University of Zurich[UZH]

Institute of Molecular Life Sciences

Epigenomics

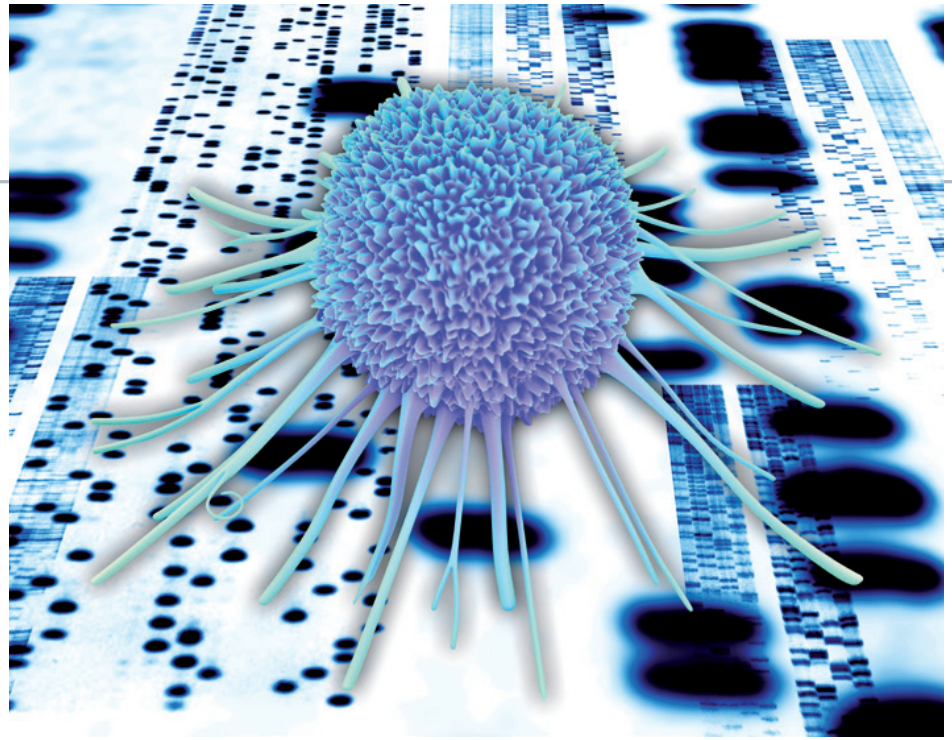
- **Part 1: Intro to epigenetics/epigenomics**
- Part 2: High-throughput technologies
- Part 3: Computational methods

Mark D. Robinson, Statistical Genomics, IMLS

A plug for (bioinformatics/ statistics in) epigenomics

There is also an intense demand for talent. In particular, epigenetics companies and individual labs need bioinformaticians as sequencing projects continue to dump terabytes of data into public databases (see *Nature* **482,** 263–265; 2012). Although this is an opportunity for job

PASIEKA/SPL



Computer reconstruction of a cancer cell on a DNA autoradiogram.

EPIGENETICS

Marked for success

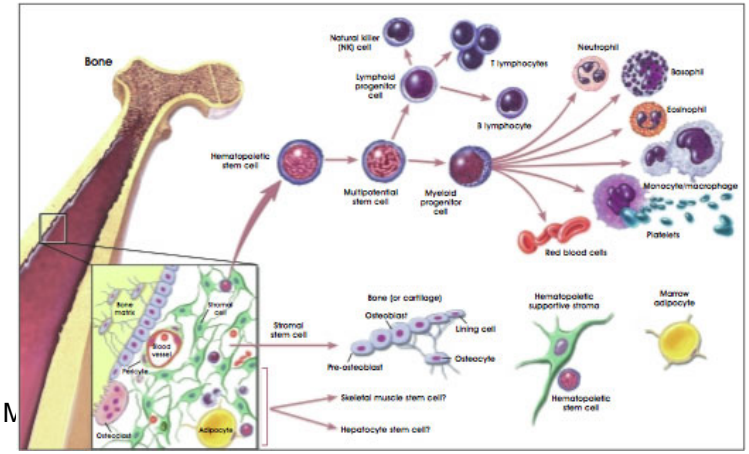
The growing field of cancer epigenetics demands computational expertise and translational research experience. Qualified practitioners are in high demand.

Genetics and Epigenetics

Genetics **can** explain differences between individuals.

Epigenetics can explain difference both ***between*** and ***within*** individuals.

Each cell type has the same DNA sequence, but **very** different epigenetic state.



Mendelism versus Lamarckism inheritance

Mendel: every individual possesses a pair of alleles each parent passes a randomly selected copy (allele) of only one of these to its offspring

Lamarck: an organism can pass on characteristics that it acquired during its lifetime to its offspring

