The Epigenome

Caroline Relton
Institute of Genetic Medicine
Newcastle University, UK
Contents

• The evolution of contemporary epigenetics
• What is epigenetics?
• The component parts of a gene
• The role of the epigenome in the regulation of gene expression
• Known and unknown features of DNA methylation
• Histone modifications and non-coding RNA
• Interaction of epigenetic components
• The promise of epigenetic epidemiology
• What can we do with the information we generate?
The story of the most important revolution in modern biology - and what it could mean for humanity
Scientific literature

Epigenetic Epidemiology

Editors Choice : Ebrahim S: Epigenetics: The next big thing

Editorial : Relton C & Davey Smith G: Is epidemiology ready for epigenetics?

Contents
• Conrad Waddington on the Epigenotype 70 years on
• Twin studies in epigenetic epidemiology
• Socio-economic circumstances and epigenetic profiles
• Two step epigenetic Mendelian randomization
• Epigenetics, epidemiology and mitochondrial disease
• “Plasticity, robustness, development and evolution”: a symposium
• The exposome
• Popular epigenetics: more than Audrey Hepburn and yellow mice?

Feb 2012
What is Epigenetics?

Conrad Waddington (1905-1975) is often credited with coining the term epigenetics in 1942 as

“the branch of biology which studies the causal interactions between genes and their products, which bring the phenotype into being”.

Timeline of important discoveries in genetic and epigenetic regulation of gene expression

Zaidi SK et al. J Biol Chem 2011;286:18355-18361
The Human Epigenome Consortium is a public/private collaboration that aims to identify and catalogue Methylation Variable Positions (MVPs) in the human genome. 

http://www.epigenome.org/
Maps and signposts

DNA code  Epigenetic markings
Using the maps and signposts

DNA code

Transcription

Epigenetic signposts

Stochastic events
When the signposts don’t function properly

**Epigenetic epidemiology** – population-based approaches to understand the mechanisms by which our genes are regulated and the consequences of variation or perturbation of epigenetic signals on health
The role of the epigenome

Environment → Epigenome → Transcriptome → Proteome → Metabolome → Phenotype
The component parts of a gene

- Intron
- Exon
- Gene body
- Promoter
- Enhancer
- Transcription start site
- Transcription factor binding sites
- CREB
- Sp1
- MYC
- TGA
- CG
- CA
- GGG
- CG
- G
- CA
- CG
- TG

Many different transcription factors
Some harbour CpG sites
Another mechanism influencing transcription: insulators (transcriptional repressors)

On the maternal chromosome the ICR is unmethylated, this allows CTCF to bind and the insulator block \textit{IGF2} transcription by the \textit{H19} enhancer.

On the paternal chromosome the ICR is methylated, this prevents CTCF binding, inactivating the insulator and allowing the \textit{H19} enhancer to transcribe \textit{IGF2}.

\textit{Adapted from} Bell AC, Felsenfeld G. \textit{Nature} 2000;405:482-485
CpG Islands

**CpG islands** – CpG rich regions of DNA that are often associated with the transcription start sites of genes and that are also found in gene bodies and intergenic regions. Around 60% of all genes contain CGIs and the rest of the genome is depleted in CGIs.

Until recently much of the work on DNA methylation has focused on CGIs at TSS’s and this has tended to shape general perceptions about the function of DNA methylation.
CGI shores, shelves and open sea

Distribution of DNA methylation patterns across the genome

(a) Autosomes

(b) Methylation at the TSS

(c) Co-methylation

(d) Methylation in CGIs, histone modifications, and TF binding sites

Molecular anatomy of CpG sites and their role in gene expression

1. TSS are usually nucleosome deplete
2. Nucleosomes flanking the TSS are marked by trimethylation of histone H3 at lysine 4 (H3K4me3)
3. Histone variants such as H2A.Z are also important in transcriptional regulation
4. Downstream of the TSS DNA is mostly CpG deplete and is mainly methylated in repetitive elements and in gene bodies
5. CGIs in gene bodies usually remain unmethylated but may acquire methylation in a tissue specific manner
6. Transcription elongation is not blocked by gene body methylation
7. Enhancers tend to be CpG poor and show incomplete methylation, suggesting dynamic state, possibly involving TET
8. Binding of insulators such as CTCF can be blocked by methylation

