated with gene repression</td>
<td>Kovesdi et al.</td>
</tr>
<tr>
<td>1989</td>
<td>1st DNA 5-mC reader, MeCP1, discovered</td>
<td>Meehan et al.</td>
</tr>
<tr>
<td>1993</td>
<td>DNA 5-mC is associated with gene repression using dmnt1 knockout mice</td>
<td>Li et al.</td>
</tr>
<tr>
<td>1998</td>
<td>Function of Dnmt3a and Dnmt3b (de novo methylation of proviral DNA and repetitive sequences) determined</td>
<td>Okano et al.</td>
</tr>
<tr>
<td>1998</td>
<td>Additional DNA 5-mC readers, MeCP2, MBD1, MBD2 &amp; MBD4, discovered</td>
<td>Hendrich &amp; Bird</td>
</tr>
<tr>
<td>2002</td>
<td>Function of Dnmt3L (de novo methylation of maternal imprinted genes) determined</td>
<td>Hata et al.</td>
</tr>
<tr>
<td>2007</td>
<td>DNA methylation of gene bodies associated with gene expression</td>
<td>Hellman &amp; Chess</td>
</tr>
<tr>
<td>2009</td>
<td>DNA 5-mC erasers, TET1-3, discovered</td>
<td>Tahiliani et al., Kriaucionis &amp; Heintz</td>
</tr>
</tbody>
</table>

**Key:** function reader writer eraser
Two classes of DNA methyltransferases (DNMTs)

Jones and Liang, 2009
Nature Review Genetics
DNA Methylation is Heritable

5′-CpG-3′
3′-GpC-5′

DNMT: DNA methyltransferase
SAM: S-adenosyl-methionine
SAH: S-adenosyl-L-homocysteine

One-carbon donors

DNMT: DNA methyltransferase
SAM: S-adenosyl-methionine
SAH: S-adenosyl-L-homocysteine
DNA demethylation pathways

**Passive**

Replication

**Active**

TET

Replication

TET

BER & replication

BER & replication

BER = base excision repair pathway

DNA methylation is not distributed evenly in the mammalian genome

- In human somatic cells, 60%-80% of all CpGs (~1% of total DNA bases) are methylated
  - Most methylation is found in “repetitive” elements

- “CpG islands”, GC-rich regions that possess a high density of CpGs, remain methylation-free
  - The promoter regions of ~70% of genes are embedded in CpG islands
Function of DNA Methylation in Mammalian System

• Host defense - endogenous parasitic sequence (repeats, etc.)
• Imprinting
• X chromosome inactivation
• Heterochromatin maintenance, chromosome stability, telomere length
• Gene expression controls
Normal pattern and function of DNA methylation
DNA methylation changes during developmental epigenetic reprogramming
Normal Patterns of DNA Methylation

CpG islands and gene expression

Pericentromeric regions - stability

Transcription

Methylated CpG

Unmethylated CpG
Mechanisms of gene silencing by methylation

Direct mechanism:
Inhibition of transcription factor binding (e.g., CTCF, UBF)
Not a universal mechanism since not all transcription factor binding sites contain CG dinucleotides

Indirect mechanism:
Inhibition mediated by methyl-CpG binding proteins MeCP1/MeCP2
Recruitment of corepressor complexes including histone deacetylases (HDAC)
Change in chromatin conformation
DNA methylation in cancer
Point mutations
Deletions
CpG island hypermethylation
Chromosomal instability
Gene amplification
Genomic hypomethylation
Translocations
Copy number changes
Deletions
Tumorigenesis
Epigenetic
Genetic
Tumorigenesis
Tumor Suppressor Gene Inactivation in Human Cancer

Knudsen's 2-hit hypothesis, 1971
Cancer is a genetic and epigenetic disease

Aberrant DNA Methylation in Cancer

CpG islands hypermethylation

Pericentromeric hypomethylation

Methylated CpG

Unmethylated CpG
Different regions of the genome are hypermethylated or hypomethylated in cancer