http://en.wikipedia.org/wiki/Mendelian_inheritance

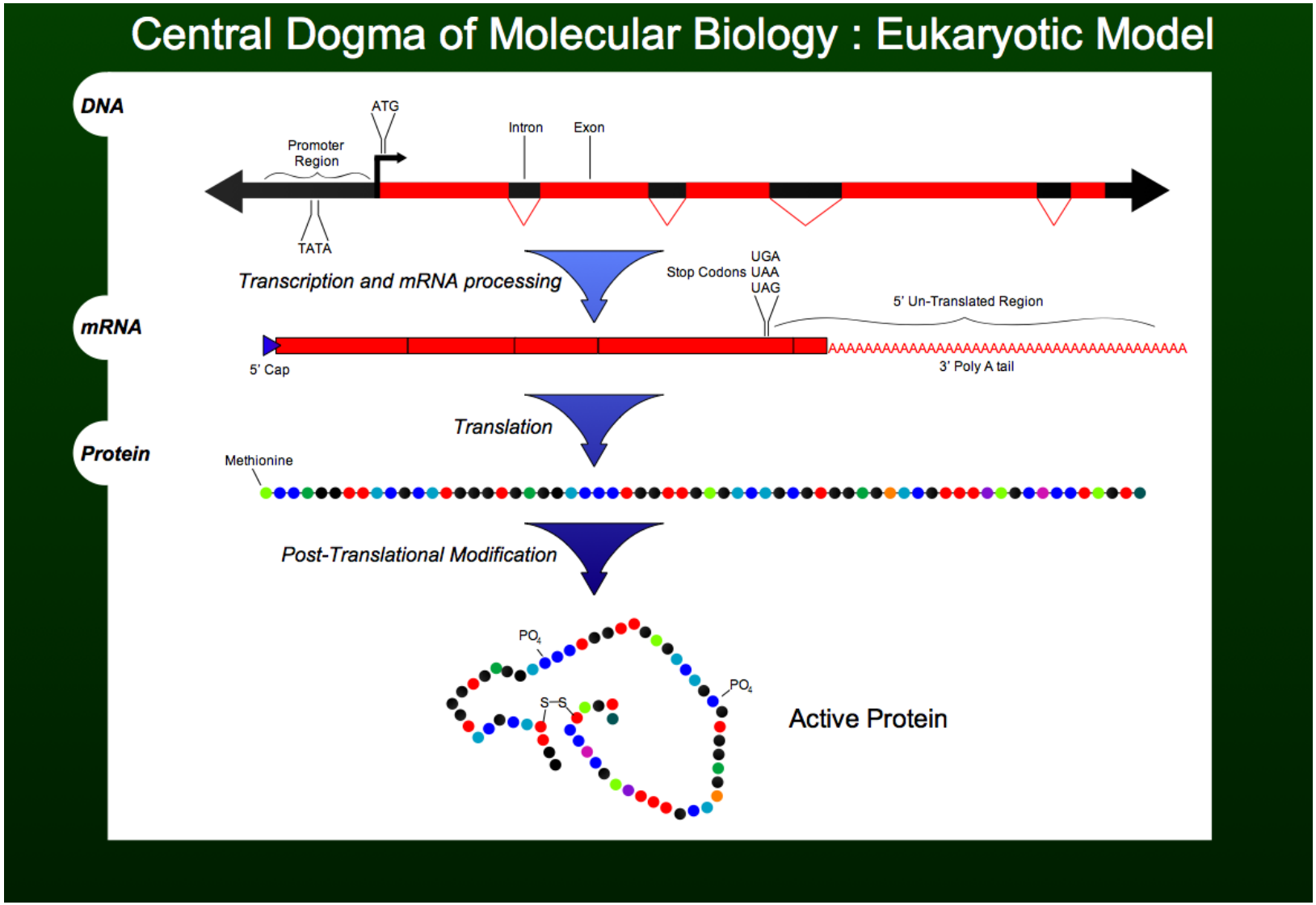
http://en.wikipedia.org/wiki/Lamarckism

Epigenetics definition

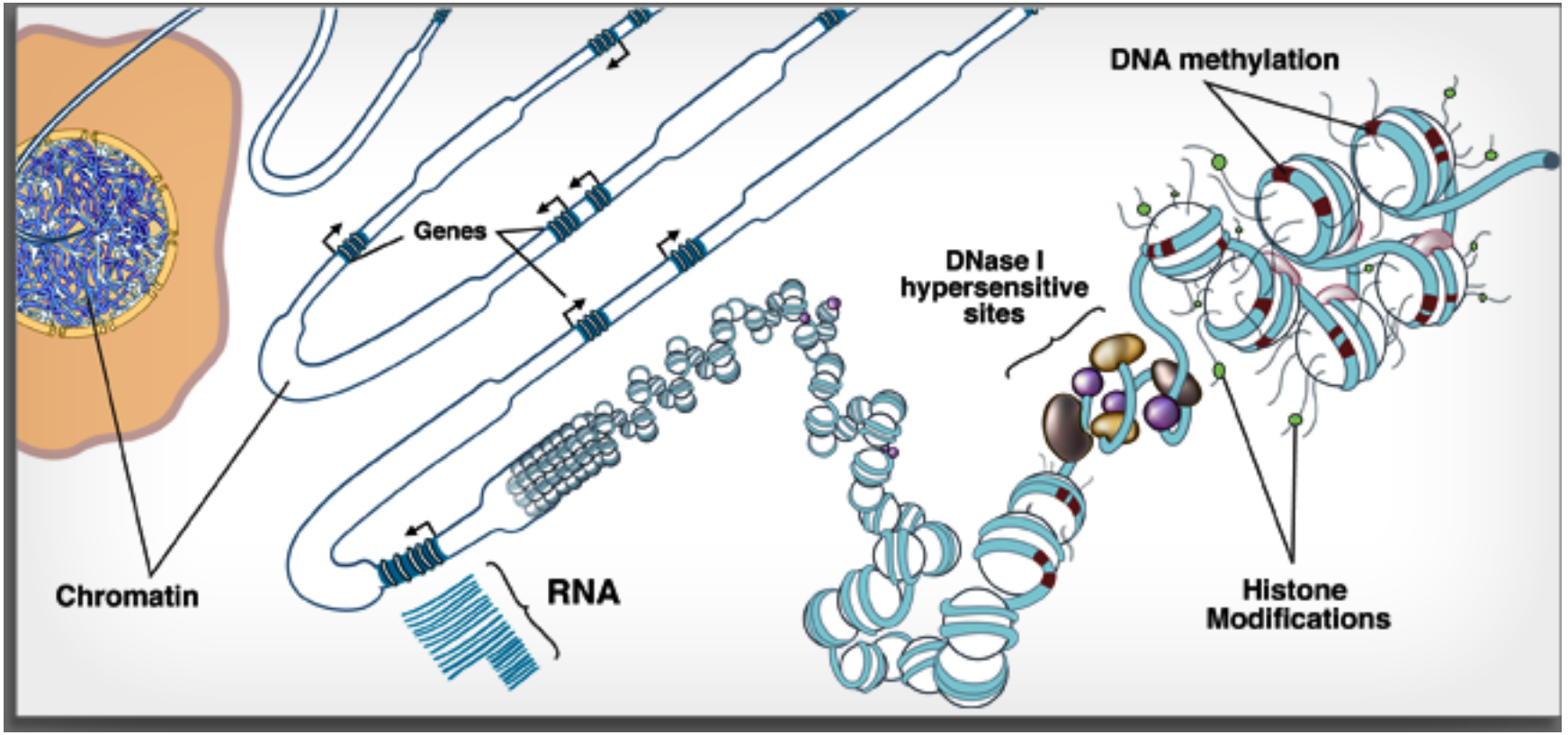
Epi - "on top of" or "in addition to"

“Epigenetics”:

- **heritable alterations in gene expression caused by mechanisms other than changes in DNA sequence.**
- the study of the mechanisms of temporal and spatial control of gene activity during the development of complex organisms
- "epigenetic code" has been used to describe the set of epigenetic features that create different phenotypes in different cells


Central Dogma of Molecular Biology : Eukaryotic Model
DNA
ATG
Promoter Region
Intron
Exon
TATA
Transcription and mRNA processing
UGA
Stop Codons UAA
UAG
5' Un-Translated Region
mRNA
5' Cap
3' Poly A tail
Translation
Protein
Methionine
Post-Translational Modification
PO4
PO4
S—S
Active Protein

Molecular basis of epigenetics



Epigenetic analogies

Computer: Two computers have the same specifications and software packages installed ("identical twins"). One user is doing word processing and email, the other is doing email and image processing. That is, the underlying instructions are common, but are being used ("expressed") differently.

Music: Genetics is the music, epigenetics is the musician's interpretation of the notes, rhythm, etc.

Television: You can fine tune the hue, brightness, contrast, etc., but you cannot change the original broadcast.

Script*: The Romeo and Juliet script is a fixed document ("genes"), but the director's interpretation ("epigenetics") can vary drastically (e.g. Baz Luhrmann 1996 Hollywood vs. Shakespeare).

*From The Epigenetics Revolution by Nessa Carey

Some compelling examples of epigenetics

1. X-inactivation
2. agouti mice (maternal diet affects coat colour)
3. Maternal **behaviour** in rats leaves a molecular footprint (that can be reversed)

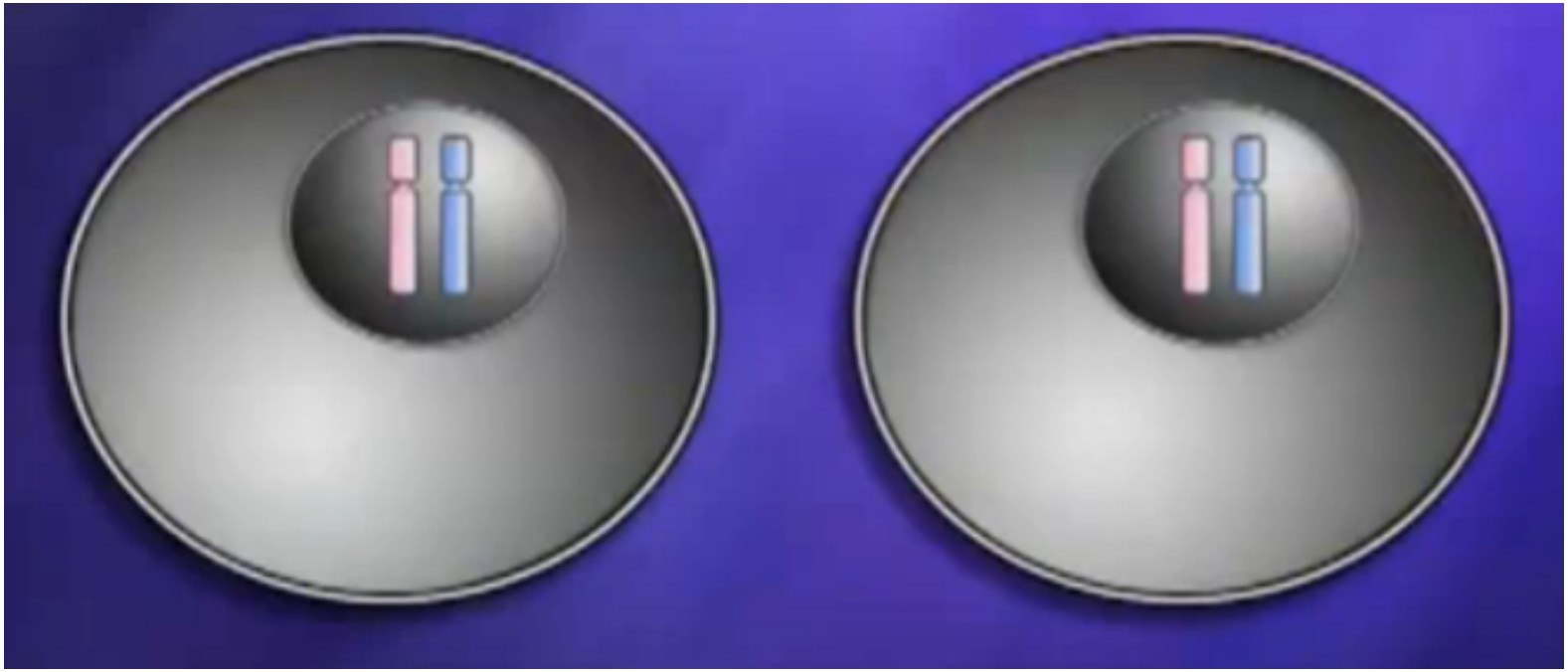
Calico cats

Why is there a patchy coat colour in calico cats?