DNA methylation and gene silencing

Adapted from Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nature Rev Genet. 29 May 2012. doi:10.1038/nrg3230
DNA methylation is negatively correlated with gene expression. (a) Methylation levels are low in the top quartile of highly expressed genes (left), and high in the bottom quartile of lowly expressed genes (right), looking across 12,670 autosomal genes. (b) Methylation levels with respect to the TSS in sets of genes categorized by gene expression levels, from highest (red) to lowest (blue), using the quartiles of gene expression with respect to gene expression means, where fitted lines represent running median levels.
Relationship between DNA methylation and genetic variation

- **Common genetic variation**
  - 6-8% of SNPs are associated with CpG site ablation or creation
  - Potential influence on co-methylation in the vicinity of SNPs
  - Differentially methylated sites identified through epigenetic studies should be assessed in light of SNP architecture
  - Are functional effects of these SNPs due to altered methylation?

- **De novo mutations**
  - 35% of mutations occur within CpG dinucleotides
  - Over 90% are C→T or G→A transitions
  - Within coding regions this equates to a 42-fold higher rate than that predicted from random mutations
Known features of DNA methylation in mammals

- Most CpG islands (CGI) are not methylated when located at transcription start sites (TSS)
- CGI methylation of the TSS is associated with long term silencing
- CGI in gene bodies are sometimes methylated in a gene specific manner
- Non-CGI methylation is more dynamic and more tissue specific than CGI methylation
- Methylation blocks transcription initiation but not elongation
- Methylation silences transposable (repeat) elements but allows the host gene to undergo elongation
- Gene body methylation contributes to somatic and germ-line mutations

*Adapted from* Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Rev Genet.* 29 May 2012. doi:10.1038/nrg3230
Unknown features of DNA methylation in mammals

- The role of non-CGI methylation in gene silencing (cause or consequence)?
- The function of methylation in a non-CpG context
- The roles of active and passive demethylation in activating genes
- The function of ‘shore’ methylation
- The role of gene body methylation in controlling splicing
- The role of 5-hydroxymethylation in the brain and other tissues
- The role of methylation in enhancer or insulator function
- Does silencing always precede methylation?

*Adapted from* Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Rev Genet.* 29 May 2012. doi:10.1038/nrg3230
Histone modifications

This figure illustrates nucleosome models and major post-translational modifications which play essential roles in gene expression regulation and disease processes.
# Co-location of methylation and histone marks

<table>
<thead>
<tr>
<th></th>
<th>Unmethylated CpG islands</th>
<th>Methylated CpG islands</th>
<th>$P$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of islands</td>
<td>13,372</td>
<td>2,583</td>
<td>$&lt;1 \times 10^{-45}$</td>
</tr>
<tr>
<td>Length</td>
<td>948.2</td>
<td>408.6</td>
<td>$&lt;1 \times 10^{-15}$</td>
</tr>
<tr>
<td>GC (%)</td>
<td>65.7</td>
<td>66.0</td>
<td>$&lt;1 \times 10^{-45}$</td>
</tr>
<tr>
<td>CpG$_{O/E}$</td>
<td>0.90</td>
<td>0.77</td>
<td>$&lt;1 \times 10^{-300}$</td>
</tr>
<tr>
<td>H3K4me3 (%)</td>
<td>92.81</td>
<td>0.08</td>
<td>$&lt;1 \times 10^{-175}$</td>
</tr>
<tr>
<td>H3K27me3 (%)</td>
<td>15.83</td>
<td>0.04</td>
<td>$&lt;1 \times 10^{-175}$</td>
</tr>
</tbody>
</table>

ncRNA

- A non-coding RNA (ncRNA) is a functional RNA molecule that is not translated into a protein
- Non-coding RNA genes include highly abundant and functionally important RNAs such as transfer RNA (tRNA) and ribosomal RNA (rRNA), as well as RNAs such as snoRNAs, microRNAs, siRNAs and piRNAs and the long ncRNAs
- The expression of many thousands of genes are regulated by ncRNAs. This regulation can occur in *trans* or in *cis*
- miRNAs are post-transcriptional regulators that bind to complementary sequences on target messenger RNA transcripts (mRNAs), usually resulting in translational repression or target degradation and gene silencing. The human genome may encode over 1000 miRNAs, which may target about 60% of mammalian genes and are abundant in many human cell types
Interaction of epigenetic components

Histone deacetylation and other modifications, particularly the methylation of H3 lysine 9 residues located in the histone tails, cause chromatin condensation and block transcriptional initiation.

