Reality Check – How far has science brought us?

Trevor Archer | U.S. NIH/ National Institute of Environmental Health Sciences

U.S. Environmental Protection Agency
Epigenetics and Cumulative Risk Workshop
Arlington, VA
September 2 – 3, 2015
What is epigenetics?

‘Epigenetics’ was coined by Waddington as an English equivalent of Entwicklungsmechanik – the branch of science that undertakes a causal analysis of embryonic development by experimental means.

In modern biology, *epigenetics* has two closely related meanings:

The study of the processes involved in the unfolding development of an organism. This includes phenomena such as X-chromosome inactivation in mammalian females, and gene silencing within an organism.

The study of *heritable* changes in gene function that occur without a change in the sequence of nuclear DNA.

Epigenetics

The study of the processes involved in the unfolding development of an organism.

Epigenetic landscapes

A depiction of the epigenetic landscape. The ball represents a cell, and the bifurcating system of valleys represents the 'chreodes' or bundles of trajectories in state space.

One genome, many cell types, many epigenetic programs

- heart
- eye
- gut
- brain
- blood
The study of *heritable* changes in *gene* function that occur without a change in the sequence of *nuclear DNA*

- Changes in gene expression without changes to DNA sequence passed on through cell division
- Cellular “memory”
- Reversible
- Encoded in chromatin
Chromatin Compacts DNA
The structure of DNA in the nucleosome

Luger et al., Nature 1997

Histone tail modifications

Nicolas Lacoste et Jacques Côté
Chromatin Compacts and Regulates DNA

Multiple chromatin enzymes are required for regulated transcription.

- Chromatin Remodeling
- Histone modifications
- DNA Methyltransferases
- Regulatory RNAs
Epigenetics

Multiple chromatin enzymes are required for Epigenetic control.

Epigenetic landscapes

Inherited with each generation

Muscle  Liver  Leg  Eye

Chromatin Remodeling

Histone Modifications

DNA Methyltransferases

26S Proteosome

Regulatory RNAs

A depiction of the epigenetic landscape. The ball represents a cell, and the bifurcating system of valleys represents the 'chreodes' or bundles of trajectories in state space.

Molecular Mechanisms for setting up and reinforcing (epigenetic memory) alternative chromatin states.

DNA methylation-modification machinery:
- DNA methyltransferases
- Methyl-CpG Binding proteins
- Associated chromatin remodeling machinery (MeCP1/NuRD, MeCP2/Sin3a/Lsh)

Histone-modification machinery:
- Histone acetyltransferases
- Histone deacetylases
- Histone methyltransferases
- Histone demethylases
- Histone phosphorylases
- Non-Histone proteins (HP1 etc)
- Associated chromatin remodeling machinery

Function:
- Imprinting
- X inactivation
- Developmental gene expression
- Germ cell specific gene expression
- Repeat gene silencing
- Centromeric heterochromatin

Non-coding RNAs and heterochromatin (RNAi machinery)
RNAi (inhibition) was initially defined by Andrew Fire and colleagues as a process that is triggered by double strand RNA (dsRNA) and silences the expression of genes complementary to the dsRNA in Caenorhabditis elegans.
DNA methylation

- DNA methylation is:
  - a heritable epigenetic mark
  - essential for normal development

- Is a key repressive mark critical to:
  - Imprinting
  - X-inactivation
  - Suppression of repetitive elements
  - Regulation of gene expression (lineage-spec genes)

- Is a powerful mutagen
  - CpG suppression in mammalian genomes

- CpG Islands are normally free of this mark
Molecular Mechanisms for setting up and reinforcing (epigenetic memory) alternative chromatin states.

DNA methylation-modification machinery:
- DNA methyltransferases
- Methyl-CpG Binding proteins
- Associated chromatin remodeling machinery (MeCP1/NuRD, MeCP2/Sin3a/Lsh)

Histone-modification machinery:
- Histone acetyltransferases
- Histone deacetylases
- Histone methyltransferases
- Histone demethylases
- Histone phosphorylases
- Non-Histone proteins (HP1 etc)
- Associated chromatin remodeling machinery

Function:
- Imprinting
- X inactivation
- Developmental gene expression
- Germ cell specific gene expression
- Repeat gene silencing
- Centromeric heterochromatin

Non-coding RNAs and heterochromatin (RNAi machinery)
RNAi (inhibition) was initially defined by Andrew Fire and colleagues as a process that is triggered by double strand RNA (dsRNA) and silences the expression of genes complementary to the dsRNA in Caenorhabditis elegans.
Histone Code

H3K4  H3K9  H3K27

H4K16  H4K20
DNA Methylation and Histone Modifications help to compartmentalize the genome into domains of different transcriptional potentials

- Hyperacetylated histones
  - Low DNA methylation
  - H3-K4 methylation
  - H4-K16 acetylation

- Hypoacetylated histones
  - Dense DNA methylation
  - H3-K9 methylation
  - H4-K20 methylation
Epigenetic Significance for Long Term Health

Understanding how epigenetic information is deposited, maintained and processed is key to understanding gene regulation during development and disease.