Robertson, Nature Reviews Genetics, Vol6, 597
"Epigenetic cancer therapy"

**normal**
- tumor suppressor: ON

**tumor**
- tumor suppressor: OFF

**Diagram**
- 5-aza zebularine + Dnmt inhibitors: tumor suppressor OFF
- SAHA + HDAC inhibitors: tumor suppressor ON
DNA methylation and other diseases
Imprinting Diseases: 
Angelman and Prader-Willi Syndromes

Angelman Syndrome
• “Happy puppet”
• Severe mental retardation
• Absence of speech
• Happy disposition
• Excessive laughing
• Hyperactive, with jerky repetitive motions
• Red cheeks, large jaw and mouth

Prader Willi Syndrome
• Small hands and feet
• Underactive gonads, tiny external genitals
• Short stature
• Mentally retarded
• Slow-moving
• Compulsive overaters
• Obese
Imprinting Diseases: Angelman and Prader-Willi Syndromes

- Chr15 deletion, ~4Mb
  - Loss of maternal contribution
    - UBE3A
    - Angelman Syndrome
  - Loss of paternal contribution
    - SNRPN
    - Prader-Willi Syndrome
Genomic imprinting

• Imprinting is unique to mammals and flowering plants. In mammals, about 1% of genes are imprinted.
• For imprinted genes, one allele is expressed and the other is silent.
• This is typically controlled epigenetically. The expressed alleles are unmethylated and associated with loosely packed chromatin.
• Imprinted genes bypass epigenetic reprogramming.
• Imprinting is required for normal development.
Why imprinting?

- The Genetic Conflict Hypothesis
  - Many imprinted genes are involved in growth and metabolism.
  - Paternal imprinting favors the production of larger offspring, and maternal imprinting favors smaller offspring.

- Imprinted genes are under greater selective pressure.
  - No back up!
  - Any variation in single gene is expressed.
  - Closely related species have different imprinting patterns
    - Liger and Tigon

- Imprinted genes are sensitive to environmental signals.
What we don’t know about imprinting

• What targets a gene for imprinting?
  • Why are some genes expressed from both alleles and other expressed from only one allele?

• How are the imprints imposed?
  • Do males and females have different mechanisms for imprinting genes?
Rett Syndrome

• X-linked trait
• Mainly girls affected
• Normal at birth
• At 6–18 months, begin losing purposeful movement
• Persistent wringing of hands
• Loss of speech, gait
• Mental retardation ensues

Rett’s is due to defect in MeCP2

- Methyl-cytosine binding protein 2 (MeCP2) binds methylated DNA and recruits binding of a histone deacetylase

- Normal role is tightening chromatin packing, leading to gene silencing
Mouse model for Rett’s

• Mice have a gene that is homologous to MeCP2

• Knocking out the gene in mouse gives a phenotype similar to human Rett’s

• This model offers good experimental system for studying the human disease

Male mice with MeCP2 knockout develop normally for a while (middle), but at 6 weeks of age, they begin to develop neurological symptoms, such as hindlimb clasping (right).
Why is the phenotype neurological?

- The phenotype suggests that the targets are genes in the brain

- Normal neurological differentiation requires silencing of MeCP2 gene target(s)

- The target(s) of MeCP2 are not known

**LETTER**

**L1 retrotransposition in neurons is modulated by MeCP2**

Alysson R. Muotri, Maria C. N. Marchetto, Nicole G. Coufal, Ruth Oefner, Gene Yeo, Kinichi Nakashima & Fred H. Gage
DNA Methylation with age, environment
Differences in DNA methylation patterns between identical twins increase with aging