Example 1: X-inactivation

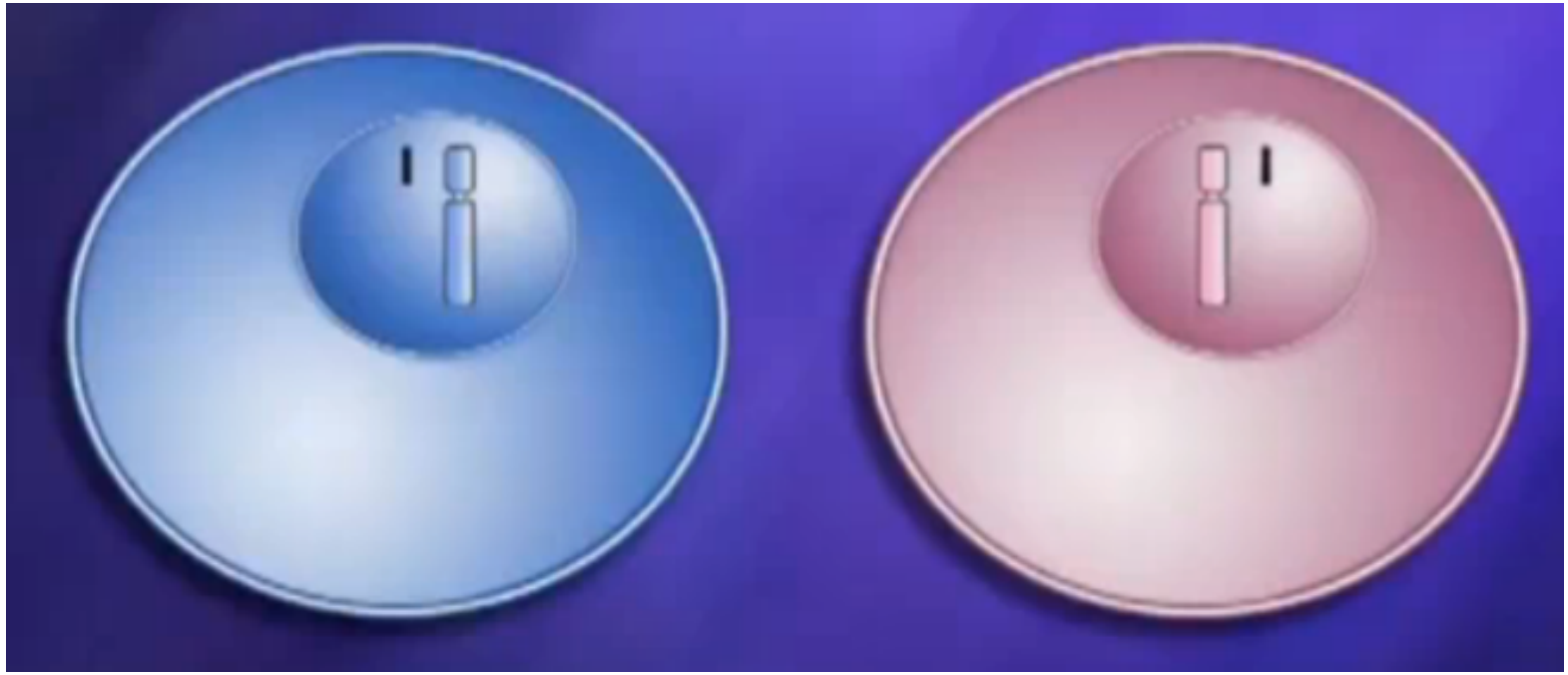
Females have 2 **X**-chromosomes, but one of them is (mostly) silenced. In early embryogenesis, either the maternal or paternal allele is silenced at random, but any subsequent cell divisions will maintain the silenced X. For example, calico coat colour is determined by an X-inactivation outcome (gene is on the X-chromosome).

X-inactivation



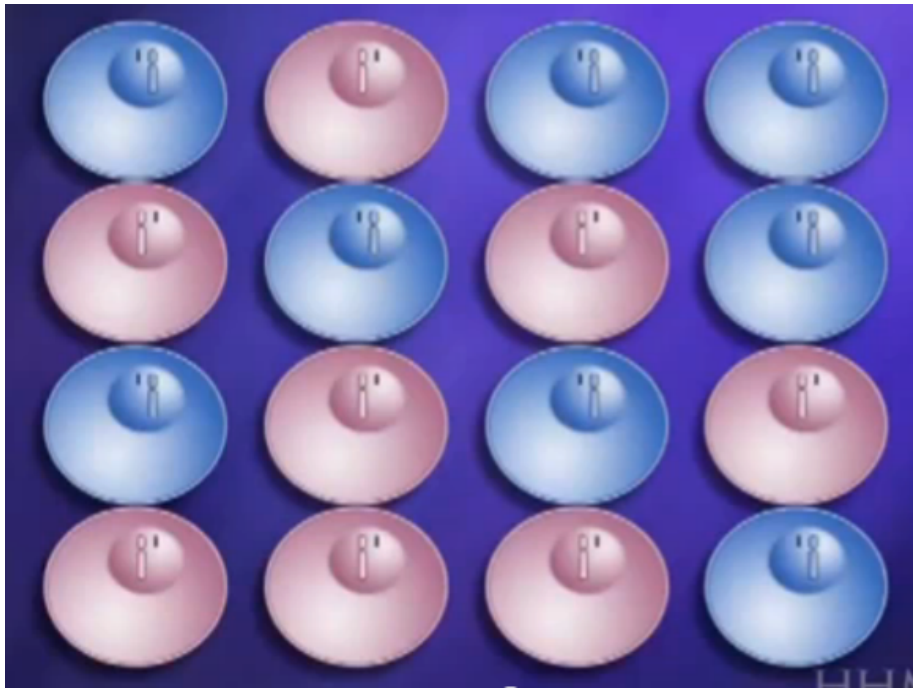
Two cells (from a female), each with 2 X-chromosomes

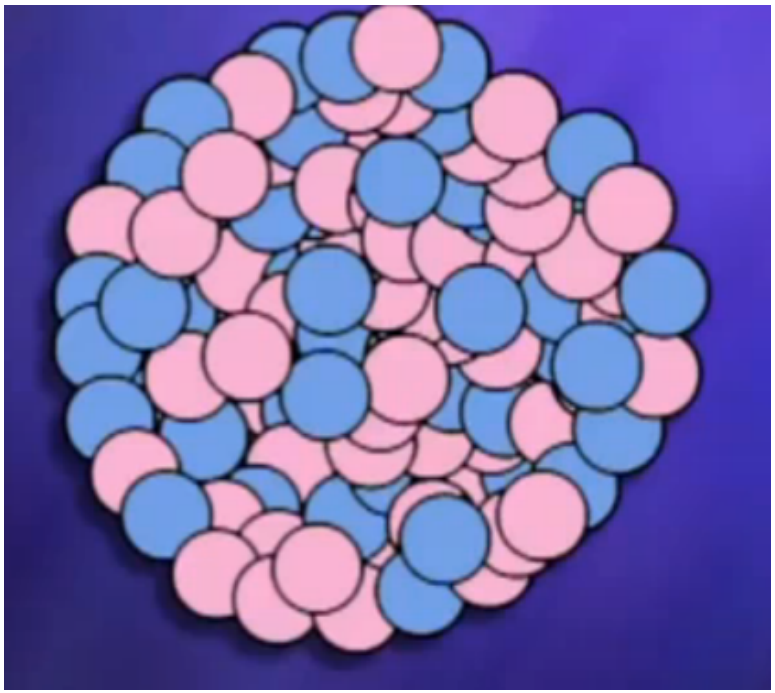
X-inactivation



One of the X chromosomes is randomly silenced.

X-inactivation

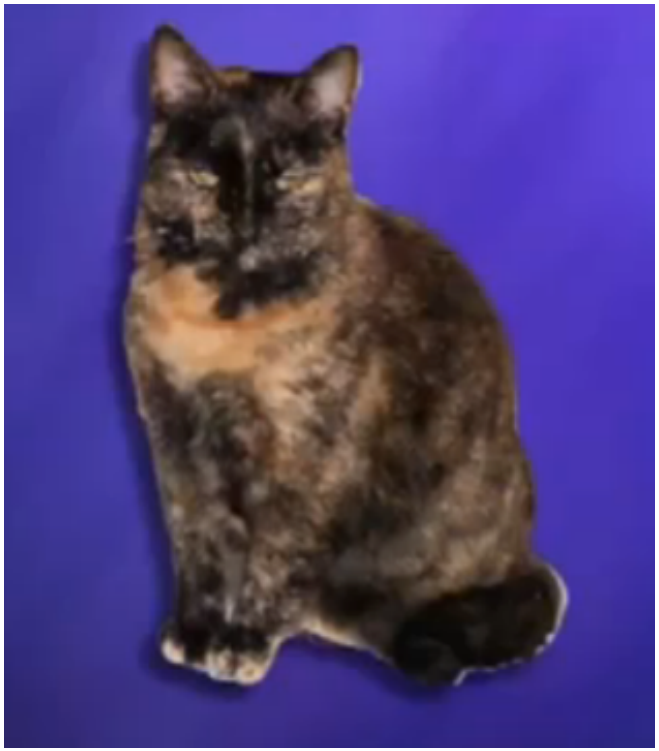




Cells divide, but preserve the inactivated X.