DNA methylation

Histone modification
One genome, many cell types, many epigenetic programs

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

Figure 1 | Tissues and cell types profiled in the Roadmap Epigenomics Consortium. Primary tissues and cell types representative of all major lineages in the human body were profiled, including multiple brain, heart, muscle, gastrointestinal tract, adipose, skin and reproductive samples, as well as immune lineages, ES cells and iPS cells, and differentiated lineages derived from ES cells.

Data sets available for each reference epigenome.

Figure 2 | Data sets available for each reference epigenome. List of 127 epigenomes including 111 by the Roadmap Epigenomics program (E001 – E113) and 16 by ENCODE (E114 – E129). See Supplementary Table 1 for a full list of names and quality scores. a – d, Tissue and cell types grouped by type of biological material (a), anatomical location (b), reference epigenome identifier (EID, c) and abbreviated name (d). PB, peripheral blood. ENCODE 2012 reference epigenomes are shown separately. e – g, Normalized strand cross-correlation quality scores (NSC) for the core set of five histone marks (e), additional acetylation marks (f) and DNase-seq (g). h, Methylation data by WGBS (red), RRBS (blue) and mCRF (green). A total of 104 methylation data sets available in 95 distinct reference epigenomes. i, Gene expression data using RNA-seq (brown) and microarray expression (yellow). j, A total of 26 epigenomes contain 184 additional histone modification marks. k, Sixty highest-quality epigenomes (purple) were used for training the core chromatin state model, which was then applied to the full set of epigenomes (purple and orange).
So…. do we have a Reference Epigenome?

DNA methylation, typically associated with repressed regulatory regions or active gene transcripts and profiled using whole-genome bisulfite sequencing (WGBS), reduced-representation bisulfite sequencing (RRBS), and mCRF-combined methylation-sensitive restriction enzyme (MRE) and immunoprecipitation based assays (Fig. 2h)

2h, Methylation data by WGBS (red), RRBS (blue) and mCRF (green). A total of 104 methylation data sets available in 95 distinct reference epigenomes.
DNA modifications with epigenetic regulatory functions and their interdependencies. Cytosine (C) is methylated to 5-methylcytosine (5mC) by DNA methyltransferases (DNMT) and then further oxidised to 5hmC, 5fC and 5caC by Tet dioxygenases. 5-Hydroxyuracil (5hmU) is produced by Tet-catalysed oxidation of thymine (T). N6-methyladenine (6mA) is likely catalysed by DNA N6 adenine methyltransferases (DAMT-1 in C. elegans), even though the biochemical activity of these enzymes remains to be characterized. The Tet-like ALKB enzymes NMAD (N6-methyl adenine demethylase 1) and DMAD (DNA 6mA demethylase) have been shown to be involved in 6mA demethylation in C. elegans and in Drosophila, respectively, possibly by using a conserved dioxygenase mechanism.

High levels of non-CpG methylation in ES cells
So…. do we have a Reference Epigenome?

Histone H3 lysine 4 trimethylation (H3K4me3), associated with promoter regions10,24; H3 lysine 4 monomethylation (H3K4me1), associated with enhancer regions10; H3 lysine 36 trimethylation (H3K36me3), associated with transcribed regions; H3 lysine 27 trimethylation (H3K27me3), associated with Polycomb repression25; H3 lysine 9 trimethylation (H3K9me3), associated with heterochromatin regions26.

Selected epigenomes also contain a subset of additional epigenomic marks, including: acetylation marks H3K27ac and H3K9ac, associated with increased activation of enhancer and promoter regions27–29 (Fig. 2f);
Using this approach, we identified **130 unique PTM sites**, which not only confirmed **63 previously known** histone PTMs, but also revealed 67 novel ones, including 28 Kcr sites, 18 lysine monomethylation (Kme) sites, 1 lysine dimethylation (Kme2) site, 4 lysine formylation (Kfo) sites, 2 lysine acetylation (Kac) sites, 8 arginine monomethylation (Rme) sites, and 6 tyrosine hydroxylation (Yoh) sites (Figure 1C).
Tissues and cell types profiled in the Roadmap Epigenomics Consortium.

Figure 1 | Tissues and cell types profiled in the Roadmap Epigenomics Consortium. Primary tissues and cell types representative of all major lineages in the human body were profiled, including multiple brain, heart, muscle, gastrointestinal tract, adipose, skin and reproductive samples, as well as immune lineages, ES cells and iPSC cells, and differentiated lineages derived from ES cells.

Epigenetic Silencing in Cancer

Histone Deacetylases
Histone Methyltransferases
DNA Methyltransferases
The ‘epigenetic’ code?

Functions:
- Imprinting
- Developmental gene expression
- Germ cell specific gene expression
- Repeat gene silencing
- Centromeric heterochromatin

Lee Langer (NIEHS)
Paul Wade (NIEHS)
Richard Meehan (MRC-UK)
Paula Vertino (Emory)
Lisa Chadwick (NIEHS)
Archer Lab Colleagues