Fraga et al
PNAS 102:10604, 2005
Maternal Bisphenol A (BPA) Exposure

- Monomer that makes up polycarbonate plastic
- Endocrine active compound
- Found in commonly used products
- Present in 95% of humans tested
- Some animal studies reveal negative health outcomes
Maternal BPA Exposure

p-value = 0.0074

Control Diet (n=60)
BPA Diet (n=73)
Maternal Genistein Supplementation

- Plant phytoestrogen
- Found in soy and soy products
- Selective estrogen receptor modulator
- Worldwide exposure varies by diet
- Chemoprevention and decreased adipocyte deposition
Maternal Genistein Supplementation


p-value = 0.0005
Maternal Nutritional Supplementation

% of Offspring

- Yellow
- Slightly Mottled
- Mottled
- Heavily Mottled
- Pseudo-agouti

- Genistein Supplemented BPA Diet (n=39)
- Control Diet (n=60)
- BPA Diet (n=73)

p-value=0.96 (control versus genistein supplementation)
Oxidative derivatives of DNA 5-mC

5-hydroxymethylcytosine (5-hmC)
5-formylcytosine (5-fC)
5-carboxylcytosine (5-fC)
Many unanswered questions remain

- Who are the 5-hmC readers?
- How is 5-hmC maintained during DNA replication?
- What are the physiological context and relative importance of the various pathways involving 5hmC?
- How does 5-hmC affect transcriptional regulation?
- What are the dynamics of 5-hmC during development?
- What role does 5-hmC play in cancer?
DNA methylation + Histone modification → Chromatin
Chromatin: DNA plus protein in cells with nuclei

- 2 each of histones: H2A, H2B, H3, and H4

DNA double helix (2-nm diameter)

Nucleosome

146 bp of DNA

Tight helical fiber (30-nm diameter)

Metaphase chromosome

“Beads on a string”
The Nucleosome core particle

Nucleosome

A

H3

H4

NH2-ARTKQTARKSTGGKAPRKQLATKAAKRSAATPGYKKPHR-

Ac-SGRGKGGKGLGKGAKRHKKVLRD-

NH2-PEPSKAPAPKGGSKKAITRAQKDDGKRRKSARAK-

NH2-SGRGKQGGKARAKAK-

K79 in H3 core

Sk-COOH
Post-translational Histone Modifications

H3 tail Modifications:
- Acetylation
- Methylation

Active

HATs ↔ HDACs ↔ KMTases

Repressive
Epigenome provides an annotation of the genome
Chromatin modifications demarcate functional elements in the genome.
Chromatin modification patterns reveal various genomic features.

Rivera & Ren, Cell 2013

### Table 3. Distinctive Chromatin Features of Genomic Elements

<table>
<thead>
<tr>
<th>Functional Annotation</th>
<th>Histone Marks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Promoters</td>
<td>H3K4me3</td>
</tr>
<tr>
<td>Bivalent/Poised Promoter</td>
<td>H3K4me3/H3K27me3</td>
</tr>
<tr>
<td>Transcribed Gene Body</td>
<td>H3K36me3</td>
</tr>
<tr>
<td>Enhancer (both active and poised)</td>
<td>H3K4me1</td>
</tr>
<tr>
<td>Poised Developmental Enhancer</td>
<td>H3K4me1/H3K27me3</td>
</tr>
<tr>
<td>Active Enhancer</td>
<td>H3K4me1/H3K27ac</td>
</tr>
<tr>
<td>Polycomb Repressed Regions</td>
<td>H3K27me3</td>
</tr>
<tr>
<td>Heterochromatin</td>
<td>H3K9me3</td>
</tr>
</tbody>
</table>
Predicting non-coding RNA?