http://www.youtube.com/watch?v=n330FzHpI90

X-inactivation

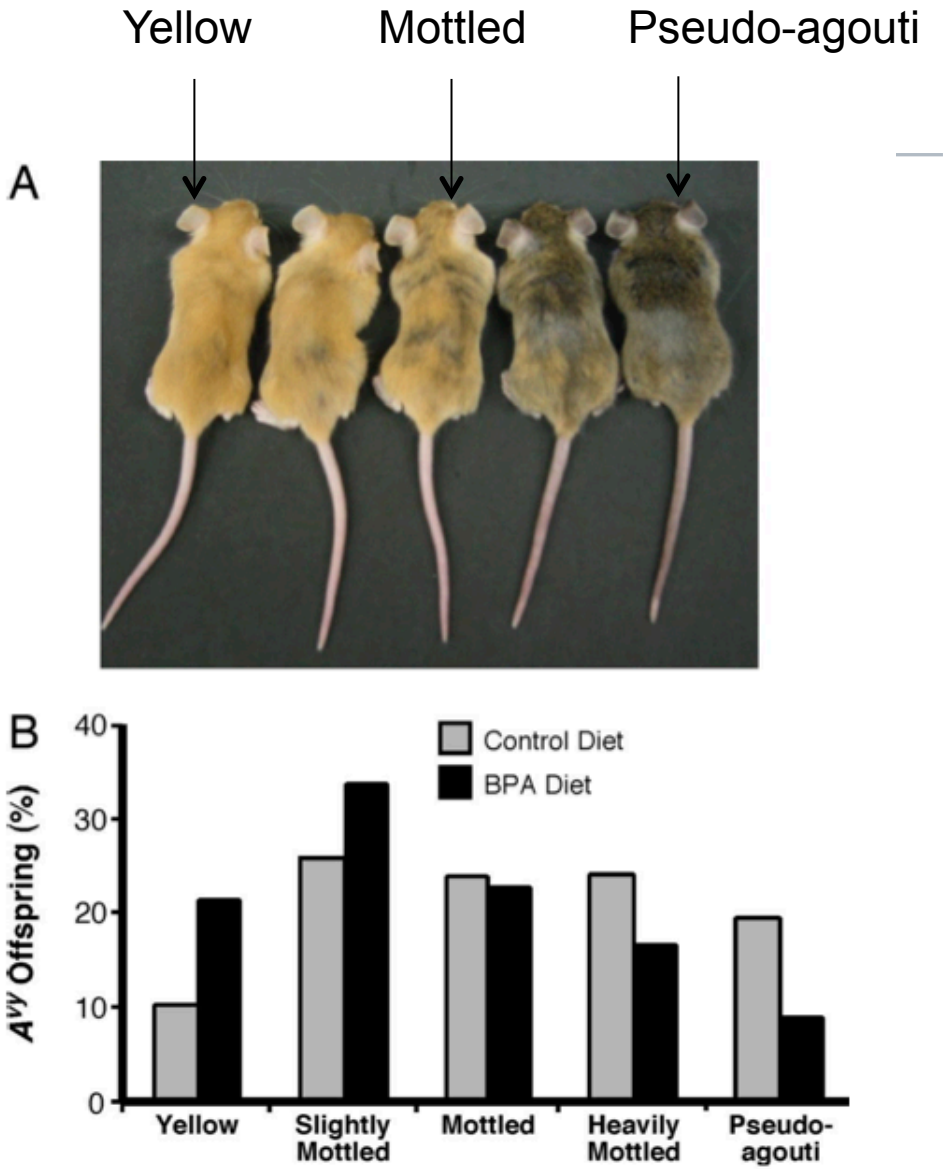


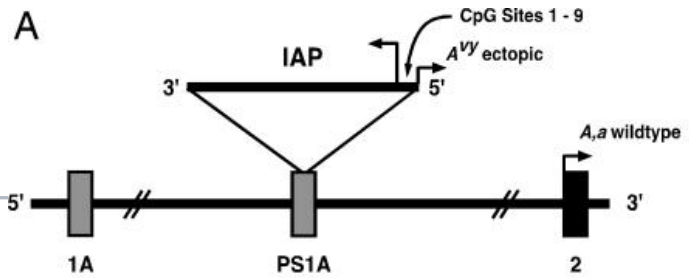
Result: patchy coat colours in female calico cats.

Example 2: Agouti mice

Observation: coat colour in offspring is strongly affected by mother's diet.

Molecularly, what is driving this?



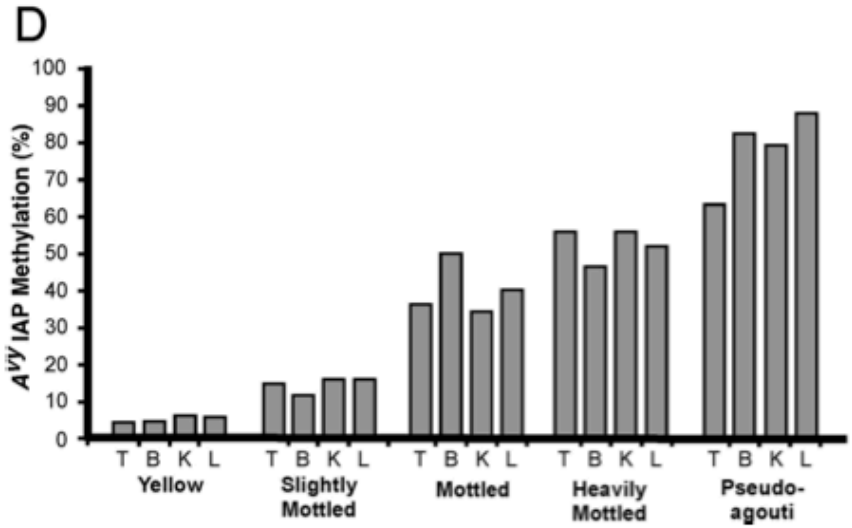


Agouti mice

Observations:

1. Methylation level (at promoter upstream of agouti gene) is strongly associated with coat colour.

2. Diet affects methylation level (in several tissues).



Maternal nutrient supplementation counteracts bisphenol A-induced DNA hypomethylation in early development

Dana C. Dolinoy*†‡, Dale Huang*, and Randy L. Jirtle*†‡§

*Department of Radiation Oncology and ‡University Program in Genetics and Genomics, Duke University, Durham, NC 27710; and †Integrated Toxicology and Environmental Health Program, Duke University, Durham, NC 27708

Edited by R. Michael Roberts, University of Missouri, Columbia, MO, and approved June 25, 2007 (received for review April 23, 2007)

Example 3: Behaviour affects DNA methylation

Epigenetic programming by maternal behavior

Ian C G Weaver[1,2], Nadia Cervoni[3], Frances A Champagne[1,2], Ana C D'Alessio[3], Shakti Sharma[1], Jonathan R Seckl[4], Sergiy Dymov[3], Moshe Szyf[2,3] & Michael J Meaney[1,2]

Here we report that increased pup licking and grooming (LG) and arched-back nursing (ABN) by rat mothers altered the offspring epigenome at a glucocorticoid receptor (GR) gene promoter in the hippocampus. Offspring of mothers that showed high levels of LG and ABN were found to have differences in DNA methylation, as compared to offspring of 'low-LG-ABN' mothers. These differences emerged over the first week of life, were reversed with cross-fostering, persisted into adulthood and were associated with altered histone acetylation and transcription factor (NGFI-A) binding to the GR promoter. Central infusion of a histone deacetylase inhibitor removed the group differences in histone acetylation, DNA methylation, NGFI-A binding, GR expression and hypothalamic-pituitary-adrenal (HPA) responses to stress, suggesting a causal relation among epigenomic state, GR expression and the maternal effect on stress responses in the offspring. Thus we show that an epigenomic state of a gene can be established through behavioral programming, and it is potentially reversible.