Histone modification can also attract DNA methyltransferases to initiate cytosine methylation, which in turn can reinforce histone modification patterns conducive to silencing.

Experiments in yeast and plants have clearly shown the involvement of RNA interference in the establishment of heterochromatic states and silencing. RNA triggering of heritable quiescence might therefore also be involved in higher organisms.

From genes to phenotype – applying epidemiological approaches
Confirming observational associations

Environment

Genome → Epigenome → Transcriptome → Proteome → Metabolome → Phenome

Functional characterization

*In silico* validation, *in vitro* assays (LCLs, other cell lines), iPSC animal models, early stage trials
Integrating multi-dimensional information

**Bioinformatics & Statistics**

Data integration
Data mining
Analysis methods

**Exposome**
- Data capture
- Questionnaires
- Environmental monitoring
- Biological sample analysis

**Genome**
- SNPs
- CNVs
- mtDNA variants
- Rare variants
- Whole genome sequencing

**Phenome**
- Questionnaires
- Clinical assessment
- Health records
- Recall
- Imaging
- Educational attainment

**Epigen**
- DNA methylation
- miRNA expression
- Chromatin analysis

**Transcriptome**
- LCL gene expression
- Tissue specific expression profiles

**Proteome**
- Orbitrap-MS intact protein analysis

**Metabolome**
- NMR metabolites
- One carbon intermediates

**Orbitrap**
- Intact protein analysis
Realising the promise of epigenetic epidemiology

- Understanding what it is we are measuring and the role of epigenetic processes in biological pathways is crucial - particularly the complex relationship with gene expression

- Epigenetic patterns are an intermediate phenotype (or biomarker) that can be used to inform us about pathways to disease – and possible interventions

- Group level associations can be useful in identifying the determinants and consequences of epigenetic variation

- Statistical methods for longitudinal modelling commonly used in epidemiology can be applied to epigenetic data to study the impact of change over time

- Epidemiological tools can be applied to help decipher causal relationships
What can we do with the information we generate?

- Stratified medicine through biomarker identification
  - Prediction
  - Prognosis
  - Targeting intervention
- Refinement of conventional risk models
- Elucidation of novel pathways in disease pathogenesis
- Modulation of the epigenome by dietary, lifestyle or pharmacotherapeutic interventions
What can we do with the information we generate?

- **Stratified medicine through biomarker identification**
  - Prediction
  - Prognosis
  - Targeting intervention
- **Refinement of conventional risk models**
- **Elucidation of novel pathways in disease pathogenesis**
- Modulation of the epigenome by dietary, lifestyle or pharmacotherapeutic interventions
DNA Methylation Patterns in Cord Blood DNA and Body Size in Childhood

Caroline L. Relton¹, Alexandra Groom¹*, Beate St. Pourcain², Adrian E. Sayers³, Daniel C. Swan⁴, Nicholas D. Embleton⁵,⁶, Mark S. Pearce⁶, Susan M. Ring⁷, Kate Northstone², Jon H. Tobias³, Joseph Trakalo⁸, Andy R. Ness⁹, Seif O. Shaheen¹⁰, George Davey Smith²

Scientists In Newcastle have code that babies have at birth overweight.

New evidence has linked the environment in the womb with increased body weight in later life.

Scientists found changes around the DNA at birth which may result from a mother’s diet or exposure to pollution or stress.

They then linked these changes to a higher Body Mass Index (BMI) in children aged about nine years of age.

But the researchers say more work is needed to definitively prove the link between these changes and obesity.

Details are published in the journal Plos One.

Childhood or adult obesity has many causes, not least childhood or adult diet, but scientists have previously linked specific genes, such as the FTO gene, with increased body weight.

Others have looked at not the genes, but associated molecular changes - what are called epigenetics - which can play a role in how a gene functions in the body, switching genes on and off.

These changes are thought to be caused in part by exposure to...
**References**