• From sequence?
  • Not clear which properties can be exploited
  • Sequence features such as promoters are too weak

• Histone modifications + conservation worked
Figure 1 | Intergenic K4–K36 domains produce multi-exonic RNAs.
How to detect epigenetic marks?
Restriction Landmark Genome Scanning (RLGS)

Unmethylated

Partial Methylation

Homozygous Methylation

NotI digest → endlabeling

NotI digest

EcoRV, 1-D, HinfI, 2-D

1st-D

2nd-D
Scaling to high coverage, high resolution

• Enrichment based methods
  • MeDIP-seq
  • MBD-seq/MethylMiner

• Restriction enzyme based methods
  • MRE-seq
  • HELP, Methyl-MAPS, Methyl-seq

• Bisulfite based methods
  • MethylC-seq
  • RRBS, bisulfite padlock

• Direct reading of modified nucleotides
  • SMRT
  • Nanopore sequencing
Enriching for methylated DNA targets
MeDIP-seq and MBD-seq

- Sonicate gDNA
- Size selection, end repair
- Adapter ligation
- Denaturation
- Size selection
- IP
Typical MeDIP data on a genome browser

5’ CpG islands are unmethylated

3’ CpG island is partially methylated
Taking advantage of methylation dependent restriction enzymes
MRE-seq

Digest gDNA (here with Hpa II)

Combine parallel digest
Size selection, end repair
Adapter ligation
Size selection
Typical MeDIP/MRE data on a genome browser

5’ CpG islands are unmethylated

3’ CpG island is partially methylated
Allele-specific methylation at imprinted genes

Chr 7:
- SNURF/SNURP
- SNURF-SNURP
- MeDIP-seq
- MRE-seq
- DMR

Chr 7:
- MEST
- MeDIP-seq
- MRE-seq
- DMR

Chr 18:
- SMAD4
- MeDIP-seq
- MRE-seq

imprinted

not imprinted
Gold standard: bisulfite sequencing
Bisulfite sequencing

Cytosine $\xrightarrow{HSO_3/\text{OH}^-} \text{Cytosinesulfonate} \xrightarrow{H_2O/\text{NH}_4^+} \text{Uracilsulfonate} \xrightarrow{\text{OH}^-/HSO_3^-} \text{Uracil}$

Bisulfite Treatment

ATTACGATCGAT-3’ $\xrightarrow{\text{m}}$ ATTACGATUGAT-3’

TAATGCTAGCTATA-5’ $\xrightarrow{\text{m}}$ TAATGCTAGUTATA-5’

PCR

ATTACGATCGATAT-3’

TAATGCTAATAT-3’ Complementary to original top (CTOT)

TAATGCTAGTTATA-5’ Complementary to original bottom (CTOB)

ATTACGATGATAT-3’ Original top (OT)

TAATGCTACTATA-5’ Original bottom (OB)

Grehl et al., Epigenomes, 2018
CpG islands
98% of the genome
1 CpG/100bp
majority methylated

<2% of the genome
1 CpG/10bp short stretches (~1000bp)
majority unmethylated

RRBS: Reduced Representations
Allow Enrichment of CpG Dinucleotides
5-mC detection with the Infinium BeadChip

**Pros**

- Relatively inexpensive
- Single-base resolution
- Internal quality controls
- Highly reproducible ($r > 0.98$)
- PCR-free protocol
- Works with FFPE samples

**Cons**

- Only covers a subset of the methylome
- Bisulfite treatment can damage DNA
- Dependent on bisulfite conversion
### 5-mC detection with the Infinium BeadChip

<table>
<thead>
<tr>
<th>Array</th>
<th>Year released</th>
<th># of Sites</th>
<th>Targeted sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>HumanMethylation27k</td>
<td>2008</td>
<td>&gt; 27k</td>
<td>• &gt;14K RefSeq genes</td>
</tr>
<tr>
<td>HumanMethylation450k</td>
<td>2011</td>
<td>&gt; 450k</td>
<td>• 99% of RefSeq genes</td>
</tr>
</tbody>
</table>
| Infinium MethylationEPIC | 2015          | > 850k     | • > 90% of sites in the 450k
• FANTOM5 and ENCODE enhancers (350k sites)
• CpG sites outside of CpG islands
• ENCODE open chromatin   |