Summary of observations: tissue-centric (hippocampus), not related to prenatal conditions / ingested food (first week of life), reversible (by change in behaviour or HDAC drug), association with transcription factor, "causal"

Maternal behaviour leaves molecular signature

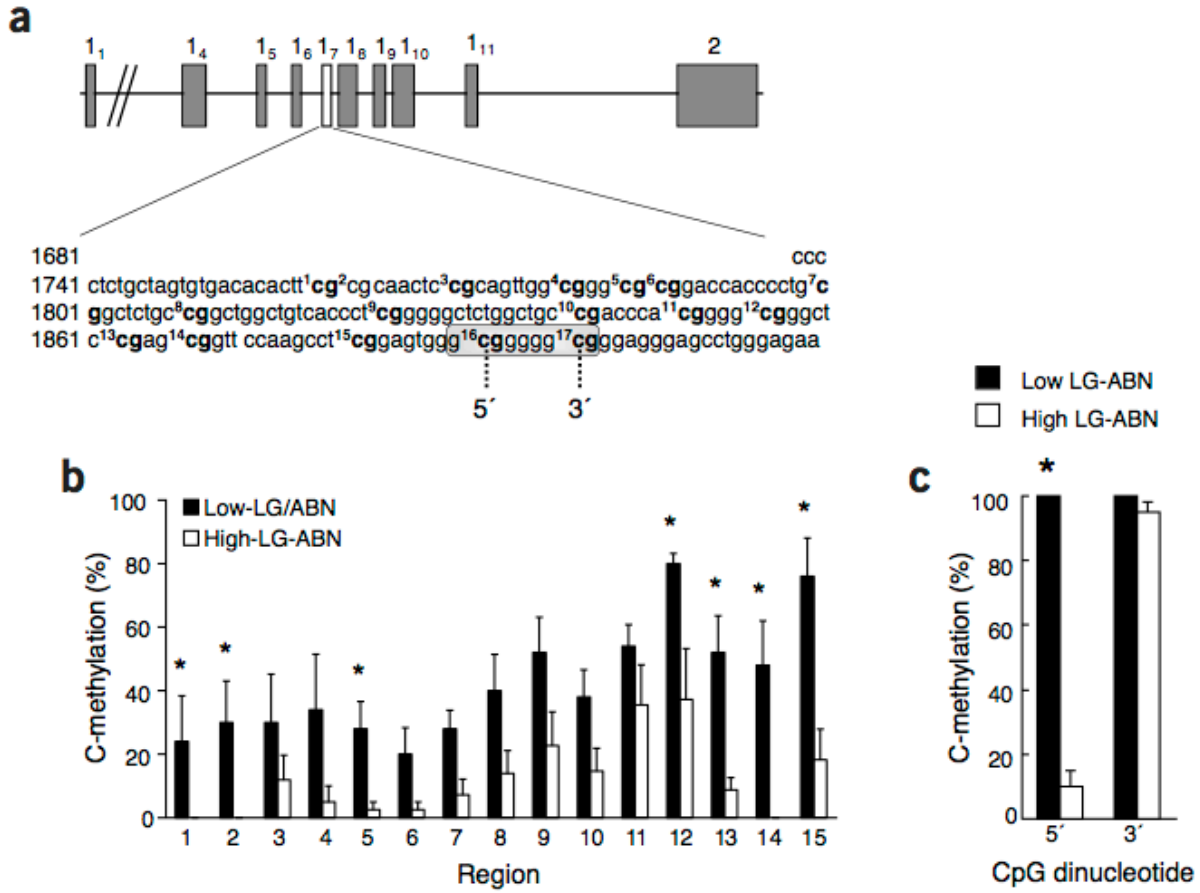


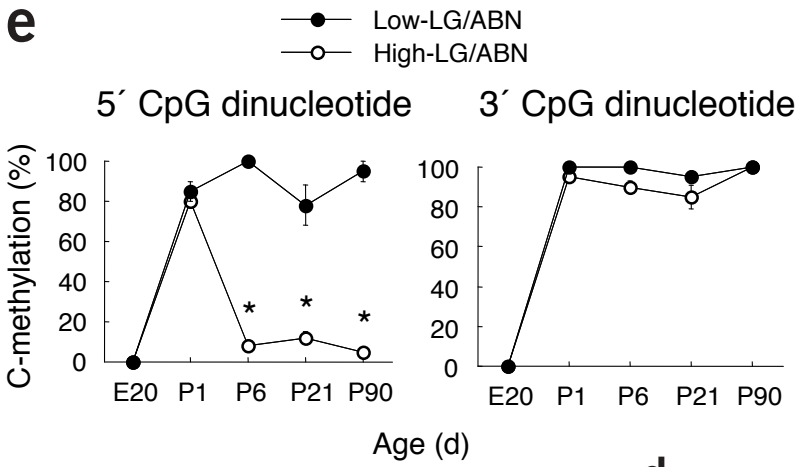
Figure 1 Maternal care alters cytosine methylation of GR promoter. (**a**) Sequence map of the exon 1_7 GR promoter including the 17 CpG dinucleotides (bold) and the NGFI-A binding region[16] (encircled). (**b**,**c**) Methylation analysis of the 17 CpG dinucleotides of the exon 1_7 GR promoter region from adult high- and low-LG-ABN offspring (6–10 clones sequenced/animal; $n = 4$ animals/group; $^*P < 0.01$). (**b**) Percentage of cytosine residues that were methylated (mean ± s.e.m.) for the first 15 CpG dinucleotides ($^*P < 0.05$). (**c**) Percentage of methylated cytosines (mean ± s.e.m.) for the 5′ (site 16) and 3′ (site 17) CpG dinucleotides within the NGFI-A binding sequence ($^*P < 0.0001$). (**d**) The effect

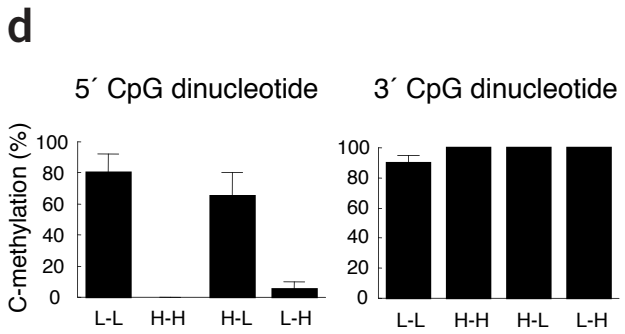
LG = licking grooming

ABN = arched-back nursing

Mothers are identifiable as “low” or “high” LG-ABN

Mothering determines status; timing is important

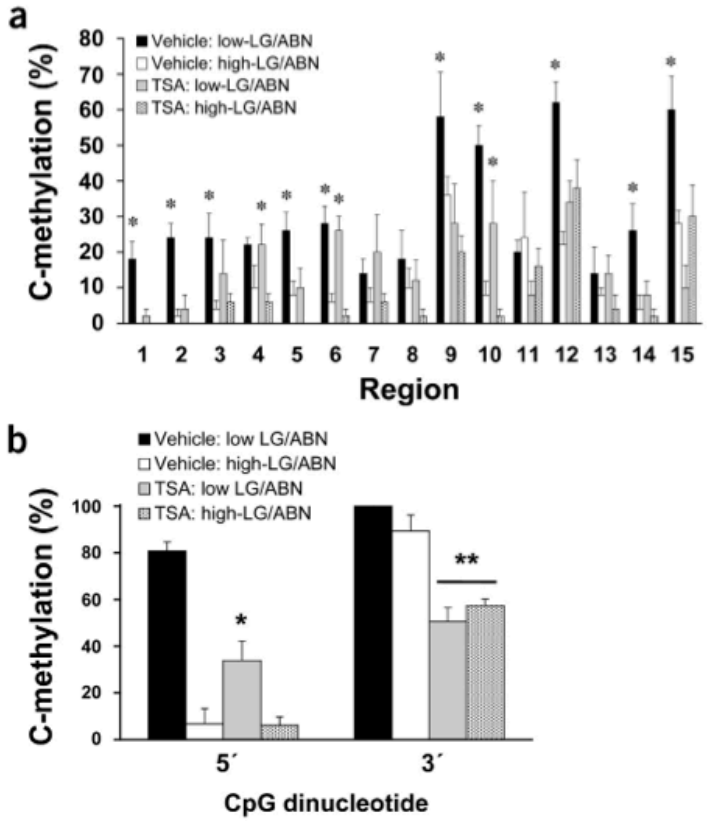




LG = licking grooming
ABN = arched-back nursing

binding sequence (*$P < 0.0001$). (**d**) The effect of cross-fostering the offspring of high- and low-LG-ABN mothers on cytosine methylation of the 5′ and 3′ CpG dinucleotides within the NGFI-A binding sequence of the exon 1_7 GR promoter gene in adult hippocampi ($n = 5$ animals/group). L-L: animals born to and reared by low-LG-ABN mothers; H-H: animals born to and reared by high-LG-ABN mothers; H-L: animals born to high-LG-ABN mothers and reared by low-LG-ABN mothers; L-H: animals born to low-LG-ABN mothers and reared by high-LG-ABN mothers. (**e**) Percentage of cytosine methylation (mean ± s.e.m.) of the 5′ and 3′ CpG dinucleotides within the NGFI-A binding region of the exon 1_7 GR promoter gene in the offspring of high- or low-LG-ABN mothers ($n = 5$ animals/group; $P < 0.001$) as a function of age. There were no differences at any postnatal age in level of cytosine methylation of the 3′ CpG (site 17).

(Maternal behaviour induced) molecular signature can be reversed by treatment



Treatment with histone deacetylase (HDAC) inhibitor.

Example 3: Maternal Behaviour affects epigenetics

A few remarks:

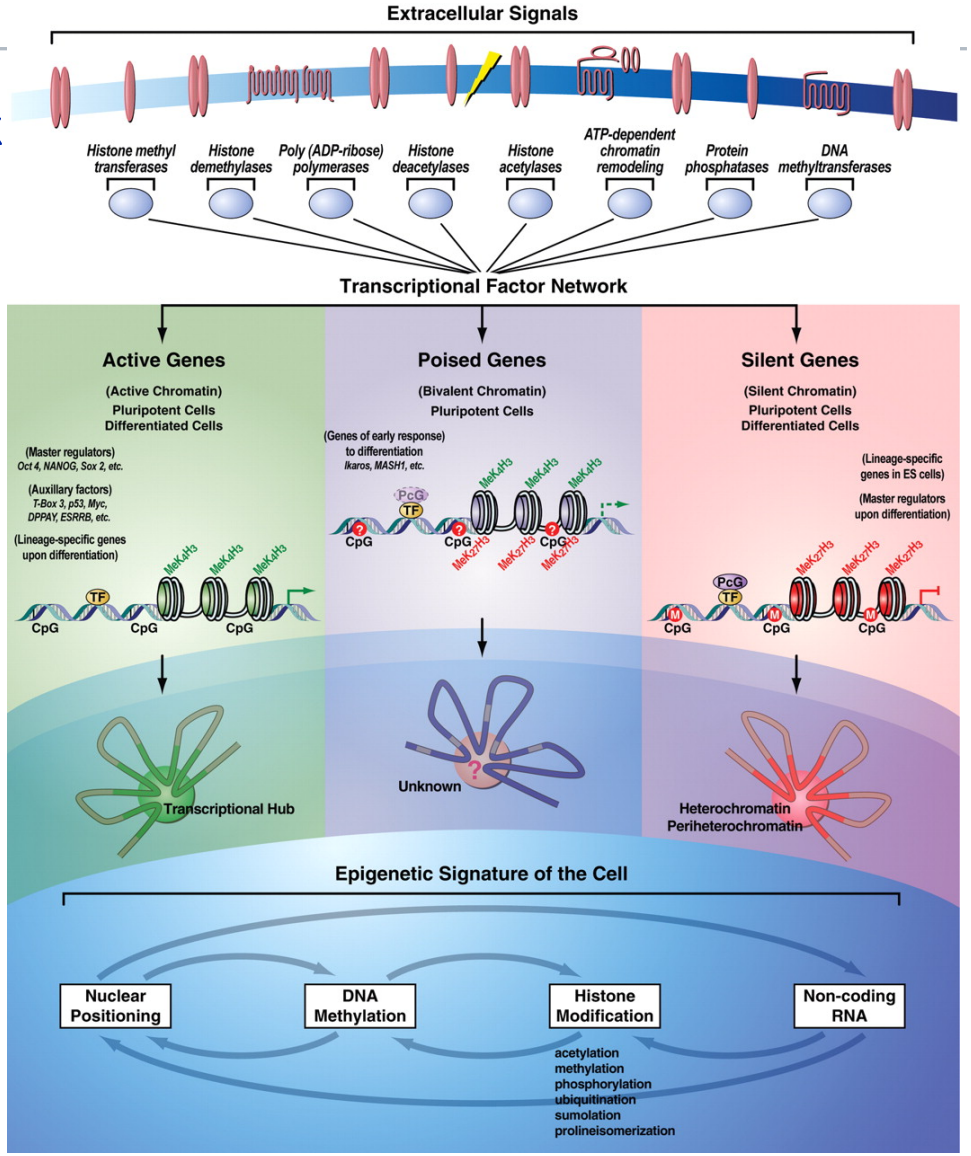
1. Environmental effect
2. Affects TF binding
3. Association to chromatin (histone) structure
4. Affects gene expression
5. Reversible

DISCUSSION

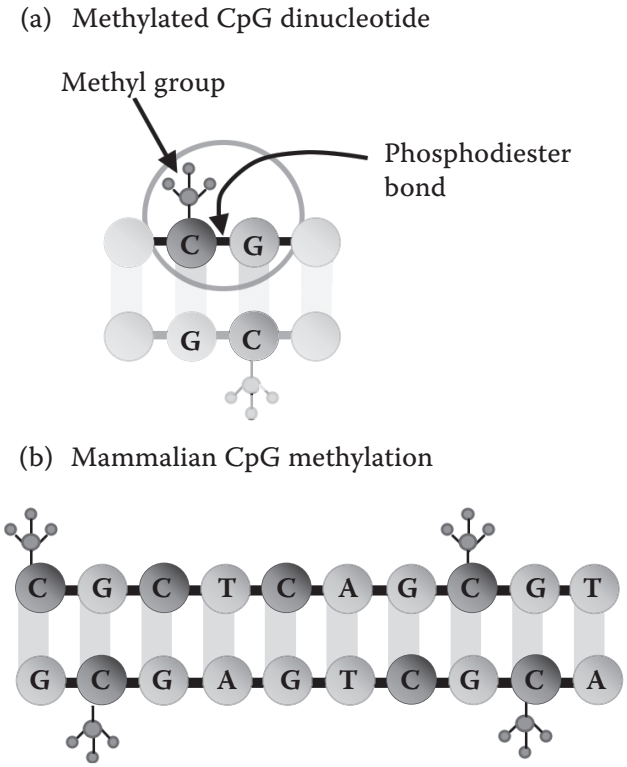
Further studies are required to determine how maternal behavior alters the epigenetic status of the exon 1_7 GR promoter. In addition, the exact causal relationship between DNA methylation and altered histone acetylation and NGFI-A binding remains to be defined. Nevertheless, our findings provide the first evidence that maternal behavior produces stable alterations of DNA methylation and chromatin structure, providing a mechanism for the long-term effects of maternal care on gene expression in the offspring. These studies offer an opportunity to clearly define the nature of gene-environment interactions during development and how such effects result in the sustained 'environmental programming' of gene expression and function over the lifespan. It is important to note that maternal

Epigenetics in concert with TF network

"… **suggests** that epigenetic players such as histone modifications, DNA methylation, the alteration of chromatin structure due to chromatin remodeling, and non-coding RNAs represent another crucial mechanism, besides the transcriptional factor network, which is designed by nature for the regulation of gene expression and cellular differentiation. Elucidating epigenetic mechanisms **promise** to have important implications for advances in stem cell research and nuclear reprogramming and **may offer** novel targets to combat human disease **potentially** leading to new diagnostic and therapeutic avenues in medicine."