**Context of the new EPIC probes**

![Image of Infinium BeadChip](image-url)

---

**www.illumina.com | Moran et al., Epigenomics (2016)**
Direct detection of modified nucleotides
Single-molecule real-time (SMRT) sequencing

SMRT sequencing discriminates between different bases by analyzing variations in polymerase kinetics

Pros
- Single-base resolution
- Measures absolute levels of many modified nucleotides
- "Raw" DNA is used
- Long reads

Cons
- Suboptimal accuracy
- Low throughput

Genome-wide detection of cytosine methylation by single molecule real-time sequencing

O. Y. Olivia Tse a,b,1, Peiyong Jiang a,b,10, Suk Hang Cheng a,b,1, Wenlei Peng b, Huimin Shang b, John Wong b, Stephen L. Chan a,b,10, Liona C. Y. Poon 1, Tak Y. Leung 1, K. C. Allen Chan a,b,10, Rossa W. K. Chiua,b,10, and Y. M. Dennis Lo a,b,10,2

Figure adapted from Clark et al., Nucleic Acids Research (2011) | Kinney et al., Epigenetic Alterations in Oncogenesis (2013)
Nanopore sequencing

Nanopore amperometry methods can discriminate between C, 5-mC, and 5-hmC due to differences in current profiles.

- **Pros**
  - Single-base resolution
  - Measures absolute levels of many modified nucleotides
  - “Raw” DNA is used
  - Long reads

- **Cons**
  - Suboptimal accuracy
  - Low throughput

Modern DNA Methylosomics

- Methylation-sensitive Restriction Digest
- 5-meC Antibody Immunoprecipitation
- Methylation-sensitive Digest
- Bisulfite (C to T; mC to C)
- Direct sequencing (Pol. kinetics)

- MRE-seq
- MeDIP-seq
- MBD-seq (MethylMiner)
- MethylC-seq
- RRBS
- SMRT Nanopore
Technologies for Interrogating Chromatin States

ChIP-seq

Zhang et al. Genome Biol 2008
Chromatin-IP Sequencing

A. Align reads

B. Infer positions of ChIP fragments

C. Count fragments at each genomic position

K4me1
K4me2
K4me3
K27me3

“active”

“repressive”

Alignable

ES (V6.5)

H3K4me3

H3K27me3

Oct4/ Tcf19/Cchcr1

Psors1c
Histone methylation and transcriptional state

**Transcribed gene**
- H3K4\(^{me3}\)
- H3K36\(^{me3}\)

**Silent developmental gene**
- K4me3
- K27me3

**Constitutive heterochromatin**
- K9me3
- K20me3

**'Poised' developmental gene**
- K4me3
- K27me3

**Markers**
- FoxP1
- Olig1
- LTR (long terminal repeat)

Histone methylation and transcriptional state
Mapping nucleosome positions

http://www.plosgenetics.org/article/info:doi/10.1371/journal.pgen.1003036
The presence of NFRs demonstrated that open promoter states are stable and common, even at genes that are transcribed so infrequently.

Cizhong Jiang, B Franklin Pugh, Nature Review Genetics, 2009 vol. 10 (3) pp. 161-72
DNase I hypersensitivity ~ Regulatory DNA

- promoters
- enhancers
- silencers
- insulators
- etc.
Digital DNaseI profiling

Precise delineation of the accessible regulatory DNA compartment
Digital DNaseI profiling: direct access to regulatory sequences
ATAC-seq
(assay for transposase-accessible chromatin)
Other interesting topics

- RNA component
  - RNAi, miRNA, X inactivation, HOTAIR, PiwiRNA
  - RNA methylation
- Reprogramming
- Cloning
- Population epigenetics
- Evolution of DNA methylation
- Evolution of epigenome
Engaging today’s epigenomic technologies and resources
Accessing the community resource

http://epigenomegateway.wustl.edu