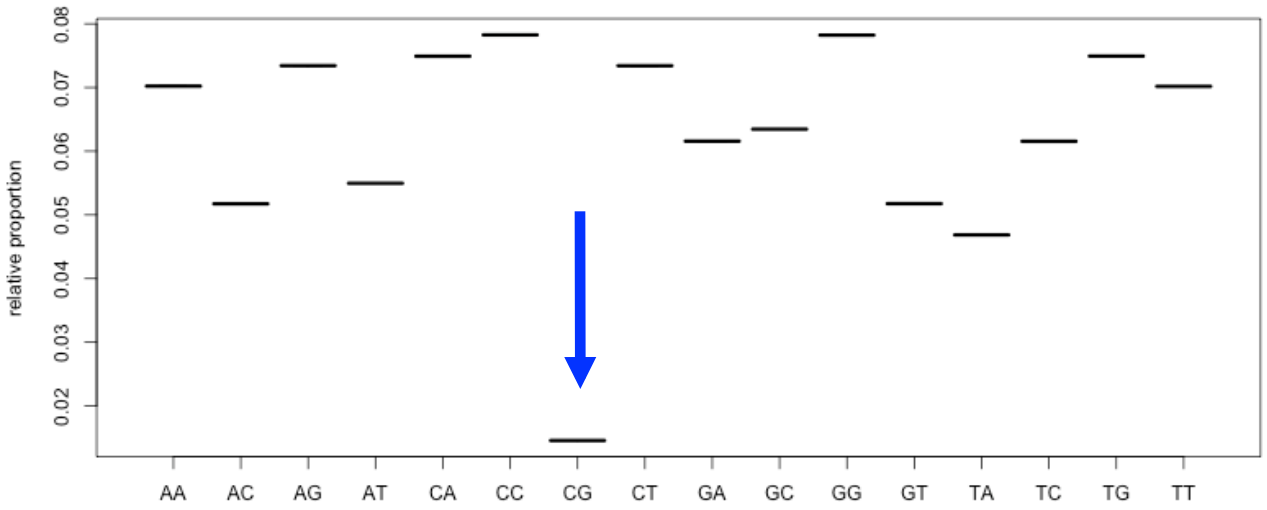
DNA methylation



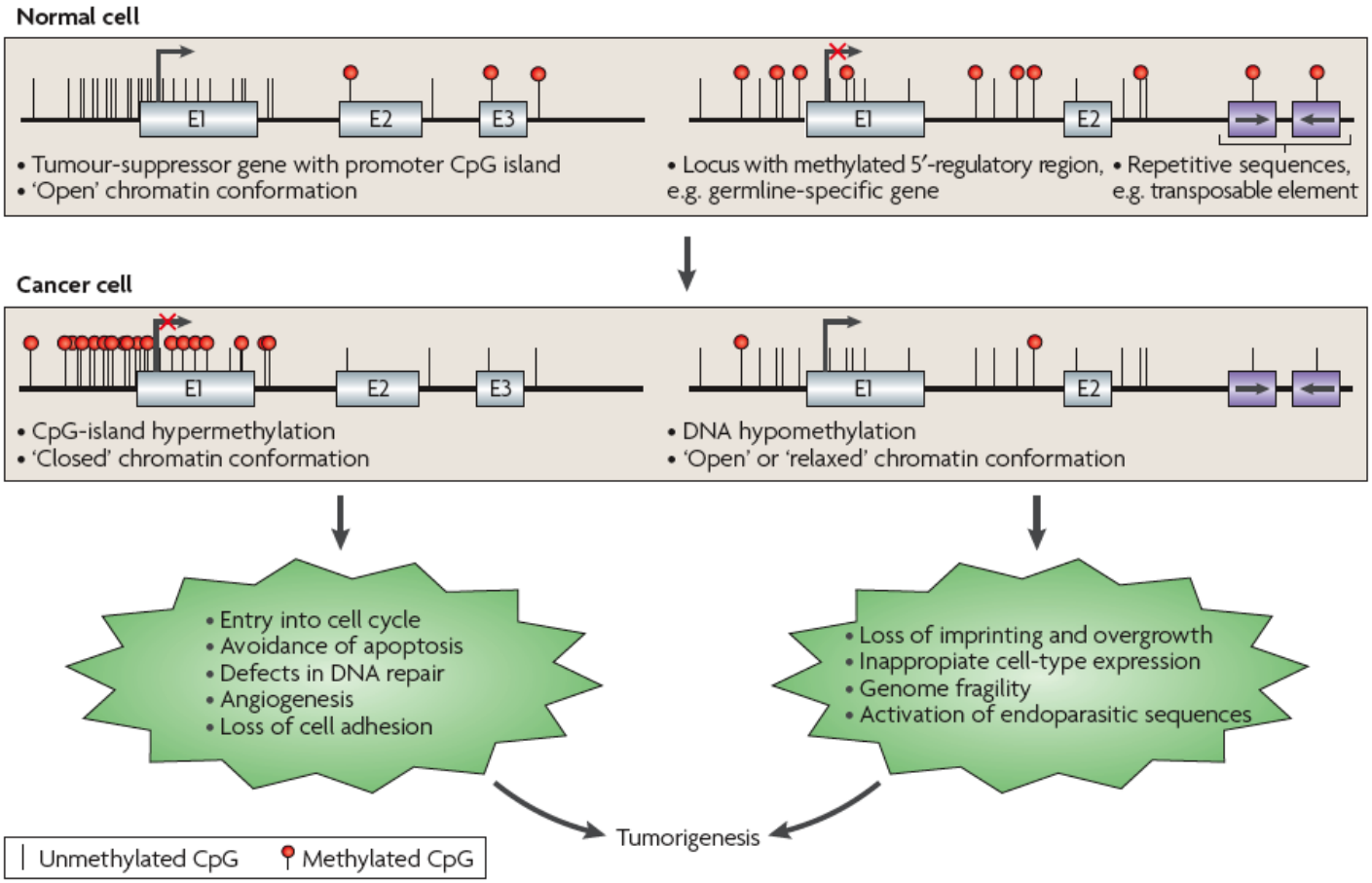
Covalent addition of methyl group (CH_3) to cytosine (almost exclusively at CpG sites in mammals); **binary status** at individual sites

CpG islands

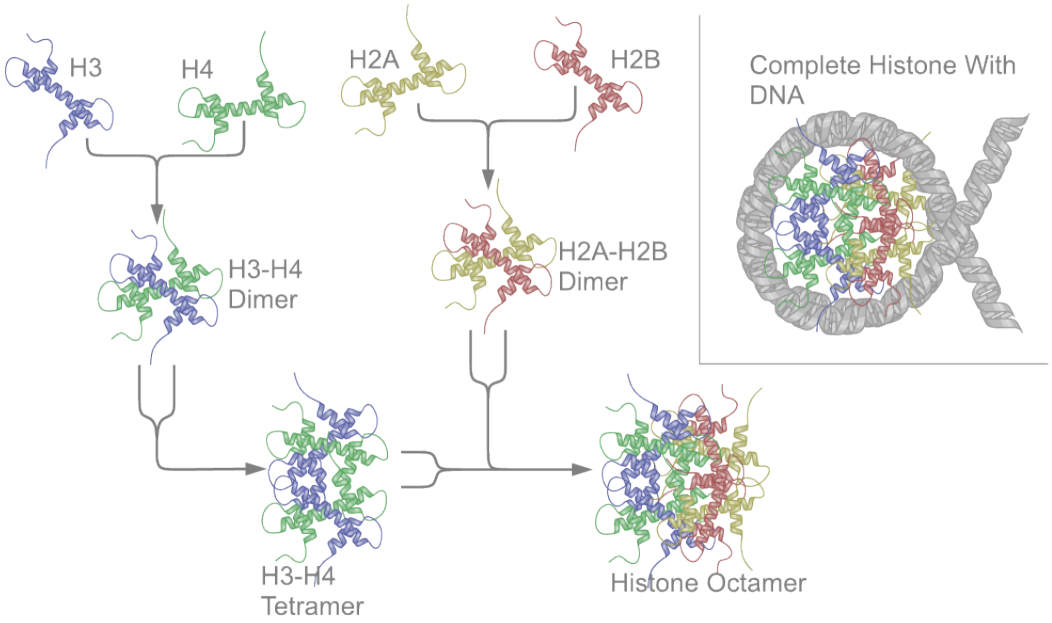
- CG dinucleotides are under-represented in the genome, but often occur in “clusters” called CpG islands (CGIs); various CGI definitions



Dogma: CpG methylation and transcription



Histone variants and post-translation modifications



Two of each of H2A, H2B, H3 and H4 form a "nucleosome", which 147bp of DNA can wrap around

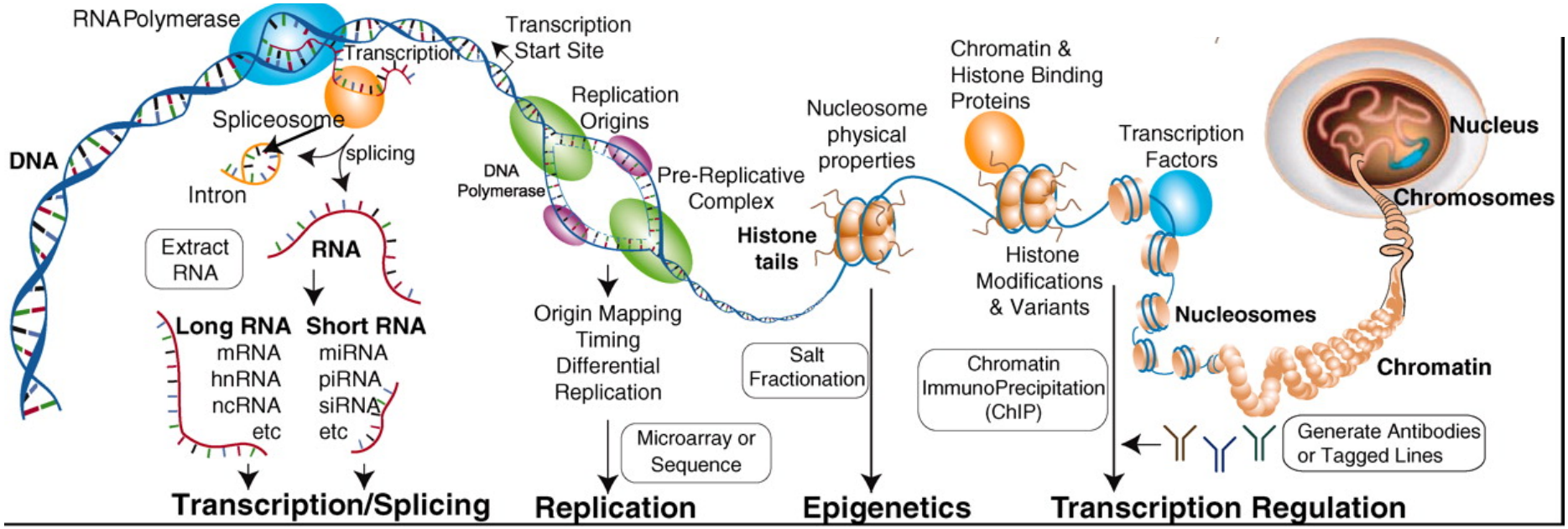
Histone variants and post-translation modifications

A very basic summary of the histone code for gene expression status is given below (histone nomenclature is described here):

Type of modification	Histone						
	H3K4	H3K9	H3K14	H3K27	H3K79	H4K20	H2BK5
mono-methylation	activation[6]	activation[7]		activation[7]	activation[7][8]	activation[7]	activation[7]
di-methylation		repression[3]		repression[3]	activation[8]		
tri-methylation	activation[9]	repression[7]		repression[7]	activation,[8] repression[7]		repression[3]
acetylation		activation[9]	activation[9]				

- H3K4me3 is found in actively transcribed promoters, particularly just after the transcription start site.
- H3K9me3 is found in constitutively repressed genes.
- H3K27me is found in facultatively repressed genes.[7]
- H3K36me3 is found in actively transcribed gene bodies.
- H3K9ac is found in actively transcribed promoters.
- H3K14ac is found in actively transcribed promoters.

Various other epigenetic (and regulator) factors



Roy et al. *Science* 2010

Genome/Epigenome Wide Association Studies (GWAS/EWAS)

GWAS – associating genotype to phentype

EWAS – association "epigenotype" to phenotype

Genetics does not explain a high amount of causality in common diseases

Challenge is far greater – there is 1 genome, but 1000s of epigenomes (100s of cell types, 10s-100s of epigenome dimensions)

But how does one conduct an EWAS? In addition to considerations that are common to both GWASs and EWASs (for example, adequate technology and sample size), the design of EWASs has specific considerations with respect to sample selection. DNAm patterns are specific to tissues and developmental stages, and they also change over time. Furthermore, EWAS associations can be causal as well as consequential for the phenotype in question — a difference from GWASs that presents considerable challenges. Here, we discuss these considerations in the context of designing and analysing an effective EWAS, keeping in mind that EWASs are likely to evolve, much like GWASs did, as information and experience accumulate.

Rakyan et al. 2011, Nature Reviews Genetics

Epigenetics and cancer

Most is known about DNA methylation. Cancers typically exhibit (of varying degrees associated with severity):

- Global DNA hypomethylation
- Region-specific hypermethylation, typically at CpG-island-associated promoters

Recent evidence highlights interruptions of epigenetic machinery from genetic mutations in cancer

University of Zurich[UZH]

Cancer Genome Sequencing -> Epigenetics

Institute of Molecular Life Sciences

Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin

Morin et al. (2010) Nature Genetics.

Ryan D Morin[1], Nathalie A Johnson[2], Tesa M Severson[1], Andrew J Mungall[1], Jianghong An[1], R
Jessica E Paul[1], Merrill Boyle[2], Bruce W Woolcock[2], Florian Kuchenbauer[2], Damian Yap[2], R Ke
Obi L Griffith[1], Sohrab Shah[2], Henry Zhu[3], Michelle Kimbara[3], Pavel Shashkin[3], Jean F Charlot[3],
Richard Corbett[1], Angela Tam[1], Richard Varhol[1], Duane Smailus[1], Michelle Moksa[1], Yongjun Z
Hong Qian[1], Inanc Birol[1], Jacqueline Schein[1], Richard Moore[1], Robert Holt[1], Doug E Horsman[4],
Steven Jones[1], Samuel Aparicio[2], Martin Hirst[1], Randy D Gascoyne[4] & Marco A Marra[1,6]

Follicular lymphoma (FL) and the GCB subtype of diffuse large B-cell lymphoma (DLBCL) derive from germinal center B cells[1]. Targeted resequencing studies have revealed mutations in various genes encoding proteins in the NF-κB pathway[2,3] that contribute to the activated B-cell (ABC) DLBCL subtype, but thus far few GCB-specific mutations have been identified[4]. Here we report recurrent somatic mutations affecting the polycomb-group oncogene[5] *EZH2*, which encodes a histone methyltransferase responsible for trimethylating lys27 of histone H3 (H3K27).

technology to sequence genomic DNA
malignant lymph node biopsy ("FL sar
individual with FL (Online Methods).
immunohistochemistry to
BCL2 and BCL6. This sa
because it had an unusu
Fig. 1), lacking the translo
scale alterations (Suppl
Tables 1 and 2). We analyz

Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes

Gillian L. Dalgliesh[1], Kyle Furge[2], Chris Greenman[1], Lina Chen[1], Graham Bignell[1], Adam Butler[1], Helen Davies[1], Sarah Edkins[1], Claire Hardy[1], Calli Latimer[1], Jon Teague[1], Jenny Andrews[1], Syd Barthorpe[1], Dave Beare[1], Gemma Buck[1], Peter J. Campbell[1], Simon Forbes[1], Mingming Jia[1], David Jones[1], Henry Knott[1], Chai Yin Kok[1], King Wai Lau[1], Catherine Leroy[1], Meng-Lay Lin[1], David J. McBride[1], Mark Maddison[1], Simon Maguire[1], Kirsten McLay[1], Andrew Menzies[1], Tatiana Mironenko[1], Lee Mulderrig[1], Laura Mudie[1], Sarah O'Meara[1], Erin Pleasance[1], Arjunan Rajasingham[1], Rebecca Shepherd[1], Raffaella Smith[1], Lucy Stebbings[1], Philip Stephens[1], Gurpreet Tang[1], Patrick S. Tarpey[1], Kelly Turrell[1], Karl J. Dykema[2], Sok Kean Khoo[3], David Petillo[3], Bill Wondergem[2], John Anema[4]. Richard J. Kahnoski[4]. Bin Tean Teh[3,5]. Michael R. Stratton[1,6] & P. Andrew Futreal[1]

Clear cell renal cell carcinoma (ccRCC) is the most common form of adult kidney cancer, characterized by the presence of inactivating mutations in the *VHL* gene in most cases[1,2], and by infrequent somatic mutations in known cancer genes. To determine further the genetics of ccRCC, we have sequenced 101 cases through 3,544 protein-coding genes. Here we report the identification of inactivating mutations in two genes encoding enzymes involved in histone modification—*SETD2*, a histone H3 lysine 36 methyltransferase, and *JARID1C* (also known as *KDM5C*), a histone H3 lysine 4 demethylase—as well as mutations in the histone H3 lysine 27 demethylase, *UTX* (*KMD6A*), that we recently reported[3]. The results highlight the role of mutations in components of the chromatin modification machinery in human cancer. Furthermore, *NF2* mutations were found in non-*VHL* mutated ccRCC, and several other probable cancer genes were identified. These results indicate that substantial genetic heterogeneity exists in a cancer type dominated by mutations in a single gene, and that systematic screens will be key to fully determining the somatic genetic architecture of cancer.

Much of our current understanding of cancer is based on the central tenet that it is a genetic disease, arising as a clone of cells that expands in an unregulated fashion because of somatically acquired mutations (*1*). These somatic mutations include base substitutions, insertions and deletions (indels) of bases, rearrangements caused by breakage and abnormal rejoining of DNA, and changes in the copy number of DNA segments. They also often include epigenetic changes that are stably inherited over mitotic DNA replication, for example, alterations in methylation of cytosine residues (*2*).

Stratton (2011) Science.

Dalgliesh et al. (2010) Nature.

H3K36me3,
H3K4me3,
H3K27me3

Epigenetic drugs

THE NEXT 10 YEARS — TIMELINE

A decade of exploring the cancer epigenome — biological and translational implications

Stephen B. Baylin and Peter A. Jones

Abstract | The past decade has highlighted the central role of epigenetic processes in cancer causation, progression and treatment. Next-generation sequencing is providing a window for visualizing the human epigenome and how it is altered in cancer. This view provides many surprises, including linking epigenetic abnormalities to mutations in genes that control DNA methylation, the packaging and the function of DNA in chromatin, and metabolism. Epigenetic alterations are leading candidates for the development of specific markers for cancer detection, diagnosis and prognosis. The enzymatic processes that control the epigenome present new opportunities for deriving therapeutic strategies designed to reverse transcriptional abnormalities that are inherent to the cancer epigenome.

Translational advances:

Biomarkers (e.g. GSTP1 in prostate cancer)

Therapeutics (e.g. azacitidine and decitabine have FDA approval for myelodisplastic syndrome, which can lead to leukemia)

FDA approval of vorinostat and romidepsin for cutaneous T cell lymphoma

HDAC inhibitors in clinical trials.

….

Not everyone is happy with these definitions

OPINION

Epigenetics: Core misconcept

Mark Ptashne[1]

Ludwig Professor of Molecular Biology, Memorial Sloan–Kettering Cancer Center, New York, NY 10021

A recent issue of PNAS opened with an unfortunate Core Concepts article titled "Epigenetics" (1). The author, a science writer, began promisingly enough:

> Despite the fact that every cell in a human body contains the same genetic material, not every cell looks or behaves the same….How does each cell retain its unique properties when, in its DNA-containing nucleus, it has the same master set of genes as every other cell?

Indeed understanding this problem has been an overarching goal of research in molecular, developmental, and, increasingly, evolutionary biology. And over the past 50 years a compelling answer has emerged from studies in a wide array of organisms. Curiously, the article ignores this body of knowledge, and substitutes for it misguided musings presented as facts.

Let me begin with a very brief overview of what drives development, a process that unfolds with essentially no changes in DNA sequence. Development of an organism from a fertilized egg is driven primarily by the actions of regulatory proteins called transcription factors. In sequential waves and combinations, these proteins bind to specific DNA sequences—called cis-regulatory sequences—associated with